GENERALISED BOOTSTRAP FOR ESTIMATING EQUATIONS

by

Snigdhansu Chatterjee
Indian Statistical Institute

and

Arup Bose
Indian Statistical Institute
Purdue University

Technical Report #00-02

Department of Statistics
Purdue University
West Lafayette, IN USA

February 2000

GENERALISED BOOTSTRAP FOR ESTIMATING EQUATIONS

SNIGDHANSU CHATTERJEE *and* ARUP BOSE
Indian Statistical Institute — *Indian Statistical Institute and Purdue University*

Abstract. We introduce a generalised bootstrap technique for estimators obtained by solving estimating equations. Special cases of this generalised bootstrap encompasses classical bootstrap of Efron, the delete-d jackknife and variations of the Bayesian bootstrap. Under fairly general conditions we establish (a) asymptotic normality of the estimator and consistency of the bootstrap estimator when parameter dimension is fixed or increasing with data size; (b) asymptotic representation of the resampling variance estimator and (c) higher order accuracy of the new generalised bootstrap for the bias-corrected, studentised estimator. The examples of M-estimation in linear regression, nonlinear regression, generalised linear models, autoregression and co-integration are discussed in details.

Key words and phrases: Estimating equations, bootstrap, jackknife, Bayesian bootstrap, asymptotic normality, dimension asymptotics. Edgeworth expansion.

AMS 2000 Subject Classification: Primary 62G09, 62E20. Secondary, 62G05, 62F12, 62F40, 62M99

1 Introduction

Consider a triangular sequence of functions $\{\phi_{ni}(Z_{ni}, \beta),\ i = 1, \ldots, n,\quad n \geq 1\}$ taking values in $I\!R^p$, $\{Z_{ni}\}$ being a sequence of observable random variables and $\beta \in \mathcal{B} \subset I\!R^p$. Here, Z_{ni} can have two components, $Z_{ni} = (y_{ni} \ :\ \mathbf{x}_{ni})^T$ where y_{ni} can have an interpretation of a response with $\mathbf{x}_i$ covariates, as in a regression framework. Assume that $E\phi_{ni}(Z_{ni}, \beta_0) = 0,\ i = 1, \ldots, n\ \ n \geq 1$ for some $\beta_0 \in \mathcal{B}$. The "parameter" β_0 is typically unknown, and its estimate $\hat{\beta}_n$ is obtained by solving the estimating equations

$$\sum_{i=1}^{n} \phi_{ni}(Z_{ni}, \beta) = 0 \tag{1.1}$$

The random variables $\phi_{ni}(Z_{ni}, \beta_0)$ form a triangular array of martingale differences, and hence we call (1.1) martingale estimating equations. Examples are abundant in different contexts in statistics where estimators are obtained by solving martingale equations, see Godambe (1991) and Basawa, Godambe and Taylor (1997) for extensive discussion on estimating equations. Often an approximate zero of (1.1) serves as a solution wherein certain kinds of semiparametric regression and density estimation problems and local estimating equations may also be treated. Later we make some remarks on this. The major objective of this paper is to estimate features of $\hat{\beta}_n$ by a new approach resampling and to study their asymptotic properties.

We define our resample estimator $\hat{\beta}_{\mathrm{B}}$ as the solution of

$$\sum_{i=1}^{n} w_{ni}\phi_{ni}(Z_{ni}, \beta) = 0 \tag{1.2}$$

where the bootstrap weights $\{w_{ni}, i = 1, \ldots, n,\quad n \geq 1\}$ is a triangular sequence of random variables, independent of $\{Z_{ni}\}$. These are the so called 'bootstrap weights'. We discuss the conditions on w_{ni} in Section 5. The concept of resampling with (1.2) may be traced back to Freedman and Peters (1984) and Rao and Zhao (1992). A major point about the present paper and the above two references is that these are on "bootstrap" using equations, as distinguished from the more traditional approach of resampling using observations. In the context of least squares estimator in linear regression, Chatterjee and Bose (1999) introduced the *uncorrelated weights bootstrap* (*UBS*) which is a generalisation of the paired bootstrap and the delete-d jackknives. The present paper uses the same kind of weights in a martingale estimating equations framework, hence we often refer to the bootstrap weights $\{w_{ni}\}$ as "UBS" weights. This is because the asymptotic uncorrelated nature of the weights is a major aspect for the *UBS* method, which serves to delineate various block bootstrap methods from the *UBS*. It may be noted that the *UBS* technique is different from the bootstraps suggested by Lele (1991) and Hu and Kalbfleisch (2000) in the context of estimating equations.

Under broad conditions on the weights and on the ϕ_{ni}, we establish the consistency of the *UBS* estimator for the distribution function of $\hat{\beta}_n$. We also obtain an asymptotic representation of the UBS variance estimator along the lines of Liu and Singh (1992) and Chatterjee and Bose (1999).We generally consider models with fixed dimension. However, in one section the bootstrap consistency is established even when the parameter dimension p tends to infinity with data size n. Typically, the asymptotic linearizations hold when $p^2/n \to 0$.

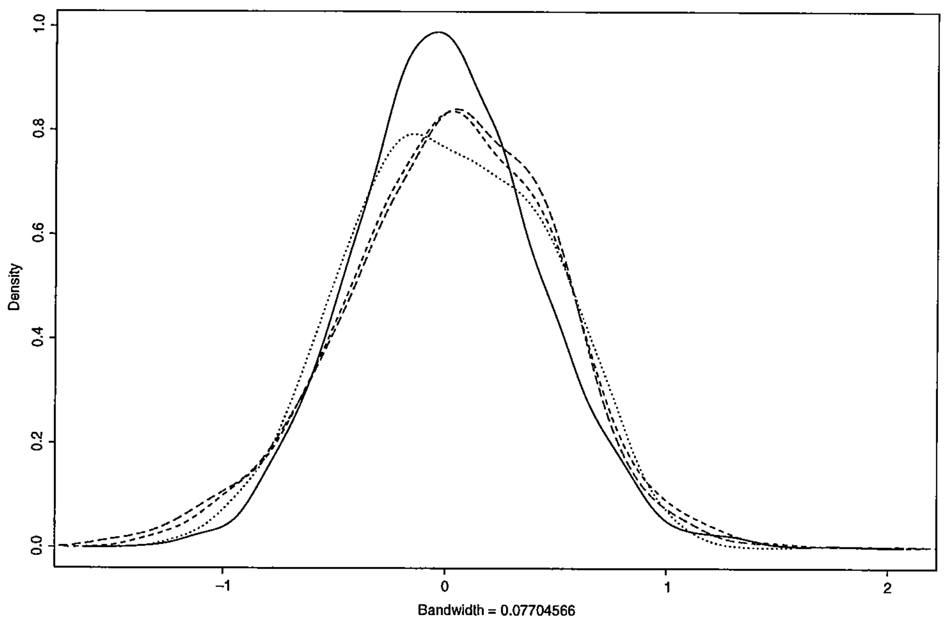
In many situations, naively resampling from the data leads to inconsistency. Consider the model $y_i = \beta_0^T g(\mathbf{x}_i) + e_i$, and β_0 is estimated by least squared error. Consider the two paired bootstraps which draw simple random sample with replacement from the pairs $(y_i, g(\mathbf{x}_i))$, and $(y_i, \mathbf{x}_i)$ respectively. These two procedures do not result in identical estimating equations. Let $\xi_i = (\xi_{i1}, \ldots, \xi_{in})$ be distributed as $Multinomial(1, 1/n, \ldots, 1/n)$. Then the first paired bootstrap yields the'squared error' term $\sum_{j=1}^n \xi_{ij}(y_j - \beta_0^T g(\mathbf{x}_j))^2$, but the second yields $\sum_{j=1}^n (\xi_{ij}^{1/2} y_j - \beta_0^T g(\xi_{ij}^{1/2} \mathbf{x}_j))^2$. The latter can lead to inconsistency, as observed in Stute, Gonzalez Manteiga and Presedo Quindimil (1998). More such examples may be found in Shao and Tu (1995). Our approach helps clarify the correct way of bootstrapping in such situations.

A popular approach to resampling is by using "residuals" by using estimated parameters as plug-in. However, typical residual based resampling techniques are not robust. For example, the classical residual bootstrap for the least squares estimator in linear regression is inconsistent under heteroscedasticity. On the other hand the paired bootstrap, is consistent under heteroscedasticity. The UBS, which may be seen as a generalisation of the paired bootstrap, is consistent under fairly general conditions. The robustness aspect is apparent from the fact that several diverse kinds of problems and several resampling techniques may be studied in a unified way using the approach of this paper. In linear regression the trade-off for robustness is in a loss of efficiency, see for example Liu and Singh (1992). Interestingly, for more general M-estimation problems the UBS can be simultaneously robust and efficient. Also, the results of Mammen (1996) serve as a caveat against using residuals in many situations.

Another visible advantage of the UBS over other comparable techniques (of Lahiri (1992), for example) is its simplicity in use. The practitioner can use the same code to obtain the estimator as well as its resamples, since the weights $\{w_{ni}\}$ are often easily obtained. This advantage can be carried further. Typically (1.1) is solved using iterative techniques like the *iteratively reweighted least squares* ($IRLS$) technique. In such cases sometimes (1.2) may be solved as one would solve a weighted least squares problem. Thus for each bootstrap Monte Carlo sample only one linear equation is solved, instead of an entire iterative scheme, thus greatly reducing the computational burden. It is obvious that the one-step method would not work well always, but whenever it does it is a great computational advantage. This aspect requires further study.

The choice of weights is also an important computational issue. As a test case we study the model $y_i = \beta_1 + \exp(-\beta_2 x_i) + e_i$, where $\beta_1 = 10$, $\beta_2 = 1.25$, x_i's are *i.i.d.* from $Uniform(0, 1)$, and e_i's are *i.i.d.* from $Normal(0, 1)$. A sample of size $n = 30$ was taken, and (β_1, β_2) estimated by using least squared errors. This requires iteration, and in order to get an initial estimate of the parameter we used the following steps: we computed $y_0 = 0.9 \min_{1 \le i \le n} y_i$, then used ordinary least squares to regress $\log(y_i - y_0)$ on $-\beta_2 x_i$, to get an estimate β_{20} of β_2. Then we took the average of $y_i - \exp -\beta_{20} x_i$ as β_{10}, the initial estimate of β_1. The constant 0.9 used in the definition of y_0 was fixed arbitrarily.

Figure 1: Different bootstrap density estimators



In Figure 1 we plotted the density of the least squares estimator of β_2, simulated from 1000 trials of the above experiment. By fixing one such dataset, we obtained different UBS estimators also for the same quantity. The bold line shows the density of the least squares estimator, the dashed line the bootstrap estimator using $Multinomial(n, 1/n, \ldots, 1/n)$ weights, the double-dash line is the one step iteration free estimator described in this section for the same bootstrap weights; and the dotted line is UBS estimator with *i.i.d.* $Uniform(1/2, 3/2)$ weights. In each case a kernel smoothing with Gaussian kernel was used, and the common bandwidth used is quoted in the figure. The figure shows that all the three bootstrap methods yield equally good results. A number of other models were tried for simulation. The broad conclusion is that the different choices of UBS weights yield similar results, but the performance of the non-iterative one-step method is variable. Also it is not clear whether the choice of bootstrap weights influences the accuracy of the one-step iteration.

In subsection 2.1 we obtain asymptotic representations of $\hat{\beta}_n$ and $\hat{\beta_{\mathrm{B}}}$ when β is real valued. This helps establish asymptotic normality and consistency of the UBS for estimating the distribution. The conditions on ϕ_{ni} are reasonably weak but there is scope for further relaxation.

For the important case where $\phi_{ni},\ i = 1, \ldots, n$, are independent, we establish an asymptotic representation of the UBS variance estimator, along the lines of Chatterjee and Bose (1999) and Liu and Singh (1992) in subsection 2.2. This implies the consistency of UBS under heteroscedasticity for a very broad class of estimators and models, including M-estimators in linear regression, generalised linear models and non-linear regressions. However, the classification of common resampling schemes as *efficient* or *robust* along the lines of Liu and Singh (1992) does not hold generally for M-estimators in linear regression. The UBS estimators are always robust, sometimes they are also efficient.

The estimators $\hat{\beta}_n$ and $\hat{\beta_{\text{B}}}$ are typically biased. In subsection 2.3 we restrict to independent ϕ_{ni} and consider the distribution of bias corrected, studentised $\hat{\beta}_n$ and show that the distribution of an appropriate studentised $\hat{\beta_{\text{B}}}$ when certain *i.i.d.* weights are used, yield a second order accurate approximation.

In subsection 2.4 we consider general p dimensional parameters ($p \to \infty$ as $n \to \infty$. Asymptotic normality and consistency of UBS results are obtained for the estimator of a linear combination of parameters.

In Section 3 we discuss several well known resampling techniques which the UBS class encompasses. Broadly speaking, the conditions for distributional consistency results are satisfied by all resampling schemes except the delete-d jackknives for which $d/n \to c \in (0,1)$ is not true. The higher order accuracy results are valid only for *i.i.d.* absolutely continuous weights, for example the typical weights used in "bootstrap clone" methods. Extensions to more general UBS weights is under investigation.

Estimating equations have been used in a wide variety of problems, including likelihood and quasi-likelihood methods, semiparametrics and nonparametrics, time series, biostatistics, stochastic processes, spatial statistics, robust inference, survey sampling to name a few. Broadly, the UBS resampling applies to all such problems, subject to adequate technical assumptions being made. In Section 4 we illustrate the use of the techniques described in this paper in the context of some specific problems like non-linear regression, M-estimation in linear models, generalised linear models, and autoregressive processes. We discuss how traditional assumptions in these problems turn out to be sufficient for the assumptions that we make in this paper. However, it would be interest to explore UBS in these and other problems individually, to further understand the nature of UBS and discover the minimal conditions under which it works in a given problem.

Our results in this paper concentrate on the case where the parameter of interest is finite dimensional, but the treatment for infinite dimensional parameters present no additional conceptual difficulty. While a detailed analysis would require an independent study, in Section 4 we show how to extend our ideas adequately for resampling for the parametric component in semiparametric models.

The results of this paper require diverse technical conditions. We put all these conditions together in an appendix (Section 5). Our different theorems require different conditions both on the model as well as on the UBS weights. The results and remarks of Section 2 are based on these technical conditions. Another appendix (Section 6) contains the proofs of the theorems.

2 Main Results

The technical conditions and the assumptions needed for the results are all given in Section 5. We do not refer to this fact in the statements of the theorems.

We establish consistency of the estimator and an asymptotic representation, in Theorem 2.1 for one dimensional parameter and in Theorem 2.5 for general p dimensional parameter, allowing p to tend to infinity with n. Asymptotic normality of the estimator follows as a corollary. The asymptotic linearised representation for the UBS estimator are given in Theorems 2.2 and 2.6 for $p = 1$ and general p respectively. The consistency of the UBS distribution estimator follows from this.

Assuming in addition that $\phi_{ni},\ i = 1, \ldots, n$ are independent, we obtain two more results. In Theorem 2.3 an asymptotic representation of the UBS variance estimator is obtained. After bias correction and studentisation of the original estimator as well as its UBS equivalent, a second order accuracy result is obtained in Theorem 2.4. For these two results we consider one dimensional parameter only for technical convenience. Further details and remarks on the results are given in the different subsections.

We adopt the notations $\phi_{(k+1)ni}(\lambda) = \frac{\partial}{\partial\lambda}\phi_{kni}(\lambda)$, for $k = 0, 1, 2, \ldots$, where $\phi_{0ni}(\lambda) \equiv \phi_{ni}(\lambda)$; $\phi_{kni}(\beta_0) = \phi_{kni}$ for $k = 0, 1, 2 \ldots$.

2.1 Asymptotics and bootstrap for $p = 1$

Let $\sum_{i=1}^n E\phi_{ni}^2 = O(a_n^2)$, and $\gamma_n = [\sum_{i=1}^n \phi_{ni}^2]^{-1/2} \sum_{i=1}^n \phi_{1ni}$.

Theorem 2.1 *Under* $(A0)$-$(A3)$ *exists a sequence* $\{\hat\beta_n\}$ *of solutions of* (1.1) *such that*

$$a_n(\hat\beta_n - \beta_0) = O_P(1) \text{ and } \quad (2.1)$$

$$\gamma_n(\hat\beta_n - \beta_0) = -(\sum_{i=1}^n E\phi_{ni}^2)^{-1/2} \sum_{i=1}^n \phi_{ni} + r_n \quad (2.2)$$

where $r_n = o_P(1)$. *If further,* $c_n = O(n)$, $b_n = O(1)$ *and* $d_n = O(1)$, *then* $r_n = O_P(a_n^{-1})$.

