*The Annals of Statistics*
1992, Vol. 20, No. 1, 370–384

# EFFICIENCY AND ROBUSTNESS IN RESAMPLING[1]

BY REGINA Y. LIU AND KESAR SINGH

*Rutgers University*

It is known that the standard delete-1 jackknife and the classical bootstrap are in general equally efficient for estimating the mean-square-error of a statistic in the i.i.d. setting. However, this equivalence no longer holds in the linear regression model. It turns out that the bootstrap is more efficient when error variables are homogeneous and the jackknife is more robust when they are heterogeneous. In fact, we can divide all the commonly used resampling procedures for linear regression models into two types: the E-type (the *efficient* ones like the bootstrap) and the R-type (the *robust* ones like the jackknife). Thus the theory presented here provides a unified view of all the known resampling procedures in linear regression.

**1. Introduction.** The standard delete-1 jackknife and the classical bootstrap are two of the most commonly used resampling procedures for estimating key functionals of sampling distributions such as mean square error, bias or skewness. Quenouille (1956) introduced the jackknife for bias reduction and Tukey (1958) used it to estimate the mean square errors. The bootstrap was introduced in Efron (1979). Some basic properties of the two procedures such as consistency and normality can be found in Miller (1974) for the jackknife and in Bickel and Freedman (1981) and Singh (1981) for the bootstrap.

The bootstrap is computationally more demanding than the jackknife. Due to this, one may think that the bootstrap should be the more efficient procedure in estimating functionals of sampling distributions. Various studies on the bootstrap and the jackknife concerning the consistency under heteroscedasticity of error variables in the linear regression have appeared [see Hinkley (1977), Wu (1986) and Shao and Wu (1987)]. However no systematic comparison has as yet been available. The goal of this paper is to provide a unified framework for classifying all known resampling procedures according to certain aspects of efficiency and robustness. The key tools in our approach are representation theorems, which express estimators as sample means, up to a negligible remainder term. For smooth statistics from i.i.d. samples, it follows from Efron (1979) and Beran (1984) that the bootstrap and the jackknife are essentially equivalent, in the sense that the asymptotic relative efficiency of their m.s.e. (mean square error) estimators is typically 1. This results from the fact that the leading terms in the asymptotic representations of the two estimators are the same. As for robustness in terms of providing consistent results in the case of non i.i.d. setting, the two procedures are

Received October 1989; revised March 1991.
[1]Research partially supported by NSF Grants DMS-88-02558 and DMS-90-04658.
*AMS* 1980 *subject classification*. 62G15.
*Key words and phrases*. Asymptotic variance, bootstrap procedures, jackknife procedures, resampling, asymptotic relative efficiency, E-type, R-type.

known to be roughly equivalent, although the bootstrap is applicable for a larger class of statistics [see Liu (1988) and Liu and Singh (1989)]. This story changes completely in the case of linear models. There the bootstrap and the jackknife are quite different, with relative advantages and disadvantages: the bootstrap is more efficient and the jackknife is more robust. In fact, we are able to classify all the known resampling procedures for linear models into two types: type E (the efficient ones like the bootstrap) and type R (the robust ones like the jackknife). For instance, Wu's external bootstrap [Wu (1986)] and the bootstrap which randomly draws the pairs of responses and covariates [Efron (1979)] are type R. This classification provides a unified view of resampling procedures for linear models. None of the existing procedures seems to have both the properties and it remains an open problem to find one. We believe an adaptive type of resampling procedure should be able to do the job. It should also be interesting to find out how various confidence intervals constructed from different methods reflect their type E or type R characteristics.

We now describe briefly the contents of the paper. Section 2 is devoted to the simple linear model $Y_i = x_i \beta + e_i$, where $Y_i$ is the observation, $e_i$ the random error and $x_i$ the only covariate. Here the intercept is assumed to be 0. Under this simple model, we divide all resampling procedures into two groups —type E and type R. In Section 3 we extend the results to the case of general linear models. Explicit formulas are provided for relative efficiency of type E versus type R in both cases.

The theory developed in this paper can be further extended to the following model:

$$Y_i = \beta X_i + e_i, \qquad i = 1, \dots, n,$$

where $(X_1, Y_1), \dots, (X_n, Y_n)$ are i.i.d. random vectors with $E(e_i|X_i) = 0$ and $\operatorname{Var}(e_i|X_i) = \sigma_i^2$. As a matter of fact, one may think that type R procedures are motivated by this more general model, while type E procedures are motivated by the submodel that $X_i$'s are i.i.d and given $X_i$'s, $e_i$'s are i.i.d. This viewpoint offers additional insight toward understanding the behaviors of the resampling methods discussed in the paper.

