Recycled Least Squares Estimation in Nonlinear Regression

Ben Boukai∗ and Yue Zhang†
Department of Mathematical Sciences, IUPUI
Indianapolis, Indiana, 46202

December 18, 2018

Abstract
We consider a re-sampling scheme for parameters’ estimates in nonlinear regression models. We provide an estimation procedure which recycles, via random weighting, the relevant parameters estimates to construct consistent estimates of the sampling distribution of the various estimates. We establish the asymptotic normality of the resampled estimates and demonstrate the applicability of the recycling approach in a small simulation study and via example.

Keywords: Bootstrapping; Resampling; Random Weights; Nonlinear Regression.

1 Introduction

One of the most commonly used approaches for estimation of the parameters in nonlinear regression models is that of the least squares method. However, these least squares estimators do not generally exhibit tractable nor optimal finite-sample properties, largely due to nonlinearity of the regression function. Consequently, statistical inference in these cases relies primarily on either, asymptotic results and/or simulations. Specifically, consider the following regression model which specifies the relationship between the observations on a response $Y$ and the corresponding $x = (x_1, x_2, \ldots, x_p)^t \in \mathbb{R}^p$, with $p > 1$, via a nonlinear function $f(\cdot)$ as

$$y_i = f(x_i; \theta) + \epsilon_i, \quad i = 1, \ldots, n$$

(1)

where $x_i$ is the $i$th fixed input which gives rise to the observation $y_i$. $\theta = (\theta_1, \theta_2, \ldots, \theta_p)^t$ is the unknown $p \times 1$ parameter from a compact parameter space $\Theta \subset \mathbb{R}^p$. Here

$$f_i(\theta) := f(x_i; \theta), \quad i = 1, \ldots, n$$

are assumed to be continuous functions in $\theta \in \Theta$ and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and identically distributed (i.i.d.) random error terms with mean 0 and unknown variance $\sigma^2$. We denote by $\theta_0$ the true, though unknown, value of $\theta$ in $\Theta$. We assume throughout that $\theta_0 \in \text{int}(\Theta)$. The $(p$-variate) minimizer, $\hat{\theta}_n$, of the residual sum of squares

$$Q_n(\theta) := \sum_{i=1}^{n} (y_i - f_i(\theta))^2,$$

(2)
is the least square estimator (LSE) of $\theta_0$. The existence and strong consistency of $\hat{\theta}_n$ as an estimator of $\theta_0$ was established by Jennrich (1969) under the rather general conditions of model (1) and Condition J below. Under some additional conditions (see Assumption A, below), Wu (1981) established the asymptotic normality of $\hat{\theta}_n$, so that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N_p(0, \sigma^2 \Sigma_0),$$

where $\Sigma_0$ is some $p \times p$-positive definite matrix (see Assumption A, below) and $N_p$ denotes the $p-$variate normal distribution.

In this paper, we are interested in the re-sampled version of the LSE $\hat{\theta}_n$ and its properties. The implementation of various bootstrapping techniques in the context of linear regression models has received much attention in the literature. Various approaches, such as the ‘naive’, the ‘residual’, the ‘pair’ and the ‘wild’ bootstrap have been thoroughly studied under various conditions (see, Efron (1979), Efron and Tibshirani (1986), Wu (1986)). However, here we focus attention on a resampling approach for the nonlinear regression model in (1) based on random weighting technique (see for example Zheng and Tu (1988), Mason and Newton (1992), or Chatterjee and Bose (2005)).

In Section 2 we review and provide for completeness some of the known results on LSE in the non-linear regression context. In Section 3 we discuss the random weights we employ and describe the re-sampling procedure we propose for estimating the sampling distribution of $\hat{\theta}_n$. We term the resulting estimates, $\hat{\theta}_n^*$, as obtained using this random-weighting scheme, as a recycled estimate of $\theta_0$. We study the conditions for its strong consistency and the asymptotic normality of the recycled version $\hat{\theta}_n^*$ of $\hat{\theta}_n$. In Section 4, we provide the results of a simulation study along with a numerical illustration. For the random weighting technique, it is also of interest to compare their relative performances for different choices of random weights and their distributions, which are also provided in Section 4 along with a closing discussion. Section 5 is dedicated to technical details and proofs.

2 On the LS Estimation

Jennrich (1969) provided the existence and strong consistency of $\hat{\theta}_n$, the LSE of $\theta_0$ for the nonlinear regression model in (1) as the minimizer of (2), under the following general condition:

Condition J: $D_n(\theta, \theta') \equiv \frac{1}{2} \sum_{i=1}^{n} (f_i(\theta) - f_i(\theta'))^2 \rightarrow D(\theta, \theta')$ uniformly $n \rightarrow \infty$, where $D(\cdot, \cdot)$ is a continuous function for any $\theta, \theta' \in \Theta$ and $D(\theta, \theta_0) = 0$ if and only if $\theta = \theta_0$.

As is readily available from (2), the least squares estimator $\hat{\theta}_n$ of $\theta_0$ is the $p$-simultaneous solution of $\nabla Q_n(\theta) := 2 \sum_{i=1}^{n} \phi_i(\theta) = 0$, with

$$\phi_i(\theta) = -(y_i - f_i(\theta)) \nabla f_i(\theta), \quad i = 1, \ldots, n,$$

(4)

and where $\nabla f_i(\theta) = (h_{i,j}), \quad j = 1, \ldots, p$ is a $p \times 1$ vector of the partial derivatives of $f_i$, with $h_{i,j} = \partial f_i(\theta)/\partial \theta_j$, $i = 1, \ldots, n$. Similarity we denote by $\nabla^2 f_i(\theta) = (a_{i,j,k}(\theta)), \quad j, k = 1 \ldots p$ the $p \times p$ matrix of partial second derivatives of $f_i$, with $a_{i,j,k}(\theta) := \partial^2 f_i(\theta)/\partial \theta_j \partial \theta_k$, $i = 1, \ldots, n$. Along with this notation, we denote by $\nabla \phi_i(\theta)$, the $p \times p$ matrix,

$$\nabla \phi_i(\theta) = \nabla f_i(\theta) \nabla f_i(\theta)^t - (y_i - f_i(\theta)) \nabla^2 f_i(\theta).$$

(5)

Additionally, we set

$$\Sigma_n^{-1}(\theta) := \frac{1}{n} \sum_{i=1}^{n} \nabla f_i(\theta) \nabla f_i(\theta)^t,$$

(6)
so that by (4) and (6),
\[ \sum_{i=1}^{n} E \left[ \phi_i(\theta_0) \cdot (\phi_i(\theta_0))^t \right] \equiv n\sigma^2 \Sigma_n^{-1}(\theta_0). \] (7)

Wu (1981) assumed that aside from Condition J, the functions \( f_i \) and their gradients satisfy the following conditions (which we will also use in the sequel):