Henceforth we work with that sequence of estimates $\{\hat\beta_n\}$ which satisfies Theorem 2.1. Assumption $(A4)$ ensures that $\sum_i X_{ni} = a_n^{-1}s_n^{-1}\phi_{ni} \overset{\mathcal{D}}{\Rightarrow} N(0, 1)$ by Theorem 3.2 of Hall and Heyde (1980). The usual case of '$n^{1/2}$-consistency' is when $a_n = n^{1/2}$. The following corollary is immediate.

Corollary 2.1 *Assume* $(A0)$-$(A4)$. *Then*

$$\gamma_n(\hat\beta_n - \beta_0) \overset{\mathcal{D}}{\Rightarrow} N(0, 1) \quad (2.3)$$

Let

$$F_n(x) = Prob[\gamma_n(\hat{\beta}_n - \beta_0) \leq x] \qquad \hat{g}_n = a_n^{-2} \sum_{i=1}^{n} \phi_{ni}^2(\hat{\beta}_n)$$

$$\hat{\gamma}_{1n} = a_n^{-2} \sum_{i=1}^{n} \phi_{1ni}(\hat{\beta}_n) \qquad \gamma_{nB} = \sigma_n^{-1}[\sum_{i=1}^{n} \phi_{ni}^2(\hat{\beta}_n)]^{-1/2} \sum_{i=1}^{n} \phi_{1ni}(\hat{\beta}_n)$$

The bootstrap estimator is obtained by solving (1.2). The next theorem is on its asymptotic representation. For any bootstrap random variable T_{Bn}, we use the notations $T_{Bn} = O_{PB}(\xi_n)$ and $T_{Bn} = o_{PB}(\xi_n)$ if for any fixed $\epsilon > 0$, the bootstrap probability conditional on the data $\mathrm{P_B}[\xi_n^{-1}|T_{Bn}| > \epsilon]$ are respectively $O_P(1)$ and $o_P(1)$.

Theorem 2.2 *Assume (A0)-(A3) and the bootstrap weights satisfy conditions* (5.1) - (5.3) . *Then there exists a sequence* $\{\hat{\beta_B}\}$ *of solutions of* (1.2) *such that*

(*a*) *for fixed* $\epsilon, \delta > 0$, $\exists K > 0$ *and integer* n_0 *such that for all* $n \geq n_0$

$$Prob[\mathrm{P_B}({\sigma_n}^{-1} a_n |\hat{\beta_B} - \hat{\beta}_n| \leq K) < 1 - \epsilon] < \delta \tag{2.4}$$

(*b*) *for some* $r_{nB} = o_{PB}(1)$,

$$\gamma_{nB}(\hat{\beta_B} - \hat{\beta}_n) = -(\sum_{i=1}^{n} E\phi_{ni}^2)^{-1/2} \sum_{i=1}^{n} W_i \phi_{ni} + r_{nB} \tag{2.5}$$

Further, if $c_n = O(n)$, $b_n = O(1)$ *and* $d_n = O(1)$, *then* $r_{nB} = O_{PB}(a_n^{-1})$.

Let $F_{Bn}(x) = \mathrm{P_B}[\gamma_{nB}(\hat{\beta_B} - \hat{\beta}_n) \leq x]$. The consistency of the UBS distribution estimator F_{Bn} for estimating F_n is a consequence of the above two theorems and Theorem 6.2. This is stated in the following corollary.

Corollary 2.2 *Let* $\{w_{ni}, i = 1, \ldots n\}$ *be exchangeable for every* n. *Assume* (*A*0)-(*A*4) *and conditions* (5.1)-(5.5) *on the bootstrap weights. Then*

$$\sup_x |F_{Bn}(x) - F_n(x)| \to 0 \textit{ in probability} \tag{2.6}$$

REMARK 1 (On conditions on ϕ_{ni}.) Assumption ($A2$) is less restrictive than the Lipschitz condition (C.1) of Lahiri (1992). We compare our assumptions and results in details with that of Lahiri (1992) in subsection 2.3 where we consider higher order accuracy of UBS.

Existence of second derivative ϕ_{2ni} of ϕ_{ni} and related assumptions may be replaced by the assumption that ϕ_{ni} admits the Taylor series approximation of the form $\phi_{ni}(a + t) = \phi_{ni}(a) + t\phi_{1ni}(a) + o(|t|)$ as $|t| \to 0$. Note that the functions ϕ_{ni} are random, so the remainder term has a probabilistic interpretation. The assumption needed in that case would be similar to assumption (h1) of Koul (1996).

The assumption (5.19) that $a_n^{-2} \sum \phi_{1ni}(\lambda) > 0$ for any λ may be replaced by requiring only $a_n^{-2} \sum \phi_{1ni}(\beta_0) > 0$. Then the proof of Theorem 2.1 remains unchanged, but the proof of Theorem 2.2 has to be modified, and the class of UBS techniques slightly restricted.

REMARK 2 (On bootstrap weights). Many schemes like the classical bootstrap and the delete-d jackknife satisfy $\sum_{i=1}^{n} w_{ni} = n$ (sometimes after scaling to ensure (5.1)). This implies that the correlation between w_{ni} and w_{nj} is $-(n-1)^{-1}$ ensuring that our condition (5.3), $c_{11} = O(n^{-1})$ is satisfied. Interestingly, different block bootstrap methods (see Kunsch (1989), for example) can also be identified with the framework of (1.2); however, there the bootstrap weights satisfy $c_{11} \to 1$. Without (5.3), the calculations involved in the proof of Theorem 2.2 may still be carried out, but the rate conditions would change and the rates of convergence are slower. Block bootstrap techniques are powerful tools used for resampling in case of weak dependence, often used in problems where assumptions on the dependency structure is minimal. However, in our framework dependencies are of a special kind owing to our 'martingale' assumptions, and block bootstrap techniques are not required here. The name *uncorrelated weights bootstrap* is prompted by the importance of this asymptotic uncorrelatedness of weights condition.

2.2 Asymptotic representation of variance estimator

Since the asymptotic distribution of $\hat{\beta}_n$ is normal, one may wish to estimate the asymptotic variance of $\hat{\beta}_n$ using bootstrap. In this subsection we obtain the asymptotic representation of the UBS variance estimator.

Liu and Singh (1992) obtained asymptotic representation of variance estimates from different resampling techniques for the least squares estimator in linear regression. They classified the resampling techniques as either *robust* against heteroscedasticity (R-class) or relatively *efficient* under homoscedasticity (E-class). An important corollary of our representation result is that such a classification does not carry over to general M-estimation problems.

For this subsection and next, we restrict our attention to the case ϕ_{ni}, $i = 1, \ldots, n$ *independent.* Without this, the result of this subsection would be much weaker, and that of the next subsection would require entirely different treatment. The UBS weights are assumed to satisfy conditions (5.1), (5.3), (5.6) and either (5.7)-(5.9) or (5.10)-(5.12) which are slightly stronger moment requirements than those of subsection 2.1.

Further, certain resampling schemes like the paired bootstrap and the delete-d jackknives effectively select subsets of the data in the resample, and model assumptions are often required to hold on these subsets (see Wu (1986), Chatterjee and Bose (1999)). This motivates the assumption (5.34). It helps us to show that under appropriate conditions the probability of a 'bad' set is small and helps to define the bootstrap estimator for this subsection.

Proposition 2.1 *Assume ϕ_{ni} are independent satisfying conditions* (5.27)-(5.30) *with* $L = 8(1+\alpha)$ *and* (5.34). *Assume $\hat{\beta}_n$ is a solution to* (1.1) *satisfying* (2.1)-(2.2) *from Theorem 2.1. Let $\mathcal{A}$ be the set on which $m^{-1}\sum_{i\in\mathcal{I}_m}\phi_{1ni}(\hat{\beta}_n) > k_1/2 > 0$ for every such choice of subset $\mathcal{I}_m$ of size m from $\{1,2,\ldots,n\}$ and for every m in $[m_0, n]$. Then $Prob[\mathcal{A}] > 1 - O(n^{-2})$.*

We define our bootstrap estimator $\hat{\beta_{\mathrm{B}}}$ to be the solution to (1.2) under the set $\mathcal{A}\cap\mathcal{W}$, and $\hat{\beta}_n$ otherwise. The set $\mathcal{A}$ is defined in Proposition 2.1, and $\mathcal{W}$ is defined in Section 5. The *UBS* variance estimate is $\mathrm{V}_{\mathrm{UBS}} = \sigma_n{}^{-2}\mathrm{E}_{\mathrm{B}}(\hat{\beta_{\mathrm{B}}} - \hat{\beta})^2$. This is used for estimating $\mathrm{V_n} = E(\hat{\beta} - \beta_0)^2$. In the statement of the next theorem we have used ϕ, ϕ_1, ϕ_2 respectively for ϕ_{ni}, ϕ_{1ni}, ϕ_{2ni}. The sums range from 1 to n. Also let $g_{1n} = n^{-1}\sum E\phi_1$, $g_{2n} = n^{-1}\sum E\phi_2$.

Theorem 2.3 *Assume $\{\phi_{ni},\ i = 1,\ldots,n\}$ are independent satisfying* (5.27)-(5.31) *and* (5.34) *with $L = 8(1+\alpha)$. Then*

$$\begin{aligned} & ng_{1n}^2(\mathrm{V_{UBS}} - \mathrm{V_n}) \\ = \ & n^{-1}\sum(\phi - E\phi)^2 - \frac{2}{n^2 g_{1n}}\sum\phi\sum\phi\phi_1 - \frac{2}{n^2 g_{1n}}\sum\phi^2\sum(\phi_1 - E\phi_1) \\ & + \frac{2}{n^3 g_{1n}^2}\sum\phi\sum\phi^2\sum\phi_2 + O_P(n^{-1}) \end{aligned} \tag{2.7}$$

REMARK. Liu and Singh (1992) showed that for least squares estimator in linear regressions, a weighted jackknife scheme is more efficient than the usual jackknife if errors are homoscedastic but it is not consistent under heteroscedasticity. For general M estimates in simple linear regression, we have $\phi_{ni} = -x_i\psi(y_i - \beta x_i)$. Bose and Kushary (1996) showed that the above comparison does not remain valid and even in common criterion functions like Tukey's biweight function the jackknife can be more efficient than the weighted jackknife. This also follows from the asymptotic representation (2.7). So the *UBS* is not only robust against heteroscedasticity of errors, but can be more efficient than the weighted jackknife in some homoscedastic cases.

2.3 Higher order accuracy of *UBS*

For M-estimators in linear regression, Lahiri (1992) established higher order accuracy of some residual based resampling techniques. It is generally understood that the paired bootstrap is not second order accurate, owing to a second order non-negligible bias term. Since the *UBS* is a generalisation of the paired bootstrap, it is also not second order accurate in general.

The solution $\hat{\beta}_n$ of (1.1) also has a bias factor that is not generally second order negligible. A natural approach is to try and estimate this bias factor and study the asymptotic properties of the bias corrected statistic.

With technical conditions stated later on, and by calculations similar to those of the proof of Theorem 2.3, we obtain that there is a solution $\hat{\beta}_n$ to (1.1) that satisfies

$$-[n^{-1}\sum \phi_1][n^{1/2}(\hat{\beta}_n - \beta_0)] = n^{-1/2}\sum \phi - 2^{-1}n^{-1/2}[n^{-1}\sum \phi_2][n^{1/2}(\hat{\beta}_n - \beta_0)]^2 + r_{n1}$$

where $Er_{n1}^2 = O(n^{-(1+\alpha)})$. The bias correction to this statistic can be done by estimating the second term on the right side of the above expression. This requires an estimate for $E[n^{1/2}(\hat{\beta}_n - \beta_0)]^2$. One such estimate would be $\mathrm{V}_{\mathrm{UBS}}$ from Theorem 2.3. However, a different variance estimator, like the simple plug-in $\hat{g}_n^2 = n^{-1}\sum \phi_{ni}^2(\hat{\beta}_n)$ may also be used. Let $g_n^2 = n^{-1}\sum \phi_{ni}^2$ and $\hat{\gamma}_{kn} = n^{-1}\sum \phi_{kni}(\hat{\beta}_n)$, $k = 1, 2$. After some routine algebra, we obtain the following bias corrected studentised estimator

$$T_n = \hat{\gamma}_{1n}\hat{g}_n^{-1}[n^{1/2}(\hat{\beta}_n - \beta_0)] - 2^{-1}n^{-1/2}\hat{\gamma}_{1n}^{-2}\hat{g}_n^{-1}\hat{\gamma}_{2n}\mathrm{V}_{\mathrm{UBS}} = -n^{-1/2}g_n^{-1}\sum \phi_{ni} + r_n \quad (2.8)$$

where $Er_n^2 = O(n^{-(1+\alpha)})$.

For $\hat{\beta}_{\mathrm{B}}$, define $\hat{g}_{nB}^2 = n^{-1}\sum W_i^2\phi_{ni}^2(\hat{\beta}_n)$ and $g_{nB}^2 = n^{-1}\sum W_i^2\phi_{ni}^2$. Then the bootstrap bias corrected, studentised statistic is given by

$$\begin{aligned} T_{nB} &= \hat{\gamma}_{1n}\hat{g}_{nB}^{-1}[\sigma_n{}^{-1}n^{1/2}(\hat{\beta}_{\mathrm{B}} - \hat{\beta}_n)] + 2^{-1}n^{-1/2}\sigma_n\hat{g}_{nB}^{-1}\hat{\gamma}_{2n}[\sigma_n{}^{-1}n^{1/2}(\hat{\beta}_{\mathrm{B}} - \hat{\beta}_n)]^2 \\ &= -n^{-1/2}g_{nB}^{-1}\sum W_i\phi_{ni} + r_{nB} \quad (2.9) \end{aligned}$$

where $\mathrm{P}_{\mathrm{B}}[|r_{nB}| > kn^{-1/2}(\log n)^{-1}] \to 0$ almost surely.

The slight difference in the 'bias correction terms' in (2.8) and (2.9) is because in the former we used an estimator for $[n^{1/2}(\hat{\beta}_n - \beta_0)]^2$, whereas in the latter case $[\sigma_n{}^{-1}n^{1/2}(\hat{\beta}_{\mathrm{B}} - \hat{\beta}_n)]^2$ was directly used. T_n and T_{nB} are approximated by studentised sample sums of possibly non-identical but independent random variables. To exploit this, we need Edgeworth expansions for such variables. Such a result is established in Section 6. Let $F_{n1}(x) = Prob[T_n \leq x]$ and $F_{nB1}(x) = \mathrm{P}_{\mathrm{B}}[T_{nB} \leq x]$.

We assume (5.26) is satisfied for some $\alpha \in (0, 1]$, and the model satisfies conditions (5.27)-(5.30) with $L = 16$ and (5.32)-(5.33). For any constant $C > 1$, let $p_i(C) = Prob[\phi_{ni}^2 \in [C^{-1}, C]]$. Additionally, assume:

Assumption A $\{\phi_{ni}\}$ is a sequence of independent, absolutely continuous random variables defined on the real line. Let the density of ϕ_{ni} be $f_{ni}(\cdot)$. Define the function $h_{ni}(x) = \min\{f_{ni}(x), f_{ni}(-x)\}$, and let $H_n(x) = n^{-1}\sum_{i=1}^n h_{ni}(x)$. We assume that for all large n, $\exists$ a real valued function $g(x) \geq 0$ such that $\mathcal{A} = \{x : H_n(x) \geq g(x)\}$ has positive measure, and $0 < \int_{\mathcal{A}} g(x)dx < \infty$.

Assumption B For some $c \in (0, 1)$, there exists a constant $C > 1$ such that $n^{-1}\sum p_i(C) \geq c$ for all large n.

For the bootstrap, we assume that the UBS weights are *i.i.d.* absolutely continuous, and satisfy (5.1), (5.7), (5.13)-(5.15). Additionally, assume:

Assumption C Assume that the distribution of W_1 has a density $f(x)$ such that for some positive numbers x_0 and b, we have $f(x) > b$ for all $x \in (-x_0, x_0)$.

Theorem 2.4 *Under the assumptions of this section, we have*

$$\sup_x |F_{nB1}(x) - F_{n1}(x)| = o(n^{-1/2}) \text{ almost surely} \tag{2.10}$$

REMARK 1 Corollary 2.2 is a weaker result compared to Theorem 2.4 which was obtained under weaker model assumptions.