**2. Type E and type R resampling in simple linear regression.** We first motivate the representation theorems upon which our classification of the resampling procedures is based. Consider a random sample $Z_1, \dots, Z_n$ from a population $F$ and a statistic $T_n = T(Z_1, \dots, Z_n)$ which is an estimator of the parameter $T_F$. Suppose we are interested in estimating $M_n$, the m.s.e. of $T_n$. The bootstrap estimator of $M_n$ is

$$M_{\mathrm{B}} = E_{\mathrm{B}}(T_{n,\mathrm{B}} - T_n)^2.$$

Here $T_{n,\mathrm{B}}$ is $T_n$ computed from a bootstrap sample of size $n$ drawn from the empirical distribution of $Z_i$'s and $E_{\mathrm{B}}$ is taken w.r.t. the bootstrap resampling. The jackknife estimator of $M_n$ is

$$M_{\mathrm{J}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} (J_i - T_n)^2.$$

Here $J_i$ is the $i$th pseudovalue, that is,

$$J_i = nT_n - (n-1)T_{ni}.$$

Here $T_{ni}$ is computed as $T_n$ with the $i$th observation $Z_i$ deleted. It follows from Efron (1979) and Beran (1984) that for a smooth functional $T_n$, we generally have

$$n(M_{\mathrm{B}} - M_n) = \frac{1}{n}\sum_{i=1}^{n} \xi_F(Z_i) + o_p(n^{-1/2})$$

and

$$n(M_{\mathrm{J}} - M_n) = \frac{1}{n}\sum_{i=1}^{n} \xi_F(Z_i) + o_p(n^{-1/2})$$

for some $\xi_F(\cdot)$ with $E\xi_F(Z_i) = 0$. Consequently,

$$n^{3/2}(M_{\mathrm{B}} - M_n) \to_{\mathscr{L}} N\big(0, E\big(\xi_F^2(Z_i)\big)\big)$$

and

$$n^{3/2}(M_{\mathrm{J}} - M_n) \to_{\mathscr{L}} N\big(0, E\big(\xi_F^2(Z_i)\big)\big).$$

Here $\to_{\mathscr{L}}$ stands for convergence in law and $N(a, b)$ for a normal distribution with mean $a$ and variance $b$. This implies that $M_{\mathrm{B}}$ and $M_{\mathrm{J}}$ are equally efficient as estimators in estimating $M_n$ when the observations are i.i.d.

The main objective of this paper is to study similar representations in linear models for the bootstrap, the jackknife and several other resampling procedures. As it turns out, there are only two types of such representations. This phenomenon leads to the following division of the resampling methods in use: type E—those which have additional efficiency in the case of homogeneous error variables and type R—those which are robust against heteroscedasticity (i.e., stay $\sqrt{n}$-consistent even when the error variables are heterogeneous). In fact, in the homogeneous case the classical bootstrap is not only more efficient as compared to type R but also, as suggested by Beran (1982) and Singh and Babu (1990), optimal in the nonparametric sense. This fact in part justifies the naming of the classical bootstrap as type E. This perhaps can also be viewed as another instance of the following familiar phenomenon pointed out in Stein (1956): An inference procedure which is consistent in a supermodel is usually not asymptotically efficient in a submodel. We now briefly describe all the resampling procedures considered for the simple linear model $Y_i = \beta x_i + e_i$, $i = 1, \dots, n$, where the $e_i$'s are independent random errors each with mean 0 and finite variance $\sigma_i^2$, $x_i$'s are covariates and $Y_i$'s are the responses. The parameter of interest is $\beta$, the slope of the regression line. Its least square estimator is $\hat{\beta} = \sum_{i=1}^n x_i Y_i / L_n$ and the variance of $\hat{\beta}$ is $v_n = \sum_{i=1}^n x_i^2 \sigma_i^2 / L_n^2$, where $L_n = \sum_{i=1}^n x_i^2$. This simple model serves as a learning device and it allows us to extend the results to the general model later.