**Assumption A:** The functions \( f(\theta) = (f_1(\theta), f_2(\theta), \ldots, f_n(\theta))^t \) are such that

1. \( \nabla f_i(\theta) \) and \( \nabla^2 f_i(\theta) \) exist for all \( \theta \) near \( \theta_0 \);
2. \( \Sigma_n^{-1}(\theta_0) \to \Sigma_0^{-1} \) a \( p \times p \) positive definite matrix, as \( n \to \infty \);
3. \( \Sigma_n^{-1}(\theta) \to \Sigma_0^{-1} \), uniformly as \( n \to \infty \) and \( |\theta - \theta_0| \to 0 \);
4. there exist a \( \delta > 0 \), such that for all \( j, k \)
   \[ \lim_{n \to \infty} \sup_{|\theta - \theta_0| \leq \delta} \frac{a_{i;jk}(\theta)}{\sup_{\theta \in \Theta_\delta} |a_{i;jk}(\theta)|} < \infty \]
   for all \( i \), where \( \Theta_\delta = \{ \theta \in \Theta, |\theta - \theta_0| \leq \delta \} \), and \( \delta \) is same as in 4 above.

Under the terms of Assumption A, Wu (1981) has established the asymptotic normality of the LSE, \( \hat{\theta}_n \), as given in (3). Furthermore, for any \( c = (c_1, c_2, \ldots, c_p)^t \), with \( ||c|| = 1 \), we have again under Assumption A, that
\[ R_{n,c} := \frac{\sqrt{n} c^t (\hat{\theta}_n - \theta_0)}{\sqrt{c^t (\Sigma_n(\theta_0)) c}} \Rightarrow N(0, \sigma^2), \text{ as } n \to \infty. \] (8)

In the next section, we describe the re-sampling scheme we use to obtain the *recycled* estimate \( \hat{\theta}_n^* \) of \( \hat{\theta}_n \). We denote by \( R_{n,c}^* \) the corresponding recycled version of \( R_{n,c} \) in (3). Let, \( P^* \) denote the *resampled* probability (conditional on the given sample data) and set
\[ F_n(u) = P(\mathcal{R}_{n,c} \leq u), \quad \text{and} \quad F_n^*(u) = P^*(\mathcal{R}_{n,c}^* \leq u), \quad \forall u \in \mathbb{R}, \] (9)
to denote the corresponding c.d.f of \( R_{n,c} \) and \( R_{n,c}^* \), respectively. We show in particular that under the conditions given in Theorem 2 below,
\[ \sup_u |F_n^*(u) - F_n(u)| \to 0 \quad \text{a.s. as } n \to \infty. \]

This result allows us to approximate the sampling distribution of \( \hat{\theta}_n \) based on the sampling distribution of its *recycled* version, \( \hat{\theta}_n^* \).
3 Recycled Estimation via Random Weighting

For each \( n \geq 1 \), we let the random weights, \( \mathbf{w}_n = (w_{1:n}, w_{2:n}, \ldots, w_{n:n})^t \), be a vector of exchangeable nonnegative random variables with \( E(w_{i:n}) = 1 \) and \( \text{Var}(w_{i:n}) := \tau_n^2 \), and let \( W_i \equiv W_{1:n} = (w_{i:n} - 1)/\tau_n \) be the standardized version of \( w_{i:n}, i = 1, \ldots, n \). In addition we also assume,

**Assumption W**: The underlying distribution of the random weights \( \mathbf{w}_n \) satisfies

1. For all \( n \geq 1 \), the random weights \( \mathbf{w}_n \) are independent of \( (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^t \);
2. \( \tau_n^2 = o(n) \), \( E(W_i W_j) = O(n^{-1}) \) and \( E(W_i^2 W_j^2) \to 1 \) for all \( i \neq j \), \( E(W_i^4) < \infty \) for all \( i \).

The following are examples of random weights that satisfy the above conditions in Assumption W.

i) Multinomial weights, \( \mathbf{w}_n \sim \text{Multinomial}(n, 1/n, 1/n, \ldots, 1/n) \), which reflect a simple random re-sampling scheme (with replacement) and essentially correspond to the classical bootstrap of Efron (1979).

ii) Dirichlet weights, \( \mathbf{w}_n \equiv n \times \mathbf{z}_n \) where \( \mathbf{z}_n \sim \text{Dirichlet}(\alpha, \alpha, \ldots, \alpha) \), with \( \alpha > 0 \) which often refer to as the Bayesian bootstrap (see Rubin (1981), and its variants as in Zheng and Tu (1988) and Lo (1991)).

We will assume throughout this paper that all the random weights we use in the sequel do satisfy Assumption W. With such random weights \( \mathbf{w}_n \) at hand, we define in similarity to \([2]\), the recycled version \( \hat{\theta}_n^* \) of \( \hat{\theta}_n \) as the minimizer of the randomly weighted least squares criterion.

\[
Q_n^*(\theta) := \sum_{i=1}^n w_{i:n}(y_i - f_i(\theta))^2, \tag{10}
\]

or alternatively as the \( p \)-simultaneous solution of \( \nabla Q_n^*(\theta) := 2 \sum_{i=1}^n w_{i:n} \phi_i(\theta) = 0 \). The next results establishes the strong consistency of the recycled estimator \( \hat{\theta}_n^* \) for \( \theta_0 \).

**Theorem 1** Let \( \hat{\theta}_n^* \) be the minimizer of \( Q_n^*(\theta) \) in \([10]\) and suppose that \( E(\epsilon_1^4) < \infty \) and that in addition to Condition J, \( d_i(\theta) := f_i(\theta) - f_i(\theta_0) \) also satisfy

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \sup_{\theta - \theta_0 \leq \delta} d_i^4(\theta) < \infty,
\]

Then, \( \hat{\theta}_n^* \to \theta_0 \) a.s., as \( n \to \infty \).

Once \( \hat{\theta}_n^* \) is obtained, the recycled version \( \mathcal{R}_{n,c}^* \) of \( \mathcal{R}_{n,c} \) in \([3]\) can readily be obtained as:

\[
\mathcal{R}_{n,c}^* := \frac{\sqrt{n} \epsilon^t (\hat{\theta}_n^* - \hat{\theta}_n)}{\tau_n \epsilon^t \Sigma_n^0(\hat{\theta}_n)^c}, \tag{11}
\]

where \( \Sigma_n^{-1}(\hat{\theta}_n) \) is as in \([5]\), evaluated at \( \hat{\theta}_n \). Below we show that under some additional conditions to those stated above, we may obtain the asymptotic normality of \( \mathcal{R}_{n,c}^* \). However, aside from the strong consistency of \( \hat{\theta}_n^* \) we will also need the following additional conditions in order to establish the consistency of the resampling estimator for estimating the sampling distribution of \( \hat{\theta}_n \).
In addition to Assumption A, we assume that

**Assumption B:** In addition to Assumption A, we assume that

1. \( \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{|\theta - \theta_0| \leq \delta} a_{i,jk}^4(\theta) < \infty, \) for all \( j,k. \)
2. \( \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{|\theta - \theta_0| \leq \delta} h_{i,j}^4(\theta) < \infty, \) for each \( j = 1, \ldots, p; \)

**Theorem 2** Let \( \hat{\theta}_n \) and \( \hat{\theta}_n^* \) be a strongly consistent least squares estimator and the recycled estimator of \( \theta_0 \) respectively. Then, under Assumptions A \& B we have, \( R_{n,c} \Rightarrow N(0, \sigma^2), \) as \( n \to \infty, \) and hence,

\[
\sup_u |F_{n}^*(u) - F_{n}(u)| \to 0 \quad a.s.
\]

**Remark:** The technique we use for the proof of Theorem 2 (see Section 5 below) can similarity be used, under a stronger version of Assumption B.2, to also establish the consistency of the (resampling) variance of \( \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)/\tau_n. \) However, this result will be presented elsewhere.