REMARK 2 Lahiri (1992) considers M-estimation in linear regression with non-random regressors and *i.i.d.* errors, whereas our model set-up is much broader. When considering M-estimators in linear regression, we allow random regressors and heteroscedastic errors in our framework. A quantity γ_n defined in Lahiri (1992, C.6), which is dependent on the design variables, appears as a term in the various convergence rates of Lahiri. But for us it may be random and thus cannot appear in the convergence rates. Lahiri obtains second order consistency of residual based bootstrap techniques. The resampling techniques we consider may be seen as generalisations of the paired bootstrap. The moment assumptions of Lahiri are of lower order than what we assume, and no bias correction is needed for estimators considered by him. The assumption of absolute continuity of weights was not required by Lahiri. By assuming absolute continuity of our weights, we ignore many important resampling estimators.

Thus Lahiri makes more relaxed assumptions than us and obtains second order accuracy in estimating the distribution of M-estimators in linear regression, for a residual based bootstrap. Our set of problems is much larger, the assumptions more stringent and the resampling schemes different from Lahiri. It is plausible that with more effort Theorem 2.4 can be proved under much weaker assumptions like strong-nonlatticeness instead of absolute continuity.

2.4 Dimension asymptotics

In this subsection we consider dimension asymptotics, that is, we allow $p \to \infty$ as the data size $n \to \infty$. Dimension asymptotics has been a major aspect of study in resampling in linear regression framework (Bickel and Freedman (1983), Mammen (1989, 1993)). Classical residual based bootstrap has been studied for the least squares estimator (Bickel and Freedman (1983)) and for general M-estimators (Mammen (1989)) using non-random design matrix. The random design case and resampling using paired bootstrap and wild bootstrap has been studied in Mammen (1993). This subsection is an attempt to explore the high dimensionality aspect in more general problems.

We obtain the consistency of UBS distribution estimator, extending the result of subsection 2.1. The asymptotic representation of variance and second order accuracy of distribution estimation may also be considered, but with additional complications in technicalities involved.

The mutual relation between n and p is inherent in the assumptions made. Suppose $\sup_{||c||=1} \sum_{i=1}^{n} E(c^T \phi_{ni})^2 = O(a_n^2)$ and $s_n^2 = p^{-2}[(\sum_{i=1}^{n} E\phi_{1ni})^{-1}c]^T[\sum_{i=1}^{n} E\phi_{ni}\phi_{ni}^T]$. The model assumptions are given in Section 5.

Theorem 2.5 *Under* $(B0)$-$(B3)$ *exists a sequence* $\{\hat{\beta}_n\}$ *of solutions of* (1.1) *such that*

(*a*) *if* $pa_n^{-1} \to 0$, *then*

$$a_n p^{-1/2} ||(\hat{\beta}_n - \beta_0)|| = O_P(1) \text{ and} \tag{2.11}$$

$$a_n^{-1} p^{-1/2} (\sum_{i=1}^{n} E\phi_{1ni})^T (\hat{\beta}_n - \beta_0) = -a_n^{-1} p^{-1/2} \sum_{i=1}^{n} \phi_{ni} + r_n \tag{2.12}$$

where $E||r_n||^2 = O(a_n^{-2}p^2)$.

(*b*) *if* $p^3 a_n^{-2} \to 0$, *then for any* $c \in I\!R^p$ *with* $||c|| = 1$,

$$s_n^{-1} c^T (\hat{\beta}_n - \beta_0) = -s_n^{-1} [(\sum_{i=1}^{n} E\phi_{1ni})^{-1} c]^T \phi_{ni} + o_P(1) \tag{2.13}$$

$$= \sum_{i=1}^{n} X_{ni} + o_P(1) \tag{2.14}$$

(*c*) *if* $pa_n^{-2} \to 0$ *and* $(B4)$ *holds, then*

$$a_n^{-1} p^{-1/2} (\sum_{i=1}^{n} \phi_{1ni})^T (\hat{\beta}_n - \beta_0) = -a_n^{-1} p^{-1/2} \sum_{i=1}^{n} \phi_{ni} + r_n \tag{2.15}$$

where $E||r_n||^2 = O(a_n^{-2}p^2)$.

Part (c) of the above theorem is actually proved under slightly weaker assumptions. The conditions listed in $(B4)$ is required only for $\lambda = \beta_0$. In many examples satisfying $(B0)$-$(B5)$, one has $a_n = n^{1/2}$ and $p^2/n \to 0$ is assumed, so that one has $\hat{\beta}_n = \beta_0 + O_P(n^{-1/2}p^{1/2})$. In many cases $p^2/n \to 0$ can be improved, even upto $p/n \to 0$. Part (c) of the above theorem achieves this under some additional conditions. The assumption $(B5)$ ensures that $\sum_i X_{ni} \stackrel{\mathcal{D}}{\Rightarrow} N(0,1)$ by Theorem 3.2 of Hall and Heyde (1980).

Corollary 2.3 *Assume conditions of Theorem 2.5* (*b*), *and* $(B5)$. *Then for any* $c \in I\!R^p$ *with* $||c|| = 1$, *we have*

$$s_n^{-1} c^T (\hat{\beta}_n - \beta_0) \stackrel{\mathcal{D}}{\Rightarrow} N(0,1) \tag{2.16}$$

The next theorem is an analog of Theorem 2.2.

Theorem 2.6 *Assume the conditions* $(B0)$-$(B3)$ *and the bootstrap weights satisfy conditions* (5.1) - (5.3) . *Then there exists a sequence* $\{\hat{\beta_B}\}$ *of solutions of* (1.2) *such that*

(*a*) *if* $p/a_n \to 0$, *then for fixed* $\epsilon, \delta > 0$, $\exists K > 0$ *and integer* n_0 *such that for all* $n \geq n_0$

$$Prob[\mathrm{P_B}({\sigma_n}^{-1} a_n p^{-1/2} |\hat{\beta_B} - \hat{\beta}_n| \leq K) < 1 - \epsilon] < \delta \tag{2.17}$$

(*b*) *if* $p/a_n \to 0$, *then*

$$\sigma_n{}^{-1} a_n^{-1} p^{-1/2} (\sum_{i=1}^{n} E\phi_{1ni})^T (\hat{\beta_B} - \hat{\beta}_n) = -a_n^{-1} p^{-1/2} \sum_{i=1}^{n} W_i \phi_{ni}(\hat{\beta}_n) + r_{nB1} \tag{2.18}$$

(2.19)

where $||r_{nB1}|| = o_{PB}(p^{-1})$.

(*c*) *if* $p^3 a_n^{-2} \to 0$, *then for any* $c \in I\!R^p$ *with* $||c|| = 1$,

$$s_n^{-1} \sigma_n{}^{-1} c^T (\hat{\beta_B} - \hat{\beta}_n) \;=\; \sum_{i=1}^{n} W_i X_{ni} + o_{PB}(1) \tag{2.20}$$

(*d*) *if* $pa_n^{-2} \to 0$ *and* (*B*4) *holds, then*

$$\sigma_n{}^{-1} a_n^{-1} p^{-1/2} (\sum_{i=1}^{n} \phi_{1ni})^T (\hat{\beta_B} - \hat{\beta}_n) = -a_n^{-1} p^{-1/2} \sum_{i=1}^{n} W_i \phi_{ni}(\hat{\beta}_n) + r_{nB1} \tag{2.21}$$

where $||r_{nB1}|| == O_{PB}(\sigma_n a_n^{-1} p)$.

Let $F_n(x) = Prob[s_n^{-1} c^T (\hat{\beta}_n - \beta_0) \le x]$. and let $F_{Bn}(x) = \mathrm{P_B}[s_n^{-1} \sigma_n{}^{-1} c^T (\hat{\beta_{\mathrm{B}}} - \hat{\beta}_n) \le x]$.

Corollary 2.4 (Proof omitted) *Let* $\{w_{ni}, i = 1, \ldots n\}$ *be exchangeable for every* n. *Assume the conditions of Theorem 2.6, part* (*c*). *Also assume* (*B*5) *and conditions* (5.4)-(5.5) *on the bootstrap weights. Then*

$$\sup_x |F_{Bn}(x) - F_n(x)| \to 0 \textit{ in probability} \tag{2.22}$$

3 Examples of UBS resampling techniques

In this section we discuss some special cases of UBS resampling techniques. In each such case, we identify a random mechanism, and show how the UBS conditions (5.1)-(5.15) relate to it. Also, some of our results assume that the weights are exchangeable or independent. In all the examples we consider, weights are exchangeable, and cases of independence will be particularly mentioned.

Example 3.1 THE CLASSICAL BOOTSTRAP AND ITS VARIATIONS

(*a*) Consider the vector of bootstrap weights $\mathbf{w}_n = (w_{n:1}, \ldots, w_{n:n})$ to be a random sample from $Multinomial(n, 1/n, \ldots, 1/n)$. These weights can be interpreted as simple random sampling with replacement of the functionals to minimise, thus this is essentially the classical bootstrap of Efron (1979). All the conditions except for (5.10) and (5.13) holds in this case. Thus these weights satisfy the conditions required for asymptotic representation of the bootstrap variance estimator and consistency of the bootstrap distribution estimator.

We now consider some variations of (a). The variations (c)-(e) have been mentioned in Praestgaard and Wellner (1993).

(b) Consider the m out of n bootstrap, where $m \to \infty$ $m/n \to 0$. This is carried out by selecting m data points randomly out of the n available. If the selection is without replacement, this scheme can be identified with the delete-$(n-m)$ jackknife. If the selection is with replacement, the weights are a random sample from $Multinomial(m, 1/n, \ldots 1/n)$. Both are special cases of UBS, and after appropriate scaling of the weights, (5.1)-(5.5) are satisfied. This establishes first order accuracy of the distribution estimator. Consistency of the variance estimator can also be obtained.

This bootstrap works in many problems where the "n out of n bootstrap" may not work. For example, as in the case of degenerate U-statistics, where the limiting distribution is non-normal. This suggests that a sub-class of UBS techniques is possibly applicable to a much broader framework, where the assumptions leading to asymptotic normality of the estimator is dropped. This requires further investigations.

(c) For double bootstrap or nested bootstrap the weights $(w_{n1}, \ldots, w_{nn})$ follow $Multinomial(n, M_{n1}/n, \ldots, M_{nn}/n)$ conditional on $M_n = (M_{n1}, \ldots, M_{nn})$, which in turn follows $Multinomial(n, 1/n, \ldots, 1/n)$. The behaviour of this bootstrap is similar to the classical bootstrap described in Example 3.1(a), but this is more likely to concentrate on fewer data points than the classical bootstrap.

(d) The Polya-Eggenberger bootstrap is similar to the double bootstrap, except that none of the conditional probabilities for the second stage sampling is zero. In this case $(w_{n1}, \ldots, w_{nn})$ follow the $Multinomial(n, D_{n1}/n, \ldots, D_{nn}/n)$ conditional on $D_n = (D_{n1}, \ldots, D_{nn})$, which follows $Dirichlet_n(\alpha, \ldots, \alpha)$ for some parameter $\alpha > 0$. The behaviour of this bootstrap is similar to the classical bootstrap and the double bootstrap.

(e) In multivariate hypergeometric bootstrap, the weights satisfy,

$$Prob[w_{n1} = k_1, \ldots, w_{nn} = k_n] = \binom{nK}{n}^{-1} \binom{K}{k_1} \cdots \binom{K}{k_n}; \quad \sum_{i=1}^{n} k_i = n, \ 0 \le k_i \le K$$

where K is some integer parameter. The behaviour of this bootstrap is similar to the classical bootstrap, or the double bootstrap or Polya-Eggenberger bootstrap.

Example 3.2 THE BAYESIAN BOOTSTRAP Consider $\mathbf{w}_n$ to be a random sample from $Dirichlet(\alpha, \ldots, \alpha)$ distribution. All the conditions on weights except for (5.10) and (5.14) are satisfied by these weights. This shows first order accuracy of distribution estimate and consistency of variance estimate for the 'Bayesian bootstrap'. If $\alpha = 4$, then (5.14) is also satisfied. Dirichlet weights are exchangeable but not independent, however the i^{th} such weight can be expressed as $nX_i/(\sum X_i)$ for some *i.i.d.* sequence X_i. This representation is often found useful in studying the second order accuracy of Dirichlet bootstrap. We discuss this further in Example 3.3.

Example 3.3 THE BOOTSTRAP CLONE METHOD Consider the weights $w_{n:i}$ to be *i.i.d.* random variables, satisfying moment and support restrictions imposed by (5.1)-(5.15). Absolute continuity and restrictions on the density of the weights may also be imposed. Such weights may be used to get first and second order accurate distribution estimates and variance estimate. Very often as a variation of this, generalised bootstrap is done by considering weights of the form $w_i = nX_i/(\sum X_i)$, for *i.i.d.* absolutely continuous random variables X_i. This is the idea behind the 'bootstrap clone' method of Lo (1991). A recent application of this method is given in James (1997). In (1.2), one may replace such w_i by X_i without loss. Thus second order accurate distribution estimates and variance estimates may be obtained for all such weights. This covers the case of Dirichlet weights also, discussed in Example 3.2.

Example 3.4 THE WEIGHTED LIKELIHOOD BOOTSTRAP In case $\phi_{ni}(\cdot)$ has a log-likelihood interpretation, the weighted likelihood bootstrap of Newton and Raftery (1994) is a special case of UBS. This can be used for approximate Bayesian inference as suggested by Newton and Raftery (1994). A completely different treatment for bootstrapping likelihoods has been pursued by Davison, Hinkley and Worton (1992).

Example 3.5 THE DELETE-d JACKKNIVES For an integer d in $\{1, \ldots, n\}$, consider the n dimensional vectors $\eta_{n:i_1,i_2,\ldots,i_d}$ where the j^{th} coordinate of $\eta_{n:i_1,\ldots,i_d}$ is 0 if j is one of $i_1, \ldots, i_d$, else it is $n/(n-d)$. Thus there are $\binom{n}{d}$ such vectors. Consider w_n to be a sample from the set of η, each of the $\binom{n}{d}$ vectors having the same probability of being selected. Such weights can be identified with the delete-d jackknife weights (see Chatterjee (1998) for more details). It can be seen that (5.1), (5.3) always holds. If $n - d \to \infty$ then (5.2) also holds, and (5.6) holds for $m_0 = n - d$.

If $d/n \to c \in (0,1)$, then (5.4)-(5.5), (5.7)-(5.9) and (5.15) also hold. If $d/n \to 0$, then (5.10)-(5.12) hold. Thus if $n - d \to \infty$, the basic representation of the jackknife estimator holds. If further, $d/n \to c \in (0,1)$, then first order consistency of the delete-d jackknife distribution estimator is established. However, if $d/n \to 0$, (5.5) is violated, and the jackknife histogram is no longer consistent distribution estimator. However, for variance estimation, the jackknives are consistent. A higher order accuracy of the jackknife distribution may not be expected in general, since a crucial skewness condition (5.14) fails to hold.

The standard interpretation associated with the jackknife, that of 'deleting d observations' and considering the rest does not make sense in time series, since that way the crucial temporal dependence among the observations is lost. Earlier, Kunsch (1989) suggested a jackknife based on block size of the block bootstrap. Our suggestion is different, we 'jackknife' the estimating equations.

Example 3.6 "DOWNWEIGHT-d JACKKNIVES" This is a new variation of the more celebrated delete-d jackknives discussed in (3.5). The modification is as follows: For an integer d in $\{1, \ldots, n\}$, consider the n dimensional vectors $\eta_{n:i_1,i_2,\ldots,i_d}$ where the j^{th} coordinate of $\eta_{n:i_1,\ldots,i_d}$ is d/n if j is one of $i_1, \ldots, i_d$, else it is $(n+d)/n$. The resampling weight vectors are a sample from the set of η, where each of the $\binom{n}{d}$ vectors have equal probability of being selected. The asymptotic properties of the weights of this resampling technique is same as that of the delete-d jackknife, however, since no observation is assigned a weight zero, model assumptions like (5.34) are often not needed.

4 Examples of martingale estimating equations

In this section we discuss some examples of estimating equations. The problems of M-estimation, non-linear regression and generalised linear models are discussed in some details.

Resampling for M-estimators was profoundly studied by Lahiri (1992), and by Mammen (1989, 1993). The latter also incorporated dimension asymptotics aspect in his study. In non-linear regression, resampling was studied in Huet and Jolivet (1989), Huet, Jolivet and Messean (1990) and Gruet, Huet and Jolivet (1993). In generalised linear models resampling was studied by Simonoff and Tsai (1988), Moulton and Zeger (1989, 1991), Sauermann (1989) and Lee (1990). Shao (1992a,b,c) studied the properties of jackknife variance estimators in non-linear regression and generalised linear models. However, almost entirely the literature has concentrated on residual based resampling. As has already been pointed out, the UBS may be seen as a generalisation of the paired bootstrap and delete-d jackknives.