2.1. *The classical bootstrap* (B) [Efron (1979), Freedman (1981)]. Let $r_i = Y_i - \hat{\beta} x_i$, $i = 1, \dots, n$ be the residuals. One draws an i.i.d. sample of size $n$

from the centered residuals $(r_i - \bar{r}_n)$, $i = 1, \dots, n$, which is denoted as $e_1^*, \dots, e_n^*$. Let $Y_i^* = \hat{\beta} x_i + e_i^*$, $i = 1, \dots, n$ be the bootstrap sample. The bootstrap least square estimator of $\beta$ is $\hat{\beta}_B = \sum_{i=1}^n x_i Y_i^* / L_n$. The bootstrap estimator of $v_n$ is defined as $V_B = E_B(\hat{\beta}_B - \hat{\beta})^2$, where $E_B$ is with respect to the bootstrap probabilities.

2.2. *The standard delete*-1 *jackknife* (J) [Miller (1974), Hinkley (1977)]. As in the case of i.i.d. models, we define

$$J_i = n\hat{\beta} - (n-1)\hat{\beta}_{(i)},$$

where $\hat{\beta}_{(i)}$ is the least square estimator based on $(x_j, Y_j)$, $j \in \{1, \dots, n\} - \{i\}$. The jackknife estimator of $v_n$ is $V_J = (1/n(n-1))\sum_{i=1}^n (J_i - \hat{\beta})^2$.

2.3. *The paired bootstrap* (PB) [Efron (1979)]. One draws an i.i.d. sample of size $n$ from the pairs $(x_i, Y_i)$, $i = 1, \dots, n$. Let $(\tilde{x}_i, \tilde{Y}_i)$, $i = 1, \dots, n$, be the resulting sample. Then the corresponding least square estimator of $\beta$ is $\hat{\beta}_{PB} = \sum_{i=1}^n \tilde{x}_i \tilde{Y}_i / \sum_{i=1}^n \tilde{x}_i^2$ and the estimator of $v_n$ is $V_{PB} = E_{PB}(\hat{\beta}_{PB} - \hat{\beta})^2$. Here $E_{PB}$ is taken with respect to the paired resampling procedure.

2.4. *External bootstrap* (ExB) [Wu (1986)]. Let $H(\cdot)$ be a distribution with mean value 0 and variance 1 and let $H(\cdot)$ be independent of the given regression model. Let $t_i, \dots, t_n$ be a random sample from $H(\cdot)$. Set the bootstrap sample to be $Y_i^* = \hat{\beta} x_i + r_i t_i$, $i = 1, \dots, n$, where $r_i = Y_i - \hat{\beta} x_i$. Then the external bootstrap least square estimator of $\beta$ is $\hat{\beta}_{ExB} = \sum_{i=1}^n x_i Y_i^* / L_n$ and the estimator of $v_n$ is $V_{ExB} = E_H(\hat{\beta}_{ExB} - \hat{\beta})^2$.

Note that this procedure is referred to as wild bootstrap in Härdle and Mammen (1988) and Mammen (1989).

2.5. *Weighted bootstrap* (WB) [Liu (1988)]. The only difference between this procedure and the classical bootstrap is that for WB, one would draw i.i.d. sample of size $n$ from the self-centered weighted residuals $\{\omega_i r_i, i = 1, \dots, n\}$ with $\omega_i = x_i\sqrt{n/L_n}$ instead of from the centered residuals $\{r_1 - \bar{r}_n, \dots, r_n - \bar{r}_n\}$. Let $e_1^*, \dots, e_n^*$ denote the resulting sample. The bootstrap sample is formed in the usual way, that is, $Y_i^* = \hat{\beta} x_i + e_i^*$, $i = 1, \dots, n$, and $\hat{\beta}_{WB}$ and $V_{WB}$ are defined accordingly.

2.6. *Weighted jackknife* (WJ). The pseudovalues $J_i$, $i = 1, \dots, n$ are defined as in the case of the standard jackknife. However, the estimator of $v_n$ is defined to be

$$\frac{L_n}{n^2(n-1)} \sum_{i=1}^n \frac{\left(J_i - \hat{\beta}\right)^2}{x_i^2}.$$

Besides these resampling procedures we later also comment on the weighted jackknife procedures introduced in Hinkley (1977) and Wu (1986), respectively.