## 4 Implementation and Numerical Results

It is clear from the above description that the *recycled* estimate \( \hat{\theta}_n^* \) of \( \theta_0, \) using the random weights \( w_n, \) is straightforward to implement. Further, in light of Theorem 2 one can also obtain a finite-sample approximation to the sampling distribution of \( \hat{\theta}_n \) using the (recycled) sampling distribution of \( \hat{\theta}_n. \) To that end, generate \( B \) independent replications of the random weight \( w_n, \) independent of the given sample data \( (y, X), \) to obtain the \( B \) recycled replications, \( \hat{\theta}_n^1, \hat{\theta}_n^2, \ldots, \hat{\theta}_n^B. \) Then obtain the \( B \) corresponding recycled replications, \( R_{n,c}^1, R_{n,c}^2, \ldots, R_{n,c}^B \) of \( R_{n,c} \) as in (11). Finally, a finite-sample approximation to the sampling distribution, \( F_{n}^* \) of \( R_{n,c}^* \) can be obtained by the empirical c.d.f of the recycled \( R_{n,c}^* \)

\[
\hat{F}_{B,n}(t) = \frac{1}{B} \sum_{b=1}^{B} I \left[ R_{n,c}^{*b} \leq t \right].
\]

When the error terms variance, \( \sigma^2, \) is unknown, we may estimate it by \( \hat{\sigma}_n^2 := Q_n(\hat{\theta}_n)/(n-p) \) to obtain the ‘studentized’ version, \( R_{n,c}^{*b} \equiv R_{n,c}/\hat{\sigma}_n, \) of \( R_{n,c}^*. \) The recycled estimates of \( \hat{\sigma}_n^2 \) can be obtained in a similar manner as \( \hat{\sigma}_{n,1}^2, \hat{\sigma}_{n,2}^2, \ldots, \hat{\sigma}_{n,B}^2, \) where \( \hat{\sigma}_{n,b}^2 := Q_n(\hat{\theta}_n^b)/(n-p), \) for \( b = 1, 2, \ldots, B. \)

### 4.1 A Simulation Study

For the simulation study, we considered two nonlinear regression models with \( p = 2 \) each, defined by the functions

**Model I:**

\[ f(x; \theta) = \theta_1 x e^{-\theta_2 x} \]

**Model II:**

\[ f(x; \theta) = \frac{\theta_1 x}{e^{\theta_2} + x} \]

The fixed values of the independent variable, \( x_i, \) \( i = 1, \ldots, n, \) were selected uniformly from the \([0,10]\) interval and we assumed that \( \epsilon_i \sim N(0, 0.25^2) , \ i.i.d.. \) In the simulations, the value of the parameter \( \theta_0 := (\theta_{10}, \theta_{20}) \) was set to \((2, 0.04)\) and \((10, 0)\), corresponding to Model I and Model
II, respectively. For each model, $M = 10,000$ recycled samples were drawn from a sample with varying sample sizes of $n = 10, 30, 50, 80, 150$. To study the properties of the recycled estimation procedure, we compared the simulated distribution of $R_{n,c}$ to that of the recycled distribution of $R_{n,c}^*$, obtained with $c = (1/\sqrt{2}, 1/\sqrt{2})^t$. In Tables 1 and 2 below (corresponding to Model I and Model II, respectively), we present the results as were obtained using random weights $w_n$, generated from the standard (a) Multinomial, (b) Dirichlet and (c) Exponential distributions. It can be seen, under either one of these random weighting schemes, the agreement between the simulated and the recycled distribution of $R_{n,c}$ improves, as the sample size increase. This is in agreement with the main results of this paper. To highlight this result, we present in Figure 1 and Figure 2 a comparison of the resulting recycled sampling distribution of $R_{n,c}$ illustrating both, the consistency and the asymptotic normality of the recycled least squares estimate in the case of Model I.

| Sample Size | Simulated Dist | Multinomial | Dirichlet | Exponential |
|-------------|----------------|-------------|-----------|-------------|
| 10          | 0.0037         | -0.2057     | -0.1276   | -0.1126     |
| S.E.        | 1.1441         | 0.8529      | 0.7193    | 0.6534      |
| $\hat{\sigma}$ | 0.2427       | 0.3022      | 0.2744    | 0.2740      |
| 30          | 0.0077         | 0.0480      | 0.0245    | 0.0353      |
| S.E.        | 1.0135         | 0.9789      | 0.8957    | 0.8624      |
| $\hat{\sigma}$ | 0.2486       | 0.3081      | 0.3052    | 0.3051      |
| 50          | -0.0010        | 0.0296      | 0.0129    | 0.0187      |
| S.E.        | 1.0153         | 1.0603      | 1.0324    | 1.0098      |
| $\hat{\sigma}$ | 0.2490       | 0.3052      | 0.3046    | 0.3045      |
| 80          | 0.0198         | 0.0244      | 0.0232    | 0.0244      |
| S.E.        | 1.0207         | 0.9400      | 0.9194    | 0.9074      |
| $\hat{\sigma}$ | 0.2492       | 0.2548      | 0.2547    | 0.2546      |
| 150         | 0.0073         | -0.0044     | 0.0137    | 0.0632      |
| S.E.        | 1.0044         | 1.0022      | 0.9988    | 0.9704      |
| $\hat{\sigma}$ | 0.2496       | 0.2507      | 0.2506    | 0.2506      |

Table 1: Simulated Recycled distributions of $R_{n,c}^*$ for Model I using various random weights showing the means and Standard Errors

To further study the recycling estimation procedure, we took $c = (1, 0)^t$ for obtaining sampling distribution for $\theta_1$ and $c = (0, 1)^t$ for that of $\theta_2$ and a 95% confidence interval was constructed in each case of three distributions for random weights: Dirichlet, Multinomial and Exponential distribution. In each case, $B = 10,000$ replications of such recycled simulations were executed to determine the percentage of times the true value of the parameter was contained in the interval estimate and the average length of the confidence interval was calculated. The simulation results for Model I, corresponding to $\theta_1$ and $\theta_2$ are provided in Table 3 and Table 4, respectively. The first column in these tables indicates the sample size $n$. Column 2 to Column 4 provide the Coverage Percentages and the average length of the confidence interval (in parentheses).
As can be seen from these Tables, using the Multinomial random weights for the recycled estimation performs slightly better than the other two, for both, Model I and Model II. However, the performance, in terms of the average coverage percentage, using the Dirichlet and the Exponential weights are similar. Even when sample size is small, the confidence interval constructed by recycled estimate covers the true parameter well, especially in the case of the Multinomial random weights.