Example 4.1 M-ESTIMATION IN LINEAR REGRESSION Let $y_i = x_i\beta_0 + e_i$ be the linear model and β_0 be the parameter to be estimated. Here $\phi_{ni}(y_i, x_i, \lambda) = -x_i\psi(y_i - \lambda x_i)$ for some function $\psi(\cdot)$. For simplicity assume that the covariate observations x_i are non-random, and the errors e_i are independent. The formulation with random covariates is similar, with some additional complexity. Suppose that the function $\psi(\cdot)$ is twice continuously differentiable, with the first and second derivatives denoted by $\psi_1(\cdot)$, $\psi_2(\cdot)$ respectively. In the following assumptions, k_i and K_i are used as notations for different constants.

($M0$) Assume that for all i, $0 < k_1 < |x_i| < K_1 < \infty$. This assumption was made in Liu and Singh (1992), for example.

($M1$) For all i, $E\psi(e_i) = 0$.

($M2$) For all i, $0 < k_2 < E\psi^2(e_i) < K_2 < \infty$, and $E\psi^4(e_i) < \infty$.

($M3$) For all i, $\psi_1(e_i) > 0$, $0 < k_3 < E\psi_1(e_i)$ and $0 < E\psi_1^2(e_i) < K_3 < \infty$.

($M4$) For some $\delta > 0$, $\exists$ a function $M(\cdot)$ such that for all i, $\sup_{|t|<\delta} |\psi_2(e_i + t)| < M(e_i)$, and $\sup_i EM^2(e_i) < \infty$.

($M5$) $\{x_i\psi(e_i)\}$ satisfy the Lindeberg condition.

The above assumptions ensure that assumptions $(A0)$-$(A5)$ hold. Note that conditions $(M0)$-$(M5)$ are only sufficient conditions. For some robust regression techniques $\psi_2(\cdot)$ is assumed to be bounded, and this ensures $(M4)$.

Example 4.2 LEAST SQUARES IN NON-LINEAR REGRESSION AND GLM In non-linear regression we have the model $y_i = f(\mathbf{x}_i, \beta_0) + e_i$ for some known function $f(\cdot)$ and in generalised least squares, we have $\mu(y_i) = \mathbf{x}_i\beta_0 + e_i$ for some link function $\mu(\cdot)$. The least squares in these two models are similar Here we have $\phi_{ni}(y_i, \mathbf{x}_i, \lambda) = -f_1(\mathbf{x}_i, \lambda)(y_i - f(\mathbf{x}_i, \lambda))$ for some function f whose k^{th} partial derivative with respect to λ is written as f_k.

We again assume non-random covariates $\mathbf{x}_i$ and independent errors e_i. We assume that for all $\mathbf{x}_i$, the functions $f_1(\mathbf{x}_i, \lambda)$, $f_2(\mathbf{x}_i, \lambda)$, $f_3(\mathbf{x}_i, \lambda)$ exist. In the following assumptions, c_i and C_i are used as notations for different constants.

$(NG0)$ Assume that for all i, $0 < c_4 < |x_i| < C_4 < \infty$.

$(NG1)$ For all i, $Ee_i = 0$.

$(NG2)$ For all i, $0 < c_2 < Ee_i^2$ and $Ee_i^4 < \infty$.

$(NG3)$ Assume that for all i, the functions $f_k(\mathbf{x}_i, \lambda)$, $k = 1, 2, 3$ are bounded.

$(NG4)$ For all i, $0 < c_1 < |f_1(\mathbf{x}_i, \lambda)| < C_1 < \infty$.

$(NG5)$ For all i, the functions $g(\mathbf{x}_i, \lambda)$ is continuous at β_0 for $g = f, f_1, f_2, f_3$.

$(NG6)$ Assume that $\{f_1(\mathbf{x}_i, \beta_0)e_i\}$ satisfies the Lindeberg condition.

Conditions $(NG0)$-$(NG6)$, ensure that assumptions $(A0)$-$(A5)$ hold. Conditions $(NG0)$-$(NG5)$ are satisfied in commonly used generalised linear models, see McCullagh and Nelder (1989). The error conditions $(NG1)$ and $(NG2)$, and $(NG6)$ are usual assumptions. Note that all the usual link functions are injective functions, and in our frame-work, the function $f(\mathbf{x}_i, \lambda)$ is actually the inverse of a link function. For the link functions described in McCullagh and Nelder (1989, pages 30-32), it is a small matter to check that conditions $(NG3)$, $(NG4)$, $(NG6)$ hold, if the covariate variable x satisfies $(NG0)$.

Non-linear regression models with martingale difference errors have been considered (Lai (1994)). Conditions of a different flavour appear in Wu (1981). The framework of both these papers overlap with ours.

Example 4.3 MODELS WITH NUISANCE PARAMETERS The following problem has been been of considerable interest (Neyman and Scott (1948)): Suppose X_i, $i = 1, \ldots, n$ are independent with mean μ and variance ${\tau_i}^2$, where μ is the parameter of interest. For ordinary least squares, $\phi_{ni} = -(X_i - \mu)$. The different sufficient conditions of our set-up can be satisfied in this example by assuming $\sup_n EX_n^{16}$ is bounded. This ensures first and second order accuracy of UBS distribution estimator, and the UBS variance representation result. A more careful analysis specialised to this problem shows that actually a much lower moment is sufficient.

Example 4.4 AUTOREGRESSIVE PROCESSES Suppose we have a stationary process $\{X_n\}$ satisfying $X_n = \theta_1 X_{n-1} + \ldots + \theta_p X_{n-p} + e_n$, where $\{e_n\}$ are *i.i.d.* and the polynomial $1 - \theta_1 z \ldots - \theta_p z^p$ has its roots outside the unit disc over the complex plane. The parameter of interest is $\beta_0 = (\theta_1, \ldots, \theta_p)$, whose estimate may be obtained by various techniques, see Brockwell and Davis (1986) for details. Assume $p = 1$ and least squares estimation technique for simplicity. Thus we have $X_t = \theta X_{t-1} + e_t$ where $|\theta| < 1$ and $\{e_n\}$ are *i.i.d.*. Here we have $\phi_{ni} = -X_{i-1}e_i$. If fourth moment of e_1 is bounded, other conditions may also be easily checked. One can use general M-estimation also in time series models. Recently Allen and Datta (1999) have used modified forms of estimated innovations for bootstrapping, which is comparable to the bootstrap of M-estimators in linear regression as in Lahiri (1992). Our technique of resampling is perhaps a simpler approach and does not use estimated residuals.

Example 4.5 CO-INTEGRATED PROCESSES This example is on multivariate time-series data. A univariate time series $\{Y_t\}$ is integrated d times (I_d) if differencing it d times results in a stationary, invertible $ARMA$ process. A multivariate time series process $\{X_t = (X_{1t}, \ldots, X_{(p+1)t})\}$ is co-integrated of order (d, b) if (i) $\{X_{kt}\}$ is I_d and (ii) there exists a non-zero vector a such that $a^T X_t$ is $I_{(d-b)}$ for $b > 0$. The vector a is called the co-integrating vector and is the parameter of interest. For a discussion on these models see Engle and Granger (1987). Co-integrated time series have applications in many economic problems. An important feature here is the study of equilibrium, that is, when different economic variables behave in a way that preserves some economic balance between them, which translated in mathematical terms, mean that the variables are I_d for some $d > 0$, and $d = b$, thus the co-integrated process is stationary.

Consider a simple example of a co-integrated process: where we have $X_{2t} = \beta_0 X_{1t} + Y_t$, where X_{1t} is $I(1)$ and Y_t is $AR(1)$. Suppose ordinary least squares on (X_{1t}, X_{2t}) is used and the parameter estimator is $\hat{\beta}_n = (\sum X_{1t}^2)^{-1} \sum X_{1t} X_{2t}$. In this example we have $a_n = n$, thus the usual $n^{1/2}$-convergence phenomenon is violated. Let $Y_t = \theta e_{t-1} + e_t$ and $X_{1t} = X_{1(t-1)} + u_t$, and assume that $\{u_t\}$ and $\{e_t\}$ are independent, each with mean zero and finite non-zero variance. Our model conditions $(A0)$-$(A4)$ are satisfied by assuming bounded fourth moment of X_{1t} and Y_t.

The asymptotics for estimating the co-integrating vector and its bootstrap may be derived from our set-up.

Example 4.6 ROBUST ESTIMATORS Robust estimators of location, scale and regression are typically M-estimators like the ones considered in Example 4.1, and may be similarly treated.

In all these examples, the assumptions $(A0)$-$(A1)$ on the function ϕ_{ni} and its first derivative is standard, and the second derivative is often bounded, hence assumption

($A2$) is satisfied. The moment assumptions in ($A3$) and the asymptotic normality assumptions of ($A4$) are also routinely satisfied.

Our set-up covers such inference based on likelihoods, where ϕ_{ni} has the interpretation of being derivative of (weighted or unweighted) log-likelihood terms. Note that the martingale structure is particularly important in such situations.

REMARK: SEMIPARAMETRIC INFERENCE PROBLEMS In semiparametric inference problems, along with the finite dimensional parameter an infinite dimensional nuisance parameter is present. One standard approach is to use a kernel based plug-in for the infinite dimensional parameter, and then use *least squares technique* for the finite dimensional parameter. Since estimation of the infinite dimensional component involves the entire data, the crucial martingale property of the estimating equations for the finite dimensional parameter estimator is lost. However, a slight generalisation of the set-up of this paper can be used for such cases, as explained below. We restrict our discussion to the problem of asymptotic normality and consistency of UBS. Higher order results and variance representation will require more careful analysis.

Assume that $\phi_{ni} = m_{ni} + r_{ni}$ where $\{(S_{nj} = \sum_{i=1}^{j} m_{ni}, \mathcal{F}_{nj}),\ j = 1, \ldots, n\}$ is a martingale for some filtration $\mathcal{F}_{nj}$, and r_{ni} is an 'error' term resulting from using an estimator $\hat{m}(\cdot)$ of the non-parametric component $m(\cdot)$ of the model. Let $R_{nn_1} = \sum_{i=n-n_1+1}^{n} r_{ni}$. The solution to $\sum\limits_{i=n-n_1+1}^{n} \phi_{ni} = 0$ and $\sum\limits_{i=n-n_1+1}^{n} m_{ni} = 0$ would be close if R_{nn_1} is small.

The theory developed in this paper can be generalised to cover this case also. By a careful reading of the proofs it can be seen that our asymptotic representation theorems only use the property $O((\sum_{i=n-n_1+1}^{n} \phi_{ni})^{2k}) = O(\sum_{i=n-n_1+1}^{n} \phi_{ni}^{2k})$ and this is available in this general set-up also.

The technical conditions of Section 5 are required to hold for m_{ni} and its derivatives only. Additional requirements are that $R_{nn_1} = o_P(n_1 a_{n_1}^{-1})$ and is $\mathcal{F}_{n,n}$ measurable. This is also easily satisfied; typically by letting $n_1 \to \infty$ and $n_1/n \to 0$.

The bootstrap estimator is obtained by solving $\sum\limits_{i=n-n_1+1}^{n} w_{ni}\phi_{ni} = 0$ for the same integer n_1. The asymptotic normality of the estimator and consistency of the UBS are also easily established using the martingale difference structure of $\{m_{ni}\}$ and asymptotic negligibility of R_{nn_1}. We mention four examples below in this more general framework. A detailed discussion of UBS used in these examples may be found in Chatterjee (1999).

Example 4.7 LOW DIMENSIONAL COMPONENT ESTIMATION IN SEMIPARAMETRIC ADDITIVE REGRESSION MODEL Consider the regression problem: $y_i = f(\mathbf{x}_i, \beta_0) + m(Z_i) + e_i$, where $f(\cdot)$ is a known function, $m(\cdot)$ is an unknown function, β_0 is finite dimensional parameter, e_i's are independent. Sometimes $f(\cdot)$ is linear, resulting in the model being a partial linear model. These models have generated interest owing to their wide applicability. See Fan, Hardle and Mammen (1998) for some results on these. The partial linear model has been considered in Shi (1998) and in much greater generality in Mammen and Van de Geer (1997).

Example 4.8 ESTIMATION IN PARAMETRICALLY LINKED ADDITIVE MODELS Consider the set-up $y_i = \mu + \sum_{j=1}^{J} m_j(X_j) + e_i$, where m_j's are unknown functions. For identifiability, set $Em_j(X_j) = 0$, $j = 1, \ldots, J$. This is the set-up of the additive non-parametric regression, studied in Hastie and Tibshirani (1990), Fan, Hardle, Mammen (1998) and may others. Motivated from non-parametric GARCH model, Carroll, Hardle and Mammen (1999) consider $m_j(x) = \beta_0^{j-1} m_1(x)$, and simultaneous estimation of $m_1(\cdot)$ and β_0. Carroll, Hardle and Mammen (1999) discuss examples from financial data and autoregression, and also cite several references for other applications.

Example 4.9 EXPONENTIAL FAMILY DENSITY ESTIMATES Suppose $y_1, \ldots, y_n$ are *i.i.d.* observations from an absolutely continuous distribution with density $g(x) = g_0(x)exp(b_0 + t(x)b_1)$. Here $g_0(x)$ is a *carrier density*, and the exponential component ensures that the density belongs to exponential family with sufficient statistic $t(x)$. The parameter b_0 is a normalising factor, whereas b_1 is the finite dimensional parameter of the exponential family. The problem of density estimation in this family have been considered in Efron and Tibshirani (1996). This semiparametric model is remarkably versatile in applicability and mathematical tractability. More classical density estimation procedures often require post-repair of the mean, variance etc. (see Jones (1993)), which are in-built in the present model. The carrier $g_0(x)$ is usually uniform density, but other densities can also be used. This is effective in capturing the local properties of the data, whereas the exponential term captures the global features.

Example 4.10 REGRESSION WITH CENSORED OBSERVATIONS This example is from typical regression problems that arise in the context of survival analysis. We consider only one of a large variety of techniques, namely, the proportional hazards model. The data generally consists of the triplet $(t_i, \delta_i, \mathbf{x}_i)$, where t_i is an observed time point, δ_i is the indicator on whether the i^{th} observation is uncensored, and $\mathbf{x}_i$ is the observation on the covariate. The proportional hazards model is given by the hazard function $\lambda(t) = \lambda_0(t) \exp(\mathbf{x}_i^T \beta_0)$ where λ_0 is a baseline hazard function and β_0 is the parameter of interest. Modelling is often done by using a plug-in for λ_0, like an exponential or Weibull hazard function. This is similar to Examples 4.8 or 4.9. The rest of the modelling technique is similar to generalised linear models. Conditions from Example 4.2 can be adapted in this set-up.

5 Appendix 1: Technical conditions

We first state the conditions on UBS weights. and then state the model conditions used for the different results of this chapter. We write $\phi_{ni}(Z_{ni}, \lambda) = \phi_{ni}(\lambda)$, and assume that for all $1 \leq i \leq n$ and all $n \geq 1$, the functions $\phi_{ni}(\lambda)$ are differentiable twice almost everywhere. Two conventions are used for the conditions listed below: Any condition stated for a random function is assumed to hold almost surely unless otherwise stated; and the usage of the term 'for all λ' signifies that the stated condition holds for all values λ in the interior of the parameter space, which is always assumed to be an open set. Also, throughout this paper, k and K, with or without suffix, are used as generic for constants.

5.1 Conditions on UBS weights

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 notations $\mathrm{P_B}, \mathrm{E_B}$ will respectively denote bootstrap probability and expectation, conditional on the data. Let $V(w_i) = \sigma_n^2$ and assume that the quantities

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

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

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

Note that if the weights are assumed to be exchangeable, then this is true.