Let $c_0$, $c_1$ and $c_2$ stand for positive generic constants. We assume throughout that $\sigma_i^2 \le c_0$ for all $i$, and we define

$$\tilde{v}_n = \sum_{i=1}^{n} \frac{\sigma_i^2}{nL_n}.$$

In Theorems 1 and 2 below, (a) and (b) refer to the following expressions:

$$\text{(a)} \qquad n(V_{n,\mathrm{E}} - \tilde{v}_n) = \sum_{i=1}^{n} \frac{e_i^2 - \sigma_i^2}{L_n} + O_p(n^{-1})$$

and

$$\text{(b)} \qquad n(V_{n,\mathrm{R}} - v_n) = \frac{n}{L_n}\frac{1}{L_n}\sum_{i=1}^{n} x_i^2(e_i^2 - \sigma_i^2) + O_p(n^{-1}),$$

respectively.

THEOREM 1 (Type E). (i) *If* $c_1 \le L_n/n \le c_2$, *then* (a) *holds with* $V_{n,\mathrm{E}} = V_{\mathrm{B}}$.
(ii) *If* $L_n/n \ge c_1$, $\sum_{i=1}^n x_i^4/n \le c_2$ *and* $\max x_i^2 = o(n)$, *then* (a) *holds with* $V_{n,\mathrm{E}} = V_{\mathrm{WJ}}$.

THEOREM 2 (Type R). (i) *If* $\sum_{i=1}^n x_i^6/n \le c_1$, $\max x_i^2 = o(n)$ *and* $L_n/n \ge c_2$, *then* (b) *holds with* $V_{n,\mathrm{R}} = V_{\mathrm{J}}$.
(ii) *If* $c_1 \le |x_i| \le c_2$ *for all* $i$, *then* (b) *holds with* $V_{n,\mathrm{R}} = V_{\mathrm{PB}}$.
(iii) *If* $c_1 \le L_n/n \le c_2$, *then* (b) *holds with* $V_{n,\mathrm{R}} = V_{\mathrm{ExB}}$.
(iv) *If* $c_1 \le L_n/n \le c_2$, *then* (b) *holds with* $V_{n,\mathrm{R}} = V_{\mathrm{WB}}$.

Before we provide the proofs of the theorems, we pause to comment on their implications. In the homogeneous case, that is, when the $e_i$'s have the same distribution and thus $\sigma_i^2 = \sigma^2$ for all $i$,

$$\sup_x \left| P\left( \frac{\sqrt{n}}{L_n} \sum_{i=1}^{n} (e_i^2 - \sigma^2) \le x \right) - \Phi\left( \frac{x}{a_n} \right) \right| \to 0,$$

where $a_n^2 = (Ee_1^4 - \sigma^4)/(L_n/n)^2$, provided that $Ee_1^4 < \infty$ and $c_1 \le L_n/n \le c_2$. Thus $a_n^2$ is the asymptotic variance (a.v.) for the E-type variance estimators. The corresponding result for the R-type variance estimates is

$$\sup_x \left| P\left( \frac{n}{L_n^2} \sqrt{n} \sum_{i=1}^{n} x_i^2 (e_i^2 - \sigma^2) \le x \right) - \Phi\left( \frac{x}{b_n} \right) \right| \to 0,$$

where $b_n^2 = (L_n/n)^{-4}[\sum_{i=1}^n x_i^4/n][Ee_i^4 - \sigma^4]$. For this central limit theorem (CLT), one needs the conditions that $Ee_1^{4+\delta} < \infty$, $|x_i| \le c_1$, for all $i$, and $L_n/n \ge c_2$ (cf. Lindeberg and Feller CLT for non-i.i.d. cases). Note that even if the $e_i$'s are identically distributed, the $x_i^2 e_i^2$'s are not. Now we obtain the following asymptotic relative efficiency (A.R.E.) in the homogeneous case:

$$\text{A.R.E.} = \frac{\text{a.v. of R-Type}}{\text{a.v. of E-Type}} = \frac{b_n^2}{a_n^2} = \frac{\sum_{i=1}^n x_i^4/n}{(L_n/n)^2}, \tag{2.1}$$

which is greater than or equal to 1 by the Cauchy–Schwarz inequality. Thus

$$\text{A.R.E.} - 1 = \frac{1}{n}\sum_{i=1}^n \frac{\left(x_i^2 - L_n/n\right)^2}{(L_n/n)^2}, \tag{2.2}$$

which is nothing but the *squared coefficient of variation of the* $x_i^2$'s. We call this quantity the *deficiency factor* (of R-type compared to E-type). This deficiency factor tends to 0 (assuming $L_n/n \geq c_1$) if and only if the variance of the $x_i^2$'s tends to 0 as $n \to \infty$. In the heterogeneous case, $v_n$ (the true variance) is typically different from $\tilde{v}_n$ (the limit of the E-type variance estimators), thus E-type estimators are inconsistent and R-type estimators are actually $\sqrt{n}$-consistent and have the asymptotic normal distribution with variance $(L_n/n)^{-4}[\sum_{i=1}^n x_i^4(Ee_i^4 - \sigma_i^4)/n]$. In fact, the consistency of the standard delete-1 (or more generally delete-$d$) jackknife estimate under heteroscedasticity is known; see, for example, Hinkley (1977) and Shao and Wu (1987).