We note that when the sample size increases, the Coverage Percentage for all three cases of the random weights are close to 0.95 (the nominal value), with a narrower average length of the confidence interval. Similarly, the results of the simulation studies conducted in the case of Model II for \( \theta_1 \) and \( \theta_2 \) are illustrated in Tables 5 and 6, respectively. The coverage percentages and average confidence interval lengths for Model I and II are also illustrated in Figure 3 and Figure 4.
Table 2: Simulated Recycled distributions of $\hat{R}_{n,c}^*$ for Model II using various random weights showing the means and Standard Errors

| Sample Size | Multinomial Weights | Dirichlet Weights | Exponential Weights |
|-------------|---------------------|------------------|---------------------|
| 10          | 0.0281              | 0.2744           | 0.1707              | 0.1482              |
| S.E.        | 1.1787              | 1.0445           | 0.7994              | 0.7225              |
| $\hat{\sigma}$ | 0.2425              | 0.2907           | 0.2731              | 0.2732              |
| 30          | 0.0150              | 0.0867           | 0.0837              | 0.0872              |
| S.E.        | 1.0321              | 1.1519           | 1.0668              | 1.0128              |
| $\hat{\sigma}$ | 0.2485              | 0.3114           | 0.3095              | 0.3093              |
| 50          | 0.0195              | -0.0112          | 0.0070              | 0.0036              |
| S.E.        | 1.0221              | 0.9080           | 0.8712              | 0.8559              |
| $\hat{\sigma}$ | 0.2489              | 0.3012           | 0.3007              | 0.3007              |
| 80          | 0.0109              | 0.0112           | -0.0120             | 0.0191              |
| S.E.        | 1.0091              | 1.0614           | 1.0323              | 1.0349              |
| $\hat{\sigma}$ | 0.2493              | 0.2536           | 0.2534              | 0.2535              |
| 150         | 0.0175              | 0.0226           | 0.0120              | -0.0020             |
| S.E.        | 1.0224              | 0.8787           | 0.8797              | 0.8756              |
| $\hat{\sigma}$ | 0.2496              | 0.2520           | 0.2519              | 0.2520              |

Table 3: Coverage percentage (average confidence interval length) for $\theta_1$ in Model I

| n  | Multinomial | Dirichlet | Exponential |
|----|-------------|-----------|-------------|
| 10 | 0.921 (0.301) | 0.861 (0.231) | 0.862 (0.231) |
| 30 | 0.921 (0.150) | 0.902 (0.142) | 0.903 (0.142) |
| 50 | 0.936 (0.116) | 0.929 (0.112) | 0.926 (0.112) |
| 80 | 0.942 (0.092) | 0.938 (0.090) | 0.936 (0.090) |
| 150| 0.941 (0.067) | 0.938 (0.067) | 0.936 (0.067) |

Table 4: Coverage percentage (average confidence interval length) for $\theta_2$ in Model I

| n  | Multinomial | Dirichlet | Exponential |
|----|-------------|-----------|-------------|
| 10 | 0.928 (0.023) | 0.841 (0.016) | 0.839 (0.016) |
| 30 | 0.934 (0.010) | 0.909 (0.009) | 0.907 (0.009) |
| 50 | 0.932 (0.008) | 0.915 (0.007) | 0.915 (0.007) |
| 80 | 0.936 (0.006) | 0.932 (0.006) | 0.929 (0.006) |
| 150| 0.951 (0.004) | 0.945 (0.004) | 0.943 (0.004) |

4.2 Illustrative Example

As an illustrative example we consider the Chwirut\textsuperscript{1} data file from NIST-nonlinear least square datasets. These data are the result of a NIST study involving ultrasonic calibration. The response variable is ultrasonic response, and the predictor variable is metal distance (Chwirut, 1979).
model was proposed as:

\[ f(x; \theta) = \frac{e^{-\theta_1 x}}{\theta_2 + \theta_3 x} \]

where \( f \) is predicted response and \( x \) is metal distance. The given sample data were resampled with various random weights (Multinomial, Dirichlet and Exponential) for a total of \( M = 10,000 \) times so as to produce the recycled estimate, \( \hat{\theta}_n^* \) of \( \theta = (\theta_1, \theta_2, \theta_3)' \) (and of \( \sigma \) too). The results are provided in Table 7 below which also provide the corresponding least squares estimates, \( \hat{\theta}_n \), and \( \sigma \) for a comparison.

### Table 5: Coverage Percentage (average confidence interval length) for \( \theta_1 \) in Model II

| n   | Multinomial | Dirichlet | Exponential |
|-----|-------------|-----------|-------------|
| 10  | 0.924 (0.977) | 0.875 (0.716) | 0.876 (0.716) |
| 30  | 0.930 (0.468) | 0.910 (0.437) | 0.910 (0.437) |
| 50  | 0.945 (0.359) | 0.934 (0.345) | 0.933 (0.345) |
| 80  | 0.942 (0.285) | 0.935 (0.278) | 0.934 (0.278) |
| 150 | 0.949 (0.207) | 0.947 (0.205) | 0.946 (0.205) |

### Table 6: Coverage Percentage (average confidence interval length) for \( \theta_2 \) in Model II

| n   | Multinomial | Dirichlet | Exponential |
|-----|-------------|-----------|-------------|
| 10  | 0.893 (0.544) | 0.806 (0.355) | 0.807 (0.354) |
| 30  | 0.897 (0.317) | 0.805 (0.347) | 0.808 (0.347) |
| 50  | 0.933 (0.237) | 0.906 (0.215) | 0.907 (0.215) |
| 80  | 0.935 (0.178) | 0.914 (0.168) | 0.916 (0.168) |
| 150 | 0.942 (0.139) | 0.929 (0.135) | 0.932 (0.135) |

Figure 3: Coverage Percentage (average confidence interval length) for \( \theta_1, \theta_2 \) in Model I.
5 Technical Details and Proofs

This section provides the technical details and proofs of the main results of the paper stated in Theorems 1 and 2. To establish these results, we repeatedly use the results stated in Lemma 1 below, which summaries similar technical steps from Wu (1981) (see also Corollary A there).