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 may be related to the model assumptions, and will be specified later. We now state some conditions on the weights. Various combinations of these conditions are used in proving our different results on UBS bootstrap throughout this chapter. The conditions are stated keeping in view that later we allow $p \to \infty$ in some results.

$$Ew_i = 1 \tag{5.1}$$

$$0 < \sigma_n^2 = o(\min a_n^2 p^{-3}, n) \tag{5.2}$$

$$c_{11} = O(n^{-1}) \tag{5.3}$$

$$c_{22} \to 1 \tag{5.4}$$

$$c_4 < \infty \tag{5.5}$$

$$\mathrm{P_B}[\mathcal{W}] = 1 - O(p^2 n^{-1}) \tag{5.6}$$

$$0 < k < \sigma_n^2 < K < \infty \tag{5.7}$$

$$c_{i_1 i_2 \ldots i_k} = O(n^{-k+1}) \quad \forall \quad i_1, i_2, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 3, \tag{5.8}$$

$$c_{i_1 i_2 \ldots i_k} = O(\min(n^{-k+2}, 1)) \quad \forall \quad i_1, i_2, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 4. \tag{5.9}$$

$$0 \leftarrow \sigma_n^2 \quad < \quad K < \infty \tag{5.10}$$

$$c_{i_1 i_2 \ldots i_k} = O(n^{-k+1}\sigma_n{}^{-1}) \quad \forall \quad i_1, i_2, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 3, \tag{5.11}$$

$$c_{i_1 i_2 \ldots i_k} = O(n^{-k+2}) \quad \forall \quad i_1, i_2, \ldots, i_k \text{ satisfying } \sum_{j=1}^{k} i_j = 4. \tag{5.12}$$

$$\mathrm{P_B}[w_i > k > 0] \quad = \quad 1 \tag{5.13}$$

$$c_3 \quad \to \quad 1 \tag{5.14}$$

$$c_8 \quad < \quad \infty \tag{5.15}$$

5.2 Assumptions for subsection 2.1

We adopt the notations $\phi_{(k+1)ni}(\lambda) = \frac{\partial}{\partial\lambda}\phi_{kni}(\lambda)$, for $k = 0, 1, 2, \ldots$, where $\phi_{0ni}(\lambda) \equiv \phi_{ni}(\lambda)$; $\phi_{kni}(\beta_0) = \phi_{kni}$ for $k = 0, 1, 2 \ldots$. Let $S_{nj} = \sum_{i=1}^{j} \phi_{ni}$.

For the sequence $\{a_n\}$ described later on in (5.17), define

$$s_n^2 \quad = \quad \sum_{i=1}^{n} E\phi_{ni}^2 / a_n^2$$

$$\gamma_n \quad = \quad [\sum_{i=1}^{n} \phi_{ni}^2]^{-1/2} \sum_{i=1}^{n} \phi_{1ni}.$$

Assume that for each $n \geq 1$, there is a sequence of σ-fields $\mathcal{F}_{n1} \subset \ldots \mathcal{F}_{nn}$, such that $\{S_{nj}, \mathcal{F}_{nj},\ j = 1, \ldots, n\}$ is a martingale sequence. The following are a set of conditions on this martingale:

$(A0)$

$$E\phi_{ni} = 0 \text{ for all } 1 \leq i \leq n, \;\; n \geq 1. \tag{5.16}$$

$$\exists \text{ a sequence } \{a_n \to \infty\} \text{ such that } K_1 > \sum_{i=1}^{n} E\phi_{ni}^2 / a_n^2 \quad > \quad k_1 > 0 \tag{5.17}$$

$(A1)$

$$0 < k_2 \quad < \quad \sum_{i=1}^{n} E\phi_{1ni} / a_n^2 < K_2 \tag{5.18}$$

$$0 \quad < \quad \sum_{i=1}^{n} \phi_{1ni}(\lambda) / a_n^2 \text{ for all } \lambda \tag{5.19}$$

$(A2)$ There exist M_{2ni}, possibly random and depending on β_0, and a sequence $\{c_n\}$ such that

$$\sup_{|\lambda - \beta_0| < \delta_0} |\phi_{2ni}(\lambda)| \quad \leq \quad M_{2ni} \text{ for some fixed } \delta_0 > 0. \tag{5.20}$$

$$E(\sum_{i=1}^{n} M_{2ni})^2 \quad = \quad O(c_n^2) \text{ where } a_n^{-4} c_n^2 = O(1). \tag{5.21}$$

$(A3)$ For some sequences $\{b_n\}$ and $\{d_n\}$,

$$E[a_n^{-1} \sum (\phi_{1ni} - E\phi_{1ni})]^2 = O(b_n^2) \text{ where } a_n^{-2} b_n^2 = o(1). \tag{5.22}$$
$$E[a_n^{-1} \sum (\phi_{ni}^2 - E\phi_{ni}^2)]^2 = O(d_n^2) \text{ where } a_n^{-2} d_n^2 = o(1). \tag{5.23}$$

$(A4)$ (Asymptotic normality assumptions)

The triangular sequence $X_{ni} = (\sum_{i=1}^n E\phi_{ni}^2)^{-1/2} \phi_{ni}$ satisfies

$$\max_i |X_{ni}| \xrightarrow{p} 0, \ \sum_{i=1}^n X_{ni}^2 \xrightarrow{p} 1, \ \sup_n E(\max_i X_{ni}^2) < \infty \tag{5.24}$$
$$\mathcal{F}_{n,i} \subset \mathcal{F}_{n+1,i} \text{ for } i = 1, \dots, n, n \geq 1 \tag{5.25}$$

5.3 Assumptions for subsections 2.2 and 2.3

In both subsections 2.2 and 2.3, we assume that ϕ_{ni}, $i = 1, \dots, n$ are *independent*. Assume that

$$\phi_{ni}(\lambda + t) = \phi_{ni}(\lambda) + \phi_{1ni}(\lambda)t + 2^{-1}\phi_{2ni}(\lambda)t^2 + R_{ni}(t, \lambda)t^2 \tag{5.26}$$

where $|R_{ni}(t, \lambda)| < k|t|^{\alpha}$ for each λ for some $0 < \alpha \leq 1$.

Thus for this restricted set-up we can use simplified expressions for some of the assumptions stated earlier. Following is the list of assumptions in this framework:

$$E\phi_{ni} = 0 \text{ for all } 1 \leq i \leq n, \ n \geq 1. \tag{5.27}$$
$$\sum_{i=1}^n E|\phi_{ni}|^L = O(n) \tag{5.28}$$
$$\sum_{i=1}^n E|\phi_{1ni}|^L = O(n) \tag{5.29}$$
$$\sum_{i=1}^n E|\phi_{2ni}|^L = O(n) \tag{5.30}$$
$$n^{-1} \sum_{i=1}^n \phi_{1ni} > k_1 > 0, \tag{5.31}$$
$$E\phi_{ni}^2 > k_1 > 0, \quad \text{for all } \lambda \tag{5.32}$$
$$\phi_{1ni}(\lambda) > k_2 > 0, \quad \text{for all } \lambda \tag{5.33}$$

In the above, the constant L is to be specified in the results. For some resampling schemes that we consider, condition (5.31) is not sufficient. Suppose m_0 is a specified integer in the range $[n/3]$ to n. For any integer m in $\{m_0, \dots, n\}$ consider the subset $\mathcal{I}_m = \{i_1, i_2 \dots, i_m\}$ of $\{1, 2, \dots, n\}$. We assume

$$m^{-1} \sum_{i \in \mathcal{I}_m} \phi_{1ni} > k_1 > 0 \tag{5.34}$$

for every such choice of subset $\mathcal{I}_m$ of size m from $\{1,2,\ldots,n\}$ and for every m in $[m_0,n]$. The more stringent assumption (5.34) is required to make resampling schemes like the paired bootstrap and the different jackknives feasible.

Note that (5.27)-(5.30) covers the conditions listed in $(A0)$ and $(A2)$-$(A3)$. The conditions (5.34) reflect the condition $(A1)$.

5.4 Assumptions for subsection 2.4

The model conditions here are extensions of the conditions $(A0)$-$(A4)$ for the p dimensional case, along with a condition on the growth rate of p with respect to n. The following notations will be used: $||c||$ is the Euclidean norm of a vector c, A^T is the transpose of the matrix A, $tr(A)$ is the trace of A, $\lambda_{max}(A)$ and $\lambda_{min}(A)$ are respectively the maximum and minimum eigenvalue of A, and for symmetric matrices A and B, $A > B$ means $A - B$ is positive definite.

Note that each ϕ_{ni} is actually a vector valued function, thus

$$\phi_{ni}(\lambda) = (\phi_{ni1}(\lambda),\ \ldots,\ \phi_{nip}(\lambda))^T$$

For each $\phi_{nia}(\lambda)$, we assume that the following Taylor expansion holds:

$$\phi_{nia}(\lambda+t) = \phi_{nia}(\lambda) + \phi^T_{1nia}(\lambda)t + 2^{-1}t^T H_{2nia}(\lambda_1)t \tag{5.35}$$

for $\lambda_1 = \lambda + ct$ for some $0 < c < 1$. We consider two separate possibilities: the dimension p is a fixed positive integer; or $p \to \infty$ as $n \to \infty$. As earlier, we often denote $\phi_{ni}(\beta_0)$, $\phi_{1nia}(\beta_0)$ etc. by ϕ_{ni}, ϕ_{1nia} etc.

Let $S_{nj} = \sum_{i=1}^{j} \phi_{ni}$. Assume that for every n, there is a sequence of σ-fields $\mathcal{F}_{n1} \subset \ldots \mathcal{F}_{nn}$, such that $\{S_{nj}, \mathcal{F}_{nj},\ j = 1,\ldots,n\}$ is a martingale sequence.

$(B0)$ Assume that

$$E\phi_{ni} = 0 \tag{5.36}$$

$$\exists \text{ a sequence } \{a_n \to \infty\} \text{ such that } K > [\sup_{||c||=1} \sum_{i=1}^{n} E(c^T\phi_{ni})^2]/a_n^2 > k > 0 \tag{5.37}$$

$(B1)$ Assume that

$$\sum_{i=1}^{n}\sum_{a=1}^{p} E||\phi_{1nia}||^2 = O(a_n^2p^2) \tag{5.38}$$

$$\sum_{i=1}^{n}\sum_{a=1}^{p} E||\phi_{1nia} - E\phi_{1nia}||^2 = O(a_n^2p^2) \tag{5.39}$$

$(B2)$ For the symmetric matrix H_{2nia} in (5.35), for some $\delta_0 > 0$ there exists a symmetric matrix M_{2nia} such that

$$\text{for all } \{t : ||t|| \leq \delta_0\} \quad H_{2nia}(\beta_0 + t) < M_{2nia} \tag{5.40}$$

$$\sum_{i=1}^{n}\sum_{a=1}^{p} E\lambda^2_{max}(M_{2nia}) = O(a_n^4pn^{-1}) \tag{5.41}$$

($B3$) Let $\phi_{1ni}(\lambda)$ be the $(p \times p)$ matrix, whose a^{th} row is given by $\phi^T_{1nia}(\lambda)$, for $a = 1, \ldots, p$. Let $\Gamma_{1n}(\lambda) = a_n^{-2} \sum_{i=1}^n \phi_{1ni}(\lambda)$. Assume that Γ_{1n} is nonsingular. Let $G_{1n} = a_n^{-2} \sum_{i=1}^n E\phi_{1ni}$. Assume

$$0 < k_2 < \lambda_{min}(2^{-1}(G_{1n} + G_{1n}^T)) \tag{5.42}$$

$$\sup_{||c||=1} ||G_{1n}^{-1} c|| = O(p) \tag{5.43}$$

($B4$) (strengthening of ($B3$)) Recall that $\Gamma_{1n}(\lambda) = a_n^{-2} \sum_{i=1}^n \phi_{1ni}(\lambda)$. Consider all $\lambda \in \{\lambda : ||\lambda - \beta_0|| < \delta_0\}$, where without loss of generality δ_0 is same as in ($B2$). Assume

$$0 < k_2 < \lambda_{min}(2^{-1}(\Gamma_{1n}(\lambda) + \Gamma_{1n}^T(\lambda)) \tag{5.44}$$

$$\sup_{||c||=1} ||\Gamma_{1n}^{-1}(\lambda) c|| = O(p) \tag{5.45}$$

($B5$) Consider any fixed vector $c \in I\!R^p$ with $||c|| = 1$. Let

$$s_n^2 = p^{-2}[(\sum_{i=1}^n E\phi_{1ni})^{-1}c]^T [\sum_{i=1}^n E\phi_{ni}\phi_{ni}^T][(\sum_{i=1}^n E\phi_{1ni})^{-1}c]$$

$$X_{ni} = -s_n^{-1}[(\sum_{i=1}^n E\phi_{1ni})^{-1}c]^T \phi_{ni}$$

Then X_{ni} satisfies

$$\max_i |X_{ni}| \xrightarrow{p} 0, \quad \sum_{i=1}^n X_{ni}^2 \xrightarrow{p} 1, \quad \sup_n E(\max_i X_{ni}^2) < \infty \tag{5.46}$$

$$\mathcal{F}_{n,i} \subset \mathcal{F}_{n+1,i} \text{ for } i = 1, \ldots, n, n \geq 1 \tag{5.47}$$

6 Appendix 2: Proofs of results

Our first result in this section is on Edgeworth expansions for studentised sample sum of independent, not necessarily identically distributed random variables with zero mean. The proof of Theorem 2.4 easily follows from this, and hence it is omitted.

Consider random variables $\{X_i\}$ satisfying the following moment conditions:

$$EX_i = 0 \text{ for all } i \tag{6.1}$$

$$0 < EX_i^2 = \tau_i^2 < K_1 < \infty \tag{6.2}$$

$$\sum EX_i^4 = O(n) \tag{6.3}$$

$$\sum EX_i^6 = o(n^2) \tag{6.4}$$

Under the above set of assumptions, we consider Edgeworth expansion for the statistic $t_n = (\sum X_i^2)^{-1/2} \sum X_i$. The standard notations $\Phi(\cdot)$ and $\phi(\cdot)$ are used for the standard normal distribution function and density function respectively.

Theorem 6.1 *For random variables $\{X_n\}$ satisfying Assumption A and* (6.1)-(6.4),

$$\sup_{y \in \mathbb{R}} |P[(\sum X_i^2)^{-1/2} \sum X_i \leq y] - \Phi(y) - n^{-1/2} p(y) \phi(y)| = o(n^{-1/2}) \tag{6.5}$$

for some polynomial $p(y)$ of degree 3 whose coefficients are continuous functions of the averages of the first three moments of X_i, $i = 1, \ldots, n$.

REMARK 1. Hall (1987) established Edgeworth expansion for the studentised sample mean of *i.i.d.* random variables with finite third moments and non-singular distributions. The non-singularity of the distribution was used to establish a Cramer condition type result (see Lemma 2.2 of Hall (1987)), using a result on the density of the absolutely continuous part of the random variable (see Lemma 2.1 of Hall (1987)). We have assumed for convenience that the random variables are all absolutely continuous. The proof of Theorem 6.1 is along the same lines as that of Hall (1987).

REMARK 2. This theorem is a partial generalisation of the main results of three important papers. Bai and Zhao (1986) considered nonstudentised sample averages from independent random variables, our result is for the studentised case. Hall (1987) established Edgeworth expansion for studentised sample average of *i.i.d.* random variables under minimal conditions, we extend it to the independent case, but not under minimal conditions. By using standard techniques, our result can be extended to functions of sample averages of independent random variables, thereby extending the result of Bai and Rao (1991). Our result may also be viewed as a variant of Theorem 20.6 of Bhattacharya and Ranga Rao (1976), where individual random variables are required to satisfy the Cramer condition. We derive a 'conditional Cramer condition', as in Lemma 2.2. of Hall (1987) in the present set-up, and for this we need assumption A. However, by following a different approach based on Edgeworth expansions for strongly non-lattice random variables, assumptions about absolute continuity of X_n's can possibly be replaced by strong non-latticeness of X_i's.

For weighted sums of *i.i.d.* random variables, a sufficient condition for ensuring assumption A is given below. The proof is easy and is omitted.

Proposition 6.1 *Suppose $\{\xi_i\}$ are i.i.d. mean zero absolutely continuous random variables with density $f(x)$ such that for some constants $x_0 > 0$ and $b > 0$, $f(x) > b > 0$ for all $x \in (-x_0, x_0)$. Let $\{a_i\}$ be a sequence of non-random constants such that for large n, and for some $\alpha > 0$; at least $[\alpha n]$ of $\{|a_i|, i = 1, \ldots n\}$ lie in $[a^{-1}, a]$ for some $a > 1$. Then $\{X_i = a_i \xi_i\}$ satisfies Assumption A.*

SKETCH OF PROOF OF THEOREM 6.1 Theorem 6.1 is proved by closely imitating the arguments of Hall (1987). Hence we give only a sketch of the proof. In Hall (1987) the X_j's are *i.i.d.* but in our case it is not so. Let $Y_j = X_j - E(X_j||X_j|)$, $p_j = P[X_j > 0||X_j|]$, $\beta_{2j} = E[Y_j^2||X_j|]$ and $\psi_j(t) = E[\exp(itY_j)||X_j|]$, where all the above expectations are conditional on $|X_j|$, as is obvious from the notation. The arguments of Hall (1987) may be followed, by substituting appropriate averages of moments (or conditional moments) in place of the moments (or conditional moments) from the single distribution considered by Hall (1987). Note that the remarks of Hall (1987, page 922, after (2.4)) do not hold in case X_i's are not *i.i.d.*. However, Assumption A allows us to overcome these issues. With a notation similar to that of Hall (1987), we may then define $\mu_2 = n^{-1}\sum EX_i^2$ and $\nu_2 = n^{-1}\sum E\beta_{2j}$ among other things, and follow his arguments. One additional technical condition that is required for the present set-up is that both μ_2 and ν_2 must have a positive lower bound. This follows from Assumption A. We also need a parallel to Lemma 2.2 of Hall (1987), which is stated below. We omit its proof, which is easy.