**Lemma 1** Let $\Theta$ be a compact subset of $\mathbb{R}^p$ and $g_1, g_2, \ldots, g_n$ functions over $\Theta$ satisfying the following assumptions:

a) $\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{\theta \in \Theta} g_i(\theta)^2 < \infty$

b) and if $\sum_{i=1}^{n} \sup_{\theta \in \Theta} g_i(\theta)^2 = \infty$, there exists a constant $M$, independent of $i$ such that

$$\sup_{s \neq t, s, t \in \Theta} \frac{|g_i(s) - g_i(t)|}{|s - t|} \leq M \sup_{s \in \Theta} |g_i(s)| < \infty$$

Then for the independent random variables $v_1, v_2, \ldots, v_n$ with $E(v_i) = 0$, $\sup_i E(v_i^2) < \infty$, we have, uniformly over $\Theta$, $\frac{1}{n} \sum_{i=1}^{n} g_i(\theta)v_i \to 0$ a.s. as $n \to \infty$. 

---

**Table 7**: The *recycled* Least Squares Estimates in the Chwirut1 example; displaying the mean and the SE of the sampling distribution based on $M = 10,000$ runs, along with the estimated standard deviation, $\sigma$.

|                  | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\sigma$ |
|------------------|------------|------------|------------|----------|
| $\hat{\theta}_n$ (LSE) | 0.1903     | 0.0061     | 0.0105     | 3.3617   |
| $\hat{\theta}^*$ (Multinomial wt.) | 0.1907     | 0.0061     | 0.0105     | 3.4002   |
| S.E.              | 0.0223     | 0.0005     | 0.0009     | 0.0528   |
| $\hat{\theta}^*$ (Dirichlet wt.) | 0.1906     | 0.0061     | 0.0105     | 3.3975   |
| S.E.              | 0.0216     | 0.0005     | 0.0009     | 0.0510   |
| $\hat{\theta}^*$ (Exponential wt.) | 0.1903     | 0.0061     | 0.0106     | 3.3974   |
| S.E.              | 0.0214     | 0.0005     | 0.0009     | 0.0517   |

---

**Figure 4**: Coverage Percentage (average confidence interval length) for $\theta_1, \theta_2$ in Model II.
The results in the next two Lemmas are concerned with the random weights.

**Lemma 2** With $W_i = (w_{i:n} - 1)/\tau_n$, $i = 1 \ldots , n$, as above, and $\bar W_n := \frac{1}{n} \sum_{i=1}^n W_i$ we have under Assumption W,

\[
\frac{1}{n} \sum_{i=1}^n W_i \overset{p}{\to} 0, \quad \text{as } n \to \infty ,
\]

\[
\frac{1}{n} \sum_{i=1}^n W_i^2 \overset{p}{\to} 1, \quad \text{as } n \to \infty ,
\]

and hence

\[
\frac{1}{n} \sum_{i=1}^n (W_i - \bar W_n)^2 \overset{p}{\to} 1, \quad \text{as } n \to \infty .
\]

**Proof:**
First note that

\[
E\left(\frac{1}{n} \sum_{i=1}^n W_i\right)^2 = \frac{1}{n^2} \sum_{i=1}^n E(W_i^2) + \frac{1}{n^2} \sum_{i \neq i} E(W_i W_i)
\]

\[
= \frac{1}{n} + \frac{1}{n^2} n(n-1)O\left(\frac{1}{n}\right) \to 0,
\]

as $n \to \infty$ so that,

\[
\frac{1}{n} \sum_{i=1}^n W_i \overset{p}{\to} 0.
\]

Also, since by Assumption W,

\[
E\left(\frac{1}{n} \sum_{i=1}^n (W_i^2 - 1)\right)^2 = \frac{1}{n^2} \sum_{i=1}^n E(W_i^2 - 1)^2 + \frac{1}{n^2} \sum_{i \neq j} E(W_i^2 - 1)(W_j^2 - 1)
\]

\[
= \frac{1}{n} EW_1^4 + \frac{1}{n} + \frac{1}{n^2} \sum_{i \neq j} EW_i^2 W_j^2 - \frac{1}{n^2} n(n-1) \to 0
\]

as $n \to \infty$, to obtain that,

\[
\frac{1}{n} \sum_{i=1}^n W_i^2 \overset{p}{\to} 1.
\]

Finally, we conclude that

\[
\frac{1}{n} \sum_{i=1}^n (W_i - \bar W_n)^2 = \frac{1}{n} \sum_{i=1}^n W_i^2 - \bar W_n^2 \overset{p}{\to} 1
\]

as $n \to \infty$. Q.E.D.

**Lemma 3** For the standardized random weights, $W_i = (w_{i:n} - 1)/\tau_n$, $i = 1 \ldots , n$, of Section 3, Assumption W.2 assures that $\lim_{K \to \infty} \lim_{n \to \infty} \sup \|(W_1 - \bar W_n)I_{|W_1 - \bar W_n| > K}\|_2 = 0$. 

11
Proof: By Assumption W.2 we have that \( \frac{1}{n} \sum_{i=1}^{n} (W_i - \overline{W}_n)^2 \overset{p}{\to} 1 > 0 \) and \( \lim_{n \to \infty} E(W_1 - \overline{W}_n)^4 \leq \lim_{n \to \infty} 8[EW_1^4 + EW_n^4] < \infty \). Hence \( (W_1 - \overline{W}_n)^2 \) is uniformly integrable and finally by Chebyshev’s inequality,

\[
\lim_{K \to \infty} P(|W_1 - \overline{W}_n| \geq K) \leq \lim_{K \to \infty} \frac{V(W_1 - \overline{W}_n)}{K^2} = \lim_{K \to \infty} \frac{1}{K^2} E(W_1 - \overline{W}_n)^2 = 0.
\]

Q.E.D.

Lemma 4 Let \( W_i = (w_{i:n} - 1)/\tau_n, \ i = 1, \ldots, n, \) be the standardized random weights as are given in Assumption W and let

\[
\xi_n \equiv \xi_n(\theta) = \frac{\tau_n}{n} \sum_{i=1}^{n} W_i(y_i - f_i(\theta))^2.
\]

Then, as \( n \to \infty, \) \( \xi_n(\theta) \overset{p}{\to} 0 \) uniformly in \( \theta \in \Theta \).

Proof: We begin by rewriting \( \xi_n \) as

\[
\xi_n = \frac{\tau_n}{n} \sum_{i=1}^{n} W_i(\epsilon_i + f_i(\theta_0) - f_i(\theta))^2
\]

\[
= \frac{\tau_n}{n} \sum_{i=1}^{n} W_i \epsilon_i^2 + \frac{\tau_n}{n} \sum_{i=1}^{n} W_i d_i^2(\theta) - \frac{2\tau_n}{n} \sum_{i=1}^{n} W_i \epsilon_i d_i(\theta)
\]

\[
\equiv \xi_{1:n} + \xi_{2:n} + \xi_{3:n}
\]

For each of these terms, we obtain using Jensen’s inequality,

\[
E^*(\xi_{1:n})^2 = \frac{\tau_n^2}{n^2} \left[ \sum_{i=1}^{n} \epsilon_i^4 + O\left(\frac{1}{n}\right) \sum_{i < j} 2\epsilon_i^2 \epsilon_j^2 \right]
\]

\[
\leq \frac{\tau_n^2}{n^2} (1 + (n - 1)O\left(\frac{1}{n}\right)) \sum_{i=1}^{n} \epsilon_i^4 \rightarrow 0 \text{ uniformly as } n \to \infty
\]

\[
E^*(\xi_{2:n})^2 \leq \frac{\tau_n^2}{n^2} \left[ \sum_{i=1}^{n} d_i^4(\theta) + O\left(\frac{1}{n}\right) \sum_{i < j} 2d_i^2(\theta) d_j^2(\theta) \right]
\]

\[
\leq \frac{\tau_n^2}{n^2} (1 + (n - 1)O\left(\frac{1}{n}\right)) \sum_{i=1}^{n} d_i^4(\theta) \rightarrow 0 \text{ uniformly as } n \to \infty
\]

\[
E^*(\xi_{3:n})^2 = \frac{4\tau_n^2}{n^2} \left[ \sum_{i=1}^{n} \epsilon_i^2 d_i^2(\theta) + O\left(\frac{1}{n}\right) \sum_{i < j} 2\epsilon_i \epsilon_j d_i(\theta) d_j(\theta) \right]
\]

\[
\leq \frac{4\tau_n^2}{n} (1 + (n - 1)O\left(\frac{1}{n}\right)) \left[ \frac{1}{n} \sum_{i=1}^{n} (\epsilon_i^2 - \sigma^2)^2 d_i^2(\theta) + \frac{\sigma^2}{n} \sum_{i=1}^{n} d_i^2(\theta) \right].
\]
where \( E^*(\xi_{3:n})^2 \) converges to 0 uniformly by Lemma 1 and Condition J.

Accordingly, we have \( E^*(\xi_{1:n})^2, E^*(\xi_{2:n})^2, E^*(\xi_{3:n})^2 \) all converge to 0 uniformly. Further, by Chebyshev's inequality, for any \( \epsilon > 0 \)

\[
P^*(|\xi_{2:n}| \geq \epsilon) \leq \frac{E^*(\xi_{2:n})^2}{\epsilon} \rightarrow 0 \quad \text{uniformly as} \quad n \rightarrow \infty,
\]

so that \( \xi_{2:n} \overset{p}{\rightarrow} 0 \) uniformly. Similarly, we show that \( \xi_{1:n}, \xi_{3:n} \) converge to 0 uniformly and in probability \( (P^*) \). Q.E.D.

Proof of Theorem 1: It is easy to verify that \( Q_n^*(\theta) := \sum_{i=1}^{n} w_{i:n}(y_i - f_i(\theta))^2 \) may be written as \( Q_n^*(\theta) = Q_n(\theta) + n\xi_n \). Since by Jennrich (1969) (and under Condition J), \( \frac{1}{n} Q_n(\theta) \rightarrow D(\theta, \theta_0) + \sigma^2 \), uniformly, it follows from Lemma 4 immediately that

\[
\theta \rightarrow \theta_0 \quad \text{uniformly in} \quad \theta.
\]

Further, under Assumption A.3-A.5, and Lemma 5 it also follows that

\[
n^{-1} \sum_{i=1}^{n} \epsilon_i \nabla^2 f_i(\theta) \rightarrow 0, \quad \text{a.s. as} \quad n \rightarrow \infty,
\]

uniformly in \( \theta \in \Theta_\delta \). Finally,

\[
\frac{1}{n} \sum_{i=1}^{n} [f_i(\theta) - f_i(\tilde{\theta}_n)] \nabla^2 f_i(\tilde{\theta}_n) \rightarrow 0, \quad \text{a.s. as} \quad n \rightarrow \infty,
\]

upon using the Cauchy-Schwarz inequality in conjunction with Assumption A.3-A.4. Q.E.D.
Lemma 6 Under the conditions of Assumptions A and B,
\[ n^{-1} \tau_n \sum_{i=1}^{n} W_i \nabla \phi_i(\tilde{\theta}_n) \to 0, \]
in probability, where \( W_i \) as in Assumption W and \( \nabla \phi_i(\theta) \) as in (3) and \( \tilde{\theta}_n \) is any sequence such that \( \tilde{\theta}_n \to \theta_0 \) a.s., as \( n \to \infty \).

Proof: It can be verify that since \((j,k)\)th entry of the random matrix in \( \frac{\tau_n}{n} \sum_{i=1}^{n} W_i \nabla \phi_i(\tilde{\theta}_n) \) is
\[ n^{-1} \tau_n \sum_{i=1}^{n} W_i \nabla \phi_{ijk}(\tilde{\theta}_n) := n^{-1} \tau_n \sum_{i=1}^{n} W_i \left( h_{i;j}(\tilde{\theta}_n)h_{i;k}(\tilde{\theta}_n) - (y_i - f_i(\tilde{\theta}_n))a_{i;jk}(\tilde{\theta}_n) \right). \]
We obtain that,
\[
E^* \left[ n^{-1} \tau_n \sum_{i=1}^{n} W_i \nabla \phi_{ijk}(\tilde{\theta}_n) \right]^2 = n^{-2} \tau_n^2 E^* \left[ \sum_{i=1}^{n} W_i^2 \nabla \phi_{ijk}(\tilde{\theta}_n)^2 + \sum_{i \neq h} W_i W_h \nabla \phi_{ijk}(\tilde{\theta}_n) \nabla \phi_{hjk}(\tilde{\theta}_n) \right] \\
\leq n^{-2} \tau_n^2 \left[ (n-1)O(\frac{1}{n}) + 1 \right] \sum_{i=1}^{n} \nabla \phi_{ijk}(\tilde{\theta}_n)^2.
\]
Since, by Assumption W, we have that \( \tau_n^2 = o(n) \), we only need to show
\[
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \nabla \phi_{ijk}(\tilde{\theta}_n)^2 < \infty.
\]
However with straightforward rearrangements, we have
\[
n^{-1} \sum_{i=1}^{n} \nabla \phi_{ijk}(\tilde{\theta}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} \left( h_{i;j}(\tilde{\theta}_n)h_{i;k}(\tilde{\theta}_n) - (y_i - f_i(\tilde{\theta}_n))a_{i;jk}(\tilde{\theta}_n) \right)^2 \\
= \frac{1}{n} \sum_{i=1}^{n} \left( h_{i;j}(\tilde{\theta}_n) \right)^2 \left( h_{i;k}(\tilde{\theta}_n) \right)^2 + \frac{1}{n} \sum_{i=1}^{n} (y_i - f_i(\tilde{\theta}_n))^2 \left( a_{i;jk}(\tilde{\theta}_n) \right)^2 \\
- \frac{2}{n} \sum_{i=1}^{n} h_{i;j}(\tilde{\theta}_n)h_{i;k}(\tilde{\theta}_n)(y_i - f_i(\tilde{\theta}_n))a_{i;jk}(\tilde{\theta}_n) \\
= \frac{1}{n} \sum_{i=1}^{n} \left( h_{i;j}(\tilde{\theta}_n) \right)^2 \left( h_{i;k}(\tilde{\theta}_n) \right)^2 + \frac{1}{n} \sum_{i=1}^{n} (y_i - f_i(\theta_0))^2 \left( a_{i;jk}(\tilde{\theta}_n) \right)^2 \\
+ \frac{1}{n} \sum_{i=1}^{n} (f_i(\theta_0) - f_i(\tilde{\theta}_n))^2 \left( a_{i;jk}(\tilde{\theta}_n) \right)^2 \\
+ \frac{2}{n} \sum_{i=1}^{n} (y_i - f_i(\theta_0))(f_i(\theta_0) - f_i(\tilde{\theta}_n)) \left( a_{i;jk}(\tilde{\theta}_n) \right)^2 \\
- \frac{2}{n} \sum_{i=1}^{n} h_{i;j}(\tilde{\theta}_n)h_{i;k}(\tilde{\theta}_n)(y_i - f_i(\tilde{\theta}_n))a_{i;jk}(\tilde{\theta}_n).
Then, under Assumption A, Assumption B and application of Lemma 1 and upon repeated applications of the Cauchy-Schwarz inequality, it follows that