Lemma 6.1 *If X_i's are absolutely continuous with densities satisfying assumption A, then*

$$\sup_{|t|>\epsilon} n^{-1}\sum E|\psi_j(t)| < 1$$

for all $\epsilon > 0$ for all large n.

Then, by following arguments similar to that of Hall (1987, pages 926-930), we obtain

$$\sup_{y\in I\!R} |P[t_n \le y] - \Phi(y) - n^{-1/2}q(y)\phi(y)| = o(n^{-1/2})$$

for some third degree polynomial $q(y)$ whose coefficients depend on moments upto third order of X_i's and $(2p_i - 1)|X_i|$'s. From Theorem 20.6 of Bhattacharya and Ranga Rao (1976) it follows that $q(y)$ must be as stated in (6.5), with coefficients depending only on the averages of moments upto third order of the random vectors $(X_i, \quad n^{-1/2}(X_i^2 - {\tau_i}^2))^T, \ i = 1, \ldots, n$, since if two second order Edgeworth expansions exist for the same quantity they may differ only by $o(n^{-1/2})$. However, now we can use the special structure of our statistic to see that the coefficients of $q(y)$ must be functions of $n^{-1}\sum EX_i^2$ and $n^{-1}\sum EX_i^3$, since $EX_i = 0$ for all i. An easy cross check is accomplished by observing that this is indeed the case when X_i's are *i.i.d* random variables. This completes the sketch of the proof of Theorem 6.1. ■

Proof of Theorem 2.1 Note that given any $\epsilon > 0$, $\exists\ K > 0$ and an integer n_0 such that for all $n \ge n_0$

$$Prob[|a_n^{-1}\sum_{i=1}^{n}\phi_{ni}| > K] < \epsilon/3 \tag{6.6}$$

This is easily established using Chebyshev's inequality, $(A0)$ and $(A3)$. Let $\gamma_{1n} = a_n^{-2}\sum_{i=1}^{n}\phi_{1ni}$. Next note that given any constants $\epsilon, C, K > 0$, for all n large enough,

$$Prob[\sup_{|t|=C} |S_n(t)| > K] < \epsilon/3 \quad \text{where} \tag{6.7}$$

$$S_n(t) = a_n^{-1}\sum_{i=1}^{n}[\phi_{ni}(\beta_0 + a_n^{-1}t) - \phi_{ni}(\beta_0)] - \gamma_{1n}t$$

Instead of (6.7), we actually prove the much stronger result

$$E[\sup_{|t|\leq C}|S_n(t)|]^2 = O(a_n^{-2}) \tag{6.8}$$

By Taylor series expansion we have

$$\sum_{i=1}^{n}[\phi_{ni}(\beta_0 + a_n^{-1}t) - \phi_{ni}(\beta_0)] = a_n^{-1}t\sum_{i=1}^{n}\phi_{1ni}(\beta_0) + \frac{1}{2}a_n^{-2}t^2\sum_{i=1}^{n}\phi_{2ni}(\beta_1)$$

where β_1 lies in between β_0 and $\beta_0 + a_n^{-1}t$. Then $S_n(t) = \frac{1}{2}a_n^{-3}t^2\sum_{i=1}^{n}\phi_{2ni}(\beta_1)$. Thus we have for all large n, $\sup_{|t|\leq C}|S_n(t)| \leq 2^{-1}a_n^{-3}C^2\sum_{i=1}^{n}M_{2ni}$ using $(A2)$. Thus

$$\begin{aligned} E[\sup_{|t|\leq C}|S_n(t)|]^2 &\leq 4^{-1}a_n^{-6}C^4E(\sum_{i=1}^{n}M_{2ni})^2 \\ &= O(a_n^{-2}) \text{ using } (A2) \end{aligned}$$

This completes the proof of (6.8). Note that

$$\inf_{|t|=C}\{a_n^{-1}t\sum_{i=1}^{n}\phi_{ni}(\beta_0 + a_n^{-1}t)\} \geq -C\sup_{|t|=C}|S_n(t)| + C^2\gamma_{1n} - Ca_n^{-1}|\sum_{i=1}^{n}\phi_{ni}| \tag{6.9}$$

¿From (6.6), (6.7) and (6.9) we have, choosing C large enough,

$$\begin{aligned} & Prob[\inf_{|t|=C}\{a_n^{-1}t\sum_{i=1}^{n}\phi_{ni}(\beta_0 + a_n^{-1}t)\} > 0] \\ \geq\ & Prob[a_n^{-1}|\sum_{i=1}^{n}\phi_{ni}| + \sup_{|t|=C}|S_n(t)| \leq C\gamma_{1n}] \\ =\ & 1 - Prob[a_n^{-1}|\sum_{i=1}^{n}\phi_{ni}| + \sup_{|t|=C}|S_n(t)| > C\gamma_{1n}] \\ \geq\ & 1 - Prob[a_n^{-1}|\sum_{i=1}^{n}\phi_{ni}| > Ck_2/4] - Prob[\sup_{|t|=C}|S_n(t)| > Ck_2/4] \\ & -Prob[a_n^{-2}|\sum(\phi_{1ni} - E\phi_{1ni})| > k_2/2] \\ \geq\ & 1 - \epsilon \text{ for all } n \text{ sufficiently large} \end{aligned}$$

Using continuity of $\sum_{i=1}^{n}\phi_{ni}(\lambda)$ in λ, this means that for fixed $\epsilon > 0$ for all n sufficiently large $\exists C$ large such that

$$\sum_{i=1}^{n}\phi_{ni}(\beta_0 + a_n^{-1}t) = 0 \text{ has a root } t = T_n \text{ in } |t| \leq C \text{ with probability} > 1 - \epsilon \tag{6.10}$$

Putting $\hat{\beta}_n = \beta_0 + a_n^{-1}T_n$, we get a solution to (1.1) which satisfies, for fixed $\epsilon > 0$, $Prob[a_n|\hat{\beta}_n - \beta_0| \leq C] \geq 1 - \epsilon$ for all n large enough. This shows (2.1). Now notice that with this C fixed, we have actually obtained in (6.8) that with $t = T_n$, $a_n \gamma_{1n}(\hat{\beta}_n - \beta_0) = -a_n^{-1} \sum_{i=1}^n \phi_{ni} + r_{n1}$, where $Er_{n1}^2 = O(a_n^{-2})$. Now a little bit more algebra and conditions $(A1)$ and $(A3)$ yields (2.2). ■

Proof of Theorem 2.2 The techniques used in proving (2.4) and (2.5) are not much different from what was used to prove (2.1) and (2.2). First fix $\epsilon, \delta > 0$. Using Chebyshev's inequality, (5.1 and (5.3) it can be easily shown that $\hat{\gamma}_{1n}^{-1} = O_P(1)$ and hence for some constant k,

$$\mathrm{P_B}[|a_n^{-1} \sum_{i=1}^n w_i \phi_{ni}(\hat{\beta}_n)| > C\hat{\gamma}_{1n}\sigma_n/2] \leq \frac{k\hat{g}_n}{C^2\hat{\gamma}_{1n}^2} \ (= U_{1C} \text{ say }) \tag{6.11}$$

Thus this bootstrap probability is bounded by a random variable depending on C, which we call U_{1C}. By choosing C large enough and using some algebra, it can be shown that $Prob[U_{1C} > \epsilon/2] < \delta/2$. Fix such a C and define

$$S_{nB}(t) = a_n^{-1} \sum_{i=1}^n w_i[\phi_{ni}(\hat{\beta}_n + a_n^{-1}t) - \phi_{ni}(\hat{\beta}_n)] - \hat{\gamma}_{1n}t$$

Using Taylor series expansion, for some $\hat{\beta}_{n1}$ between $\hat{\beta}_n$ and $\hat{\beta}_n + a_n^{-1}t$, we have

$$\begin{aligned} & S_{nB}(t) \\ = \ & a_n^{-2}t \sum_{i=1}^n w_i \phi_{1ni}(\hat{\beta}_n) - \hat{\gamma}_{1n}t + \frac{1}{2}a_n^{-3}t^2 \sum_{i=1}^n w_i \phi_{2ni}(\hat{\beta}_{n1}) \\ = \ & \sigma_n a_n^{-2}t \sum_{i=1}^n W_i \phi_{1ni}(\hat{\beta}_n) + \frac{1}{2}a_n^{-3}t^2 \sum_{i=1}^n w_i \phi_{2ni}(\hat{\beta}_{n1}) \\ = \ & S_{nB1}(t) + S_{nB2}(t) \text{ say} \end{aligned}$$

Consequently,

$$\begin{aligned} & \mathrm{P_B}[\sup_{|t|=C\sigma_n} |S_{nB}(t)| > C\hat{\gamma}_{1n}\sigma_n/2] \\ \leq \ & \mathrm{P_B}[\sup_{|t|=C\sigma_n} |S_{nB1}(t)| > C\hat{\gamma}_{1n}\sigma_n/4] + \mathrm{P_B}[\sup_{|t|=C\sigma_n} |S_{nB2}(t)| > C\hat{\gamma}_{1n}\sigma_n/4] \\ = \ & U_{2C} \text{ say} \end{aligned} \tag{6.12}$$

Now

$$\begin{aligned} & \mathrm{P_B}[\sup_{|t|=C\sigma_n} |S_{nB1}(t)| > C\hat{\gamma}_{1n}\sigma_n/4] \\ = \ & \mathrm{P_B}[\sigma_n^2 a_n^{-2} C |\sum_{i=1}^n W_i \phi_{1ni}(\hat{\beta}_n)| > C\hat{\gamma}_{1n}\sigma_n/4] \\ \leq \ & \frac{k\sigma_n^2 \sum \phi_{1ni}^2(\hat{\beta}_n)}{a_n^4 \hat{\gamma}_{1n}^2} \\ = \ & O_P(\sigma_n^2 a_n^{-2}) \text{ using } (A1), (A2) \text{ and } (A3) \\ = \ & o_P(1) \end{aligned}$$

$$
\begin{aligned}
& \mathrm{P_B}[\sup_{|t|=C\sigma_n} |S_{nB2}(t)| > C\hat{\gamma}_{1n}\sigma_n/4] \\
= \; & \mathrm{P_B}[\sup_{|t|=C\sigma_n} |\frac{1}{2}a_n^{-3}t^2\sum_{i=1}^n w_i\phi_{2ni}(\hat{\beta}_{n1})| > C\hat{\gamma}_{1n}\sigma_n/4] \\
\leq \; & \mathrm{P_B}[\frac{1}{2}a_n^{-3}\sup_{|t|=C\sigma_n} t^2\sum_{i=1}^n w_i|\phi_{2ni}(\hat{\beta}_{n1})| > C\hat{\gamma}_{1n}\sigma_n/4] \text{ since } w_i \geq 0 \text{ for all } i \\
= \; & \mathrm{P_B}[\frac{1}{2}a_n^{-3}\sup_{|t|=C\sigma_n} t^2\sum_{i=1}^n w_i|\phi_{2ni}(\hat{\beta}_{n1})| > C\hat{\gamma}_{1n}\sigma_n/4](I_{\{|\hat{\beta}-\beta_0|<\delta_0/2\}} + I_{\{|\hat{\beta}-\beta_0|\geq\delta_0/2\}}) \\
\leq \; & \mathrm{P_B}[\frac{1}{2}a_n^{-3}\sup_{|t|=C\sigma_n} t^2\sum_{i=1}^n w_i|\phi_{2ni}(\hat{\beta}_{n1})| > C\hat{\gamma}_{1n}\sigma_n/4]I_{\{|\hat{\beta}-\beta_0|<\delta_0/2\}} + I_{\{|\hat{\beta}-\beta_0|\geq\delta_0/2\}} \\
\leq \; & \mathrm{P_B}[\frac{1}{2}a_n^{-3}\sup_{|t|=C\sigma_n} t^2\sum_{i=1}^n w_i|\phi_{2ni}(\hat{\beta}_{n1})| > C\hat{\gamma}_{1n}\sigma_n/4]I_{\{|\hat{\beta}-\beta_0|<\delta_0/2\}} + O_P(a_n^{-1}) \\
\leq \; & \mathrm{P_B}[\frac{1}{2}a_n^{-3}\sup_{|t|=C\sigma_n} t^2\sum_{i=1}^n w_iM_{2ni} > C\hat{\gamma}_{1n}\sigma_n/4] + O_P(a_n^{-1}) \text{ using } (A2) \\
\leq \; & \mathrm{P_B}[\frac{1}{2}a_n^{-3}C^2\sigma_n^2\sum_{i=1}^n w_iM_{2ni} > C\hat{\gamma}_{1n}\sigma_n/4] + O_P(a_n^{-1}) \\
\leq \; & 2a_n^{-3}C\sigma_n\hat{\gamma}_{1n}^{-1}\sum M_{2ni} + O_P(a_n^{-1}) \\
= \; & O_P(\sigma_n a_n^{-3}c_n) + O_P(a_n^{-1}) \text{ using } (A2) \text{ and } (A3) \\
= \; & o_P(1)
\end{aligned}
$$

So for all n large enough, $Prob[U_{2C} > \epsilon/2] < \delta/2$. Also, as in (6.9),

$$
\begin{aligned}
& \inf_{|t|=C\sigma_n} \{a_n^{-1}t\sum_{i=1}^n w_i\phi_{ni}(\hat{\beta}_n + a_n^{-1}t)\} \\
\geq \; & -C\sigma_n \sup_{|t|=C\sigma_n} |S_{nB}(t)| + C^2\sigma_n^2\hat{\gamma}_{1n} - C\sigma_n|a_n^{-1}\sum_{i=1}^n w_i\phi_{ni}(\hat{\beta}_n)| \qquad (6.13)
\end{aligned}
$$

Thus

$$
\begin{aligned}
& \mathrm{P_B}[\inf_{|t|=C\sigma_n} \{a_n^{-1}t\sum_{i=1}^n w_i\phi_{ni}(\hat{\beta}_n + a_n^{-1}t)\} > 0] \\
\geq \; & \mathrm{P_B}[|a_n^{-1}\sum_{i=1}^n w_i\phi_{ni}(\hat{\beta}_n)| + \sup_{|t|=C\sigma_n} |S_{nB}(t)| \leq C\hat{\gamma}_{1n}\sigma_n] \\
= \; & 1 - \mathrm{P_B}[|a_n^{-1}\sum_{i=1}^n w_i\phi_{ni}(\hat{\beta}_n)| + \sup_{|t|=C\sigma_n} |S_{nB}(t)| > C\hat{\gamma}_{1n}\sigma_n] \\
\geq \; & 1 - \mathrm{P_B}[|a_n^{-1}\sum_{i=1}^n w_i\phi_{ni}(\hat{\beta}_n)| > C\hat{\gamma}_{1n}\sigma_n/2] - \mathrm{P_B}[\inf_{|t|=C\sigma_n} |S_{nB}(t)| > C\hat{\gamma}_{1n}\sigma_n/2] \\
= \; & 1 - U_{1C} - U_{2C}
\end{aligned}
$$

This establishes (2.4), since $Prob[U_{iC} > \epsilon/2] < \delta/2,\ i = 1, 2$ from 6.11) and (6.12). By modifying slightly the above argument it is easy to establish that

$$\sup_{|t| \leq C\sigma_n} |S_{nB}(t)| = \sigma_n r_{nB1} \tag{6.14}$$

where $\mathrm{P_B}(a_n |r_{nB1}| > \epsilon) = O_P(1)$ for any $\epsilon > 0$, then it follows that $\hat{\gamma}_{1n}\sigma_n^{-1} a_n(\hat{\beta_{\mathrm{B}}} - \hat{\beta}_n) = -a_n^{-1}\sum_{i=1}^{n} W_i \phi_{ni}(\hat{\beta}_n) + O_{PB}(a_n^{-1})$ With some more algebra, (2.5) follows. ■

In order to prove Corollary 2.2, we need Lemma 4.6 of Praestgaard and Wellner (1993), which is quoted below.