$$\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \nabla \phi_{ijk}(\tilde{\theta}_n)^2 < \infty,$$

as required. Q.E.D.

**Lemma 7** Under the conditions of Assumptions A and B,

$$\frac{1}{n} \sum_{i=1}^{n} \phi_i(\hat{\theta}_n)\phi_i(\hat{\theta}_n)^t \rightarrow \sigma^2 \Sigma_0^{-1}.$$

**Proof:**

Rewrite as \(\phi_i(\theta)\) as,

\[
\phi_i(\theta) = -(y_i - f_i(\theta))\nabla f_i(\theta) = -[e_i + f_i(\theta_0) - f_i(\theta)]\nabla f_i(\theta).
\]

Then we have,

\[
\frac{1}{n} \sum_{i=1}^{n} \phi_i(\hat{\theta}_n)\phi_i(\hat{\theta}_n)^t = \frac{1}{n} \sum_{i=1}^{n} e_i^2 \nabla f_i(\hat{\theta}_n)\nabla f_i(\hat{\theta}_n)^t + \frac{1}{n} \sum_{i=1}^{n} [f_i(\theta_0) - f_i(\hat{\theta}_n)]^2 \nabla f_i(\hat{\theta}_n)\nabla f_i(\hat{\theta}_n)^t + \frac{2}{n} \sum_{i=1}^{n} e_i [f_i(\theta_0) - f_i(\hat{\theta}_n)] \nabla f_i(\hat{\theta}_n)\nabla f_i(\hat{\theta}_n)^t = A_1 + A_2 + A_3.
\]

Upon using the Cauchy-Schwarz inequality in conjunction with Assumption B.1, it follows that \(A_2 \rightarrow 0\) a.s. and \(A_3 \rightarrow 0\) a.s., as \(n \rightarrow \infty\). Finally, by decomposing the term \(A_1\) as

\[
A_1 = \frac{1}{n} \sum_{i=1}^{n} (e_i^2 - \sigma^2) \nabla f_i(\hat{\theta}_n)\nabla f_i(\hat{\theta}_n)^t + \frac{\sigma^2}{n} \sum_{i=1}^{n} \nabla f_i(\hat{\theta}_n)\nabla f_i(\hat{\theta}_n)^t = A_{11} + A_{12},
\]

it is easy to see that \(A_{12} \rightarrow \sigma^2 \Sigma_0^{-1}\), by Assumption A.3, and whereas the first term \(A_{11} \rightarrow 0\), by using again the Cauchy-Schwarz inequality in conjunction with Assumption B.1.

Q.E.D.

**Lemma 8** Let \(\hat{\theta}_n\) be the LSE for \(\theta_0\). Under the conditions of Assumptions A and Assumption B we have as \(n \rightarrow \infty\),

\[
-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^t \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} W_i \phi_i(\hat{\theta}_n) \Rightarrow N_p(0, \sigma^2 e^t \Sigma_0 e),
\]

as well as,

\[
-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^t W_i \phi_i(\hat{\theta}_n) \Rightarrow N_p(0, e^t \Sigma_0^{-1} e).
\]
Proof:

Let

\[ a_{ni} = c^t \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} \phi_i(\hat{\theta}_n). \]

Then clearly,

\[ \bar{a}_n := \frac{1}{n} \sum_{i=1}^{n} a_{ni} = c^t \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\hat{\theta}_n) = 0. \]

By Lemma 7 we have,

\[ \frac{1}{n} \sum_{i=1}^{n} \phi_i(\hat{\theta}_n)^t \phi_i(\hat{\theta}_n) \to \sigma^2 \Sigma^{-1}. \]

Thus, by Lemma 5,

\[ \frac{1}{n} \sum_{i=1}^{n} (a_{ni} - \bar{a}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} a_{ni}^2 \]

\[ = c^t \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\hat{\theta}_n) \phi_i(\hat{\theta}_n)^t \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} c \]

\[ \to \sigma^2 c^t \Sigma_0 c. \]

Accordingly, by Lemma 4.6 of Praestgaard and Wellner (1993), combined with Lemma 2 and Lemma 3 below, we obtain the first assertion,

\[ -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} c^t \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} W_i \phi_i(\hat{\theta}_n) \Rightarrow N_p(0, \sigma^2 c^t \Sigma_0 c). \]

As for the second assertion, we let

\[ b_{ni} = c^t \phi_i(\hat{\theta}_n). \]

Then,

\[ \bar{b}_n := \frac{1}{n} \sum_{i=1}^{n} b_{ni} = c \frac{1}{n} \sum_{i=1}^{n} \phi_i(\hat{\theta}_n) = 0. \]

Again by Lemma 7,

\[ \frac{1}{n} \sum_{i=1}^{n} (b_{ni} - \bar{b}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} b_{ni}^2 \]

\[ = c^t \frac{1}{n} \sum_{i=1}^{n} \phi_i(\hat{\theta}_n) \phi_i(\hat{\theta}_n)^t c \to \sigma^2 c^t \Sigma_0^{-1} c. \]

Finally, again by Lemma 4.6 of Praestgaard and Wellner (1993), combined with Lemma 2 and Lemma 3 below, we obtain that,

\[ -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} c^t W_i \phi_i(\hat{\theta}_n) \Rightarrow N_p(0, \sigma^2 c^t \Sigma_0^{-1} c). \]

Q.E.D.
We are now ready to present the proof of the main results stated in Theorem 2.