Theorem 6.2 (Praestgaard and Wellner (1993)) *Let $\{n\}$ be a sequence of natural numbers, let $\{a_{nj}\}$ be a triangular array of constants, and let $W_{nj},\ \ j = 1, \ldots, n,\ \ n \in \{n\}$ be a triangular array of row-exchangeable random variables such that*

$$n^{-1}\sum_{j=1}^{n}(a_{nj} - \bar{a}_n)^2 \to \tau^2 > 0, \quad n^{-1}\max_{j=1,\ldots,n}(a_{nj} - \bar{a}_n)^2 \to 0$$

$$n^{-1}\sum_{j=1}^{n}(W_{nj} - \bar{W}_n)^2 \xrightarrow{p} c^2 > 0, \quad \lim_{K\to\infty}\limsup_{n\to\infty} E(W_{nj} - \bar{W}_n)^2 I_{\{|W_{nj} - \bar{W}_n| > K\}} = 0$$

Then

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(a_{mj}W_{nj} - \bar{a}_n\bar{W}_n) \Rightarrow N(0, c^2\tau^2) \tag{6.15}$$

Proof of Corollary 2.2 It can be checked that with $a_{mj} = -n^{1/2}a_n^{-1}s_n{-}1\phi_{ni}$, the conditions of Theorem 6.2 is satisfied with $\tau^2 = 1,\ c^2 = 1$. Using Lemma 3.1 and 4.7 of Praestgaard and Wellner (1993), the conditions of Theorem 6.2 are verified.

In place of Theorem 6.2, slight variation of the model conditions and assumptions on resampling weights allows us to use Theorem 3.3 of Arenal-Gutierrez and Matran (1996). In some circumstances this is a more general result. ■

Proof of Theorem 2.3 The details of the algebra involved in this proof is similar to those of Theorems 2.1 and 2.2, and we omit many of the details.

Let us concentrate on the set $\mathcal{A} \cap \mathcal{W}$ only, since the contribution from the complement of this set is negligible. Under the conditions of the theorem, we have that

$$ng_{1n}^2 \mathrm{V_n} = n^{-1}\sum E\phi^2 + O(n^{-1})$$

Define

$$\begin{aligned} U_{nB}(t) \quad &= \quad {\sigma_n}^{-1}n^{-1/2}\sum_{i=1}^{n} w_i[\phi_{ni}(\hat{\beta}_n + \sigma_n n^{-1/2}t) - \phi_{ni}(\hat{\beta}_n)] \\ &\quad -n^{-1}t\sum_{i=1}^{n} w_i\phi_{1ni}(\hat{\beta}_n) - 2^{-1}\sigma_n n^{-3/2}t^2\sum_{i=1}^{n} w_i\phi_{2ni}(\hat{\beta}_n) \end{aligned}$$

Using (5.26), and proceeding along similar lines as with $S_{nB}(t)$ in the proof of Theorem 2.2, we can show that

$$\mathrm{E_B}[\sup_{|t|\le C\sigma_n} |U_{nB}(t)|]^2 = O_P(n^{-2})$$

Now, under $\mathcal{A}\cap\mathcal{W}$, we may plug in $t = \sigma_n{}^{-1}n^{1/2}(\hat{\beta_{\mathrm{B}}} - \hat{\beta})$ in $U_{nB}(t)$, and after quite some algebra we arrive at

$$\begin{aligned}
& -g_{1n}\sigma_n{}^{-1}n^{1/2}(\hat{\beta_{\mathrm{B}}} - \hat{\beta}) \\
= \;& n^{-1/2}\sum W_i\phi_{ni} - n^{-2}g_{1n}^{-1}\sum\phi_{ni}\sum W_i\phi_{1ni} \\
& -n^{-2}g_{1n}^{-1}\sum(\phi_{1ni} - E\phi_{1ni})\sum W_i\phi_{ni} + n^{-5/2}g_{1n}^{-2}\sum\phi_{ni}\sum\phi_{2ni}\sum W_i\phi_{ni} \\
& +\sigma_n n^{-3/2}g_{1n}^{-1}\sum W_i\phi_{ni}\sum W_i\phi_{1ni} + 2^{-1}\sigma_n n^{-3/2}g_{2n}g_{1n}^{-2}(\sum W_i\phi_{ni})^2 + R_{nB} \\
= \;& C_n + T_{1n} + T_{2n} + T_{3n} + T_{4n} + T_{5n} + R_{nB} \text{ say}
\end{aligned}$$

where $\mathrm{E_B}R_{nB}^2 = O_P(n^{-(1+\alpha)})$. This derivation follows from obtaining Taylor series expansion about β_0 for quantities involving $\hat{\beta}_n$ and $\hat{\beta_{\mathrm{B}}}$, then using the moment properties and asymptotic expansion of $\hat{\beta}_n - \beta_0$.

Now it can be easily checked that $\mathrm{E_B}C_n^2 = O_P(1)$, and $\mathrm{E_B}T_{in}^2 = O_P(n^{-1})$, for $i = 1,\ldots,5$. In the cross product, by direct computation $\mathrm{E_B}C_nT_{in} = O_P(n^{-1})$ for $i = 4, 5$, and hence

$$ng_{1n}^2\mathrm{V_{UBS}} = \mathrm{E_B}C_n^2 + 2\mathrm{E_B}C_n(T_{1n} + T_{2n} + T_{3n}) + O_P(n^{-1})$$

Now the rest of the proof follows by calculating the above moments. ■

Proof of Theorem 2.5 ¿From (5.37), one can easily show $\sum_{i=1}^n E(||\phi_{ni}||^2) = O(a_n^2p)$, by using any orthonormal basis of $I\!R^p$. Then we show that given any $\epsilon > 0$, $\exists\ K > 0$ and an integer n_0 such that for all $n \ge n_0$

$$Prob[|a_n^{-1}p^{-1/2}\sum_{i=1}^n\phi_{ni}| > K] < \epsilon/2 \tag{6.16}$$

This is easily established using Chebyshev's inequality. Let

$$S_n(t) = a_n^{-1}p^{-1/2}\sum_{i=1}^n[\phi_{ni}(\beta_0 + a_n^{-1}p^{1/2}t) - \phi_{ni}(\beta_0)] - G_{1n}^Tt$$

Let $M_{1n} = \sum_{i,j=1}^n\sum_{a=1}^p(\phi_{1nia} - E\phi_{1nia})(\phi_{1nja} - E\phi_{1nja})^T$. Then $E\lambda_{max}(M_{1n}) = O(a_n^2p^2)$ follows from $(B1)$. Also, $(B3)$ ensures that G_{1n} is non-singular. Because of (5.35), using $(B1)$ and $(B2)$ we have that

$$||S_n(t)||^2 \le 2a_n^{-4}||t||^2\lambda_{max}(M_{1n}) + 2^{-1}a_n^{-6}p||t||^4\sum_{i,j=1}^n\sum_{a=1}^p\lambda_{max}(M_{2nia})\lambda_{max}(M_{2nja})$$

from which we have that

$$[\sup_{||t||\leq C}||S_n(t)||]^2 \leq 2a_n^{-4}C^2\lambda_{max}(M_{1n}) + 2^{-1}a_n^{-6}pC^4\sum_{i,j=1}^{n}\sum_{a=1}^{p}\lambda_{max}(M_{2nia})\lambda_{max}(M_{2nja})$$

Now using $(B1)$ and $(B2)$ we have that $E[\sup_{||t||\leq C}||S_n(t)||]^2 = O(a_n^{-2}p^2)$. Note that

$$\inf_{|t|=C}\{a_n^{-1}p^{-1/2}t^T\sum_{i=1}^{n}\phi_{ni}(\beta_0+a_n^{-1}p^{1/2}t)\} \geq -C\sup_{|t|=C}||S_n(t)||+C^2l_{1n}-Ca_n^{-1}p^{-1/2}||\sum_{i=1}^{n}\phi_{ni}||$$

where $l_{1n} = \lambda_{min}(2^{-1}(G_{1n} + G_{1n}^T))$. By choosing C large enough, from the previous calculations we have that

$$\begin{aligned}
& Prob[\inf_{|t|=C}\{a_n^{-1}p^{-1/2}t^T\sum_{i=1}^{n}\phi_{ni}(\beta_0 + a_n^{-1}p^{1/2}t)\} > 0] \\
\geq\ & Prob[a_n^{-1}p^{-1/2}||\sum_{i=1}^{n}\phi_{ni}|| + \sup_{|t|=C}||S_n(t)|| < Cl_{1n}] \\
=\ & 1 - Prob[a_n^{-1}p^{-1/2}||\sum_{i=1}^{n}\phi_{ni}|| + \sup_{|t|=C}||S_n(t)|| > Cl_{1n}] \\
\geq\ & 1 - Prob[a_n^{-1}p^{-1/2}||\sum_{i=1}^{n}\phi_{ni}|| > Ck_2/2] - Prob[\sup_{|t|=C}||S_n(t)|| > Ck_2/2] \\
\geq\ & 1-\epsilon \text{ for all } n \text{ sufficiently large}
\end{aligned}$$

Using continuity of $\sum_{i=1}^{n}\phi_{ni}(\lambda)$ in λ, by Theorem 6.4.3 of Ortega and Rheinboldt (1970) this means that for fixed $\epsilon > 0$ for all n sufficiently large $\exists C$ large such that

$$\sum_{i=1}^{n}\phi_{ni}(\beta_0 + a_n^{-1}p^{1/2}t) = 0 \text{ has a root } T_n \text{ in } |t| \leq C \text{ with probability} > 1-\epsilon \quad (6.17)$$

Putting $\hat{\beta}_n = \beta_0 + a_n^{-1}p^{1/2}T_n$, we get a solution to (1.1) which satisfies, for fixed $\epsilon > 0$, $Prob[a_np^{-1/2}||\hat{\beta}_n - \beta_0|| \leq C] \geq 1-\epsilon$ for all n large enough. This shows (2.11). Now notice that with this C fixed, we have actually shown that with $t = T_n$

$$a_np^{-1/2}G_{1n}(\hat{\beta}_n - \beta_0) = -a_n^{-1}p^{-1/2}\sum_{i=1}^{n}\phi_{ni} + r_{n1}$$

where $Er_{n1}^2 = O(a_n^{-2}p^2)$. The representations (2.14) now follows easily, and (2.15) is obtained by slightly modifying the above steps. ■

Corollary 2.3 is immediate.

Proof of Theorem 2.6 We prove (2.17) and (2.21) first. Since $\sum_{i=1}^{n}\phi_{ni}(\hat{\beta}_n) = 0$, we have for some constant $k > 0$

$$\begin{aligned}
& \mathrm{P_B}[\sigma_n{}^{-1}p^{-1/2}a_n^{-1}||\sum w_i\phi_{ni}(\hat{\beta}_n)|| > K] \\
=\ & \mathrm{P_B}[p^{-1/2}a_n^{-1}||\sum W_i\phi_{ni}(\hat{\beta}_n)|| > K] \\
\leq\ & kK^{-2}p^{-1}a_n^{-2}\sum||\phi_{ni}(\hat{\beta}_n)||^2
\end{aligned}$$

Using (5.35), we have that $\phi_{nia}(\hat{\beta}_n + t) = \phi_{nia} + \phi_{1nia}^T t + 2^{-1} t^T H_{2nia}(\beta_2) t$ for some $\beta_2 = \beta_0 + c(\hat{\beta}_n - \beta_0)$. On the set $\{||\hat{\beta}_n - \beta_0|| < \delta_0/2\}$, with a little algebra, this leads to

$$\begin{aligned}\sum_{i=1}^{n} ||\phi_{ni}(\hat{\beta}_n)||^2 &\leq k[\sum_{i=1}^{n} ||\phi_{ni}||^2 + ||\hat{\beta}_n - \beta_0||^2 \sum_{i=1}^{n}\sum_{a=1}^{p} ||\phi_{1nia}||^2 \\ &\quad + ||\hat{\beta}_n - \beta_0||^4 \sum_{i=1}^{n}\sum_{a=1}^{p} \lambda_{max}^2(M_{2nia})]\end{aligned}$$

Thus

$$\begin{aligned}&\mathrm{P_B}[\sigma_n{}^{-1} p^{-1/2} a_n^{-1} ||\sum w_i \phi_{ni}(\hat{\beta}_n)|| > K] \\ \leq\ & kK^{-2} p^{-1} a_n^{-2} [\sum_{i=1}^{n} ||\phi_{ni}||^2 + ||\hat{\beta}_n - \beta_0||^2 \sum_{i=1}^{n}\sum_{a=1}^{p} ||\phi_{1nia}||^2 \\ & + ||\hat{\beta}_n - \beta_0||^4 \sum_{i=1}^{n}\sum_{a=1}^{p} \lambda_{max}^2(M_{2nia})] = U_K\end{aligned}$$

Note that $U_K = K^{-2} O_P(1)$, so that for fixed $\delta_1, \delta_2 > 0$, by choosing K large enough we have,

$$Prob[\mathrm{P_B}[\sigma_n{}^{-1} p^{-1/2} a_n^{-1} ||\sum w_i \phi_{ni}(\hat{\beta}_n)|| > K] > \delta_1] < \delta_2$$

¿From (5.35) we have that

$$\phi_{nia}(\hat{\beta}_n + \sigma_n p^{1/2} a_n^{-1} t) = \phi_{nia}(\hat{\beta}_n) + \sigma_n p^{1/2} a_n^{-1} \phi_{1nia}^T t + 2^{-1} \sigma_n^2 p a_n^{-2} t^T H_{2nia}(\beta_1) t$$

where $\beta_1 = \hat{\beta}_n + c p^{1/2} a_n^{-1} t$ for some $0 < c < 1$. Let

$$S_{nB}(t) = \sigma_n{}^{-1} p^{-1/2} a_n^{-1} \sum_{i=1}^{n} w_i [\phi_{ni}(\hat{\beta}_n + \sigma_n p^{1/2} a_n^{-1} t) - \phi_{ni}(\hat{\beta}_n)] - \Gamma_{1n}^T(\hat{\beta}_n) t$$

We have the a^{th} element of the vector $S_{nB}(t)$ given by

$$\begin{aligned}&S_{nBa}(t) \\ =\ & a_n^{-2} \sum w_i \phi_{1nia}^T(\hat{\beta}_n) t + 2^{-1} \sigma_n p^{1/2} a_n^{-3} t^T \sum w_i H_{2nia}(\beta_1) t - \Gamma_{1n}^T(\hat{\beta}_n) t \\ =\ & a_n^{-2} \sigma_n \sum W_i \phi_{1nia}^T(\hat{\beta}_n) t + 2^{-1} \sigma_n p^{1/2} a_n^{-3} t^T \sum w_i H_{2nia}(\beta_1) t\end{aligned}$$

On the set $\{||t|| \leq C\} \cap \{||\hat{\beta}_n - \beta_0|| < \delta_0/2\}$, we thus have for large n

$$\begin{aligned}&||S_{nB}(t)||^2 \\ \leq\ & 2\sigma_n^2 a_n^{-4} C^2 \lambda_{max}(\sum_{a=1}^{p}\sum_{i,j=1}^{n} W_i W_j \phi_{1nia}(\hat{\beta}_n) \phi_{1nja}(\hat{\beta}_n)^T) \\ & + \sigma_n^2 p a_n^{-6} C^4 \sum_{a=1}^{p} (\sum_{i=1}^{n} w_i \lambda_{max}(M_{2nia}))^2 \\ =\ & T_1 + T_2 \text{ say}\end{aligned}$$

Define $b_{n,p} = \sigma_n^{-1} p^{-1} a_n$. Thus we have

$$\begin{aligned}
& \mathrm{P_B}[b_{n,p} \sup_{||t|| \le C} ||S_{nB}(t)|| > 2K] \\
\le & \sum_{j=1}^{2} \mathrm{P_B}[b_{n,p}^2 T_j > K] + O_P(a_n^{-1} p^{1/2})
\end{aligned}$$