Proof of Theorem 2: We denote by $S_n(\theta) := \nabla Q_n(\theta)$ and $S^*_n(\theta) := \nabla Q^*_n(\theta)$ the gradients of $Q_n(\theta)$ and $Q^*_n(\theta)$ in (2) and (10), respectively, so that

$$S_n(\theta) = 2 \sum_{i=1}^{n} \phi_i(\theta),$$

and

$$S^*_n(\theta) = 2 \sum_{i=1}^{n} w_{i:n} \phi_i(\theta),$$

where as given in (4),

$$\phi_i(\theta) = -(y_i - f_i(\theta))\nabla f_i(\theta), \quad i = 1 \ldots n.$$  

By the mean-value theorem, there exists a $\lambda_n \in [0, 1]$ such that

$$S^*_n(\hat{\theta}_n) = S^*_n(\tilde{\theta}_n) + \nabla S^*_n(\tilde{\theta}_n) \cdot (\hat{\theta}_n - \tilde{\theta}_n),$$

(12)

where $\hat{\theta}_n = (1 - \lambda_n)\tilde{\theta}_n^* + \lambda_n \tilde{\theta}_n$. Since $\hat{\theta}_n$ and $\tilde{\theta}_n^*$ are such that $S_n(\hat{\theta}_n) = 0$ and $S^*_n(\tilde{\theta}_n^*) = 0$, the equation in (12) may be re-written as

$$\sum_{i=1}^{n} w_{i:n} \phi_i(\hat{\theta}_n) = \sum_{i=1}^{n} w_{i:n} \nabla \phi_i(\hat{\theta}_n) \cdot (\hat{\theta}_n - \tilde{\theta}_n^*),$$

where $\nabla \phi_i(\theta)$ is as defined in (5). Since $W_i \equiv (w_{i:n} - 1)/\tau_n$, we trivially obtain that

$$\tau_n \sum_{i=1}^{n} W_i \left[ \sum_{i=1}^{n} w_{i:n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} \phi_i(\hat{\theta}_n) = \hat{\theta}_n - \tilde{\theta}_n^*.$$  

Thus, we have

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c^t \left[ \frac{1}{n} \sum_{i=1}^{n} w_{i:n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} W_i \phi_i(\hat{\theta}_n) = \tau_n^{-1} \sqrt{n} c^t (\tilde{\theta}_n^* - \hat{\theta}_n),$$

(13)

where by Lemma 5 and Lemma 6

$$\frac{1}{n} \sum_{i=1}^{n} w_{i:n} \nabla \phi_i(\hat{\theta}_n) = \frac{\tau_n}{n} \sum_{i=1}^{n} W_i \nabla \phi_i(\hat{\theta}_n) + \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \mathcal{N} \Sigma_0^{-1}.$$  

(14)

Hence,

$$\tau_n^{-1} \sqrt{n} c^t (\tilde{\theta}_n^* - \hat{\theta}_n) = - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c^t \left[ \frac{1}{n} \sum_{i=1}^{n} w_{i:n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} W_i \phi_i(\hat{\theta}_n)$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c^t \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla \phi_i(\hat{\theta}_n) \right]^{-1} W_i \phi_i(\hat{\theta}_n).$$
By Lemma 5, Lemma 8 and (14),
\[ \tau_n^{-1}\sqrt{n}c^t(\hat{\theta}^*_n - \hat{\theta}_n) \Rightarrow \mathcal{N}_p(0, c^t\Sigma_0 c). \]

Finally, when combined with Assumption A.2, we have that
\[ R^*_n,c \Rightarrow \mathcal{N}(0, \sigma^2), \]
as \( n \to \infty \).

\[ \text{Q.E.D.} \]

References

[1] Bates, D. M. and Watts, D. G., *Nonlinear Regression Analysis and its Applications*, Wiley, New York, 2007.

[2] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist.*, 9, 1196-1217.

[3] Chatterjee, S. and Bose, A. (2005). Generalized bootstrap for estimating equations, *Ann. Statist.*, 33, 414-436.

[4] Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

[5] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife, *Ann. Statist.*, 7, 1-26.

[6] Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap*. New York: Chapman & Hall.

[7] Eicker, F. (1963). Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions, *Ann. Math. Statist.*, 34, 447-456.

[8] Flachaire, E. (2005). Bootstrapping Heteroskedastic Regression Models: Wild Bootstrap vs. Pairs Bootstrap, *CSDA*, 49, 361-476.

[9] Fan, J. and Mei, C. (1991). The convergence rate of randomly weighted approximation for errors of estimated parameters of AR(I) models, *Xian Jiaotong DaXue Xuebao*, 25, 1-6.

[10] Freedman, D. A. (1981). Bootstrapping Regression Models, *Ann. Statist.*, 9, 1218-1228.

[11] Hartigan, J. A. (1969). Using subsample values as typical value, *J. Amer. Statist. Assoc.*, 64, 1303-1317.

[12] Ito, K. and Nisio, M. (1968). On the convergence of sums of independent Banach space valued random variables, *Osaka J. Math.*, 5, 33-48.

[13] Jennrich, I. R. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Statist.*, 40, 633-643.
[14] Lo, A. Y. (1987). A Large Sample Study of the Bayesian Bootstrap, *Ann. Statist.*, **15**, 360-375.

[15] Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function, *Sankhya A*, **53**, 320-333.

[16] Mallet, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika* **73**, 645-656.

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

[18] Mason, D. M. and Newton, M. A. (1992), A Rank Statistics Approach to the Consistency of a General Bootstrap, *Ann. Statist.*, **20**, 1611-1624.

[19] Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *J. Roy. Statist. Soc. Ser. B*, **56**, 3-48.

[20] Praestgaard, J. and Wellner, J. A. (1993). Exchangeably Weighted Bootstraps of the General Empirical Process, *Ann. Probab.*, **21**, 2053-2086.

[21] Quenouille, M. (1949). Approximate tests of correlation in time-series. *Mathematical Proceedings of the Cambridge Philosophical Society*, **45(03)**, 483.

[22] R Core Team (2012). *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL [http://www.R-project.org/](http://www.R-project.org/)

[23] Rao, C. R. and Zhao, L. (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap, *Sankhya A*, **54**, 323-331.

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

[25] Shao, J. and Tu, D.S. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York. Singh, K. (1981). On the asymptotic accuracy of Efron’s bootstrap, *Ann. Statist.*, **9**, 1187-1195.

[26] Singh, K. (1981). On the asymptotic accuracy of Efron’s bootstrap, *Ann. Statist.*, **9**, 1187-1195.

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

[28] Wu, C. F (1981). Asymptotic Theory of Nonlinear Least Squares Estimation, *Ann. Statist.*, **9**, 501-513.

[29] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussions), *Ann. Statist.*, **14**, 1261-1350.

[30] Yu, K. (1988). The random weighting approximation of sample variance estimates with applications to sampling survey, *Chinese J. Appl. Prob. Statist.*, **3**, 340-347.

[31] Zheng, Z. (1987). Random weighting methods, *Acta Math. Appl. Sinica*, **10**, 247-253.

[32] Zheng, Z. and Tu, D. (1988). Random weighting method in regression models. *Sci. Sinica, Ser. A*. 

19