Now on the set $\{||\hat{\beta}_n - \beta_0|| < \delta_0/2\}$,

$$\begin{aligned}
& \mathrm{P_B}[2\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 \lambda_{max}(\sum_{a=1}^{p} \sum_{i,j=1}^{n} W_i W_j \phi_{1nia}(\hat{\beta}_n) \phi_{1nja}(\hat{\beta}_n)^T) > K] \\
\le & 2\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} \mathrm{E_B} \lambda_{max}(\sum_{a=1}^{p} \sum_{i,j=1}^{n} W_i W_j \phi_{1nia}(\hat{\beta}_n) \phi_{1nja}(\hat{\beta}_n)^T) \\
\le & 2\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} \sum_{a=1}^{p} \sum_{i,j=1}^{n} tr(\mathrm{E_B} W_i W_j \phi_{1nia}(\hat{\beta}_n) \phi_{1nja}(\hat{\beta}_n)^T) \\
\le & 2\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} \sum_{a=1}^{p} \sum_{i=1}^{n} tr(\phi_{1nia}(\hat{\beta}_n) \phi_{1nia}(\hat{\beta}_n)^T) \\
& + 2\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} c_{11} tr(\sum_{a=1}^{p} \sum_{i,j=1,\ i \ne j}^{n} \phi_{1nia}(\hat{\beta}_n) \phi_{1nja}(\hat{\beta}_n)^T) \\
\le & 2\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} \sum_{a=1}^{p} \sum_{i=1}^{n} ||\phi_{1nia}(\hat{\beta}_n)||^2 \\
& + 2\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} c_{11} \sum_{a=1}^{p} \sum_{i,j=1,\ i \ne j}^{n} ||\phi_{1nia}(\hat{\beta}_n)|| ||\phi_{1nja}(\hat{\beta}_n)|| \\
\le & k\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} \sum_{a=1}^{p} \sum_{i=1}^{n} ||\phi_{1nia}(\hat{\beta}_n)||^2 \\
\le & k\sigma_n^2 b_{n,p}^2 a_n^{-4} C^2 K^{-1} \sum_{a=1}^{p} \sum_{i=1}^{n} [||\phi_{1nia}||^2 + ||\hat{\beta}_n - \beta_0||^2 \lambda_{max}^2(M_{2nia})] \\
= & O_P(1)
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{P_B}[\sigma_n^2 p b_{n,p}^2 a_n^{-6} C^4 \sum_{a=1}^{p} (\sum_{i=1}^{n} w_i \lambda_{max}(M_{2nia}))^2 > K] \\
= & \sigma_n^2 p b_{n,p}^2 a_n^{-6} C^4 K^{-1} \sum_{a=1}^{p} \mathrm{E_B}(\sum_{i=1}^{n} w_i \lambda_{max}(M_{2nia}))^2 \\
= & \sigma_n^2 p b_{n,p}^2 a_n^{-6} C^4 K^{-1} (n + k\sigma_n^2) \sum_{a=1}^{p} \sum_{i=1}^{n} \lambda_{max}^2(M_{2nia}) \\
= & O_P(1)
\end{aligned}$$

Thus we have that $\mathrm{P_B}[b_{n,p} \sup_{||t|| \le C} ||S_{nB}(t)|| > 2K] = O_P(1)$.

Now observe that

$$\inf_{|t|=C} \{\sigma_n{}^{-1} a_n^{-1} p^{-1/2} t^T \sum_{i=1}^{n} w_i \phi_{ni}(\hat{\beta}_n + \sigma_n a_n^{-1} p^{1/2} t)\}$$
$$\ge -C \sup_{|t|=C} ||S_n(t)|| + C^2 \hat{l}_{1n} - C\sigma_n{}^{-1} a_n^{-1} p^{-1/2} || \sum_{i=1}^{n} w_i \phi_{ni}(\hat{\beta}_n)||$$

where $\hat{l}_{1n} = \lambda_{min}(2^{-1}(\Gamma_{1n}(\hat{\beta}_n) + \Gamma_{1n}^T(\hat{\beta}_n)))$. By choosing C large enough, from the previous calculations we have that on the set $\{||\hat{\beta}_n - \beta_0|| < \delta_0/2\}$,

$$\begin{aligned}
& \mathrm{P_B}[\inf_{|t|=C} \{\sigma_n{}^{-1} a_n^{-1} p^{-1/2} t^T \sum_{i=1}^{n} w_i \phi_{ni}(\hat{\beta}_n + \sigma_n a_n^{-1} p^{1/2} t)\} > 0] \\
\ge \ & \mathrm{P_B}[\sigma_n{}^{-1} a_n^{-1} p^{-1/2} || \sum_{i=1}^{n} \phi_{ni}(\hat{\beta}_n)|| + \sup_{|t|=C} ||S_n(t)|| < C\hat{l}_{1n}] \\
= \ & 1 - \mathrm{P_B}[a_n^{-1} p^{-1/2} || \sum_{i=1}^{n} \phi_{ni}|| + \sup_{|t|=C} ||S_n(t)|| > C\hat{l}_{1n}] \\
\ge \ & 1 - \mathrm{P_B}[\sigma_n{}^{-1} a_n^{-1} p^{-1/2} || \sum_{i=1}^{n} w_i \phi_{ni}|| > Ck_2/2] - \mathrm{P_B}[\sup_{|t|=C} ||S_n(t)|| > Ck_2/2]
\end{aligned}$$

Thus for fixed $\delta_1, \delta_2 > 0$, we have that for C large enough for all large n,

$$\begin{aligned}
& Prob[\mathrm{P_B}[\inf_{|t|=C} \{\sigma_n{}^{-1} a_n^{-1} p^{-1/2} t^T \sum_{i=1}^{n} w_i \phi_{ni}(\hat{\beta}_n + \sigma_n a_n^{-1} p^{1/2} t)\} > 0] < 1 - \delta_1] \\
\le \ & Prob[\mathrm{P_B}[\sigma_n{}^{-1} a_n^{-1} p^{-1/2} || \sum_{i=1}^{n} w_i \phi_{ni}|| > Ck_2/2] > \delta_1/2] \\
& + Prob[\mathrm{P_B}[\sup_{|t|=C} ||S_n(t)|| > Ck_2/2] > \delta_1/2] + O(a_n^{-1} p^{1/2}) \\
\le \ & \delta_2
\end{aligned}$$

Using continuity of $\sum_{i=1}^{n} w_i \phi_{ni}(\lambda)$ in λ, by Theorem 6.4.3 of Ortega and Rheinboldt (1970) this means that for fixed $\epsilon, \delta > 0$ for all n sufficiently large $\exists C$ large such that the bootstrap probability that $\sum_{i=1}^{n} w_i \phi_{ni}(\hat{\beta}_n + \sigma_n a_n^{-1} p^{1/2} t) = 0$ has a root T_n in $|t| \le C$ is $< 1 - \epsilon$ with a probability $< \delta$. Putting $\hat{\beta_{\mathrm{B}}} = \hat{\beta}_n + \sigma_n a_n^{-1} p^{1/2} T_n$, we get a solution to (1.2) which satisfies, for fixed $\epsilon, \delta > 0$, $Prob[\mathrm{P_B}[\sigma_n{}^{-1} a_n p^{-1/2} ||\hat{\beta_{\mathrm{B}}} - \hat{\beta}_n|| \le C] < 1 - \epsilon] < \delta$ for all n large enough. This shows (2.17). Now notice that with this C fixed, we have actually shown that with $t = T_n$

$$\sigma_n{}^{-1} a_n p^{-1/2} \Gamma_{1n}(\hat{\beta}_n)(\hat{\beta_{\mathrm{B}}} - \hat{\beta}_n) = -a_n^{-1} p^{-1/2} \sum_{i=1}^{n} W_i \phi_{ni}(\hat{\beta}_n) + r_{nB1}$$

where $||r_{nB1}|| = O_{PB}(\sigma_n a_n^{-1} p)$. This shows (2.21). With slight modification to the the definition $S_{nB}(t)$, and some more calculations one can obtain

$$\sigma_n{}^{-1} a_n p^{-1/2} G_{1n}(\hat{\beta_{\mathrm{B}}} - \hat{\beta}_n) = -a_n^{-1} p^{-1/2} \sum_{i=1}^{n} W_i \phi_{ni}(\hat{\beta}_n) + r_{nB1}$$

where $||r_{nB1}|| = O_{PB}(p^{-1})$, without using $(B4)$. Now using (5.35) again, after some algebra, (2.18) is obtained. The representation (2.20) follows easily from this. ■

Acknowledgment: We are grateful to Professor J. D. Kalbfleisch for providing us with a preprint of his work with F. Hu.

7 References

ALLEN, M. AND DATTA, S. (1999) A note on bootstrapping M-estimators in ARMA models, *Jour. Time Series Anal.*, **20**, 365-379.

ARENAL-GUTIERREZ, E. AND MATRAN, C. (1996) A zero-one law approach to the central limit theorem for the weighted bootstrap mean, *Ann. Prob.*, **24**, 532-540.

BAI, Z. AND RAO, C. R. (1991) Edgeworth expansion of a function of sample means, *Ann. Statist.*, **19**, 1295-1315.

BAI, Z. AND ZHAO, L. (1986) Edgeworth expansions of distribution function of independent random variables, *Scientica Sinica A*, **29**, 1-22.

BASAWA, I. V., GODAMBE, V. P. AND TAYLOR, R. L. (1997) *Selected proceedings of the symposium on estimating functions*, IMS lecture notes monograph series, Volume 32, Institute of Mathematical Statistics, Hayward, California.

BHATTACHARYA, R. N. AND RANGA RAO, R. (1976) *Normal approximation and asymptotic expansions*, John Wiley, New York.

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

BOSE, A. AND KUSHARY, D. (1996) Jackknife and weighted jackknife estimation of the variance of M-estimators in linear regression, *Tech. Report No. 96-12*, Dept. of Statistics, Purdue University.

BROCKWELL, P. J., AND DAVIS, R. A. (1986) *Time series: theory and methods*, Springer-Verlag, New York.

CARROLL, R. J., HARDLE, W. AND MAMMEN, E. (1999) Estimation in an additive model when components are linked parametrically, *preprint available from http://sfb.wiwi.hu-berlin.de/.*

CHATTERJEE, S. (1998) Another look at the jackknife: further examples of generalised bootstrap. *Statist. Prob. Letters*, **40**, 307-319.

CHATTERJEE, S. (1999) *Generalised bootstrap techniques*, Ph. D. thesis submitted at the Indian Statistical Institute, Calcutta.

CHATTERJEE, S. AND BOSE, A. (1999a) Variance estimation in high dimensional regression models, To appear in *Statistica Sinica.*

DAVISON, A. C., HINKLEY, D. V., AND WORTON, B. J., (1992) Bootstrap likelihoods, *Biometrika*, **79**, 113-130.

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

EFRON, B. AND TIBSHIRANI, R. (1996) Using specially designed exponential families for density estimation, *Ann. Statist.*, **24**, 2431-2461.

ENGLE, R. F. AND GRANGER, C. W. J. (1987) Co-integration and error-correction: representation, estimation and testing, *Econometrica*, **55**, 251-276.

FAN, J., HARDLE, W. AND MAMMEN, E. (1998) Direct estimation of low dimensional components in additive models, *Ann. Statist.*, **26**, 943-971.

FREEDMAN, D. A. AND PETERS, S. C. (1984), Bootstrapping a regression equation: some empirical results, *J. Amer. Statist. Assoc.*, **79**, 97-106.

GODAMBE, V. P. (1991) (edited) *Estimating functions*, Clarendon Press, Oxford.

GRUET, M. A., HUET, S. AND JOLIVET, E. (1993) Practical use of bootstrap in regression, *Computer Intensive Methods in Regression*, 150-166, Stat. Comp. Physica, Heidelberg.

HALL, P. (1987) Edgeworth expansion for Student's t statistic under minimal moment conditions, *Ann. Probab.*, **15**, 920-931.

HALL, P. AND HEYDE, C. C. (1980) *Martingale limit theory and its application*, Academic Press, New York.

HASTIE, T. J. AND TIBSHIRANI, R. (1990) *Generalised additive models*, Chapman and Hall, New York.

HU, F. AND KALBFLEISCH, J. D. (2000) The estimating function bootstrap, to appear in the *Can. Jour. Statist.*

HUET, S. AND JOLIVET, E. (1989) Exactitude au second order des intervalles de confiance bootstrap pour les parametres d'un modele de regression non lineaire, *C. R. Acad. Sci. Paris Ser. I. Math.*, **308**, 429-432.

HUET, S., JOLIVET, E. AND MESSEAN, A. (1990) Some simulation results about confidence intervals and bootstrap methods in non-linear regression. *Statistics*, **21**, 369-432.

JAMES, L. F.(1997) A study of a class of weighted bootstraps for censored data, *Ann. Statist.*, **25**, 1595-1621.

JONES, M. C. (1993) Kernel density estimation when the bandwidth is large, *Australian Jour. Statist.*, **35**, 319-326.

KOUL, H. L. (1996) Asymptotics of some estimators and sequential residual empiricals in nonlinear time series, *Ann. Statist.*, **24**, 380-404.

KUNSCH, H. R. (1989) The jackknife and the bootstrap for general stationary observations, *Ann. Statist.*, **17**, 1217-1241.

LAHIRI, S. N. (1992) Bootstrapping M-estimators of a multiple linear regression parameter, *Ann. Statist.*, **20**, 1548-1570.

LAI, T. L. (1994) Asymptotic properties of nonlinear least squares estimates in stochastic regression models, *Ann. Statist.*, **22**, 1917-1930.

LEE, K. W. (1990) Bootstrapping logistic regression models with random regressors, *Comm. Statist. A*, **19**, 2527-2539.

LELE, S. (1991) Resampling using estimating equations, in *Estimating equations*, edited by V. P. Godambe, Clarendon Press, Oxford.

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

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.

MAMMEN, E. AND VAN DE GEER, S. (1997) Penalised quasi-likelihood estimation in partial linear models, *Ann. Statist*, **25**, 1014-1035.

MCCULLAGH, P. AND NELDER, J. A. (1989) *Generalised Linear Models*, second edition, Chapman and Hall, London.

MOULTON, L. H. AND ZEGER, S. L. (1989) Analysing repeated measures on generalised linear models via the bootstrap, *Biometrics*, **45**, 381-394.

MOULTON, L. H. AND ZEGER, S. L. (1991) Bootstrapping generalised linear models, *Compu. Statist. Data Anal.*, **11**, 53-63.

NEWTON, M. A. AND RAFTERY, A. E. (1994) Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion), *Jour. Royal Statist. Soc., B*, **56**, 3-48.

NEYMAN, J. AND SCOTT, E. L. (1948) Consistent estimates based on partially consistent observations, *Econometrica*, **16**, 1-32.

ORTEGA, J. M. AND RHEINBOLDT, W. C. (1970) *Iterative solution of nonlinear equations in several variables*, Academic Press, New York.

PRAESTGAARD, J. AND WELLNER, J. A. (1993) Exchangeably weighted bootstrap of the general empirical process, *Ann. Prob.*, **21**, 2053-2086.

RAO, C. R. AND ZHAO, L. C. (1992) Approximations to the distribution of M-estimates in linear models by randomly weighted bootstrap, *Sankhyā, Series A*, **54**, 323-331.

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

SHAO, J. (1992a) One-step jackknife for M-estimators computed using Newton's method, *Ann. Inst. Statist. Math.*, **44**, 687-701.

SHAO, J. (1992b) Consistency of least squares estimator and its jackknife variance estimator in nonlinear models, *Canadian J. Statist.*, **20**, 415-428.

SHAO, J. (1992c) Jackknifing in generalised linear models, *Ann. Inst. Statist. Math.*, **44**, 673-686.

SHAO, J. AND TU, D. (1995) *The jackknife and bootstrap*, Springer-Verlag, New York.

SHI, J. (1998) Edgeworth expansion in partial linear models, *Acta Math. Sinica (New Series)*, **14**, 303-314.

SIMONOFF, J. S. AND TSAI, C. (1989) Jackknifing and bootstrapping quasi-likelihood estimators, *J. Statist. Compu. Simul*, **30**, 213-232.

STUTE, W., GONZALEZ MANTEIGA, W., AND PRESEDO QUINDIMIL, M. (1998) Bootstrap approximations in model checks for regression, *Jour. Amer. Statist. Assoc.*, **93**, 141-149.

WU, C. F. J. (1981) Asymptotic theory of nonlinear least squares estimation, *Ann. Statist.*, **9**, 501-513.

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

THEORETICAL STATISTICS AND MATHEMATICS UNIT,
INDIAN STATISTICAL INSTITUTE,
203 B. T. ROAD, CALCUTTA 700035,
INDIA
res9514@isical.ac.in
FAX:91 33 577 3071

DEPARTMENT OF STATISTICS,
PURDUE UNIVERSITY,
1399 MATHEMATICAL SCIENCES BUILDING,
WEST LAFAYETTE, IN 47907-1399
abose@stat.purdue.edu
765 494 0558