PROOF OF THEOREM 1. (i) Clearly $V_B = L_n^{-1}(1/n)\sum_{i=1}^n(r_i - \bar{r}_n)^2 = L_n^{-1}[(1/n)\sum_{i=1}^n r_i^2 - \bar{r}_n^2]$. Note that $r_i = e_i - (\hat{\beta} - \beta)x_i$. Under the assumed conditions, we have $\hat{\beta} - \beta = O_p(n^{-1/2})$ and $\bar{r}_n = O_p(n^{-1/2})$. Thus the result follows.

(ii) Following the definition of the pseudovalue $J_i$, we have

$$\begin{aligned} J_i &= n\hat{\beta} - (n-1)\hat{\beta}_{(i)} \\ &= \beta + n\frac{\sum_{j=1}^n x_j e_j}{L_n} - (n-1)\frac{\sum_{j\neq i} x_j e_j}{\sum_{j\neq i} x_j^2} \\ &= \beta + \frac{\sum_{j=1}^n x_j e_j}{L_n} + \frac{n}{L_n}x_i e_i - \frac{x_i e_i}{L_n} + \frac{n}{L_n^2}(1 + o(1))x_i^3 e_i \\ &\quad - \frac{n}{L_n^2}(1 + o(1))\left(\sum_{j=1}^n x_j e_j\right)x_i^2, \end{aligned} \tag{2.3}$$

provided that $\max x_i^2 = o(L_n)$. Thus

$$\frac{J_i - \hat{\beta}}{x_i} = \frac{ne_i}{L_n} - \frac{e_i}{L_n} + \frac{n}{L_n^2}(1 + o(1))x_i^2 e_i - \frac{n}{L_n^2}(1 + o(1))\left(\sum_{j=1}^n x_j e_j\right)x_i.$$

The representation therefore follows. □

PROOF OF THEOREM 2. (i) The result can be obtained immediately from (2.3).

(ii) Since $\hat{\beta}_{\mathrm{PB}} = \Sigma_{i=1}^{n}\tilde{x}_i\tilde{Y}_i/\Sigma_{i=1}^{n}\tilde{x}_i^2$, it can be shown that

$$\hat{\beta}_{\mathrm{PB}} - \hat{\beta} = \frac{\Sigma_{i=1}^{n}\tilde{x}_i\tilde{e}_i}{\Sigma_{i=1}^{n}\tilde{x}_i^2} - \frac{\Sigma_{i=1}^{n}x_i e_i}{\Sigma_{i=1}^{n}x_i^2},$$

where $\tilde{e}_i = \tilde{Y}_i - \beta\tilde{x}_i$. Moreover,

$$\begin{aligned}\hat{\beta}_{\mathrm{PB}} - \hat{\beta} &= \frac{n}{L_n}\left[\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{e}_i - \frac{1}{n}\sum_{i=1}^{n}x_i e_i\right] + \sum_{i=1}^{n}\tilde{x}_i\tilde{e}_i\left[\frac{1}{\Sigma_{i=1}^{n}\tilde{x}_i^2} - \frac{1}{L_n}\right]\\ &= C_n + D_n, \quad \text{say.}\end{aligned}$$

Since $\tilde{x}_i\tilde{e}_i$, $i = 1,\dots,n$ can be regarded as a random sample from the empirical population of $x_i e_i$, $i = 1,\dots,n$. It follows that

$$E_{\mathrm{PB}}C_n^2 = \frac{n^2}{L_n^2}\frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\left(x_i e_i - \frac{1}{n}\sum_{j=1}^{n}x_j e_j\right)^2,$$

where $E_{\mathrm{PB}}$ is the expectation taken with respect to the paired bootstrap resampling procedure. Thus

$$n\left[E_{\mathrm{PB}}C_n^2 - v_n\right] = \frac{n}{L_n}\frac{1}{L_n}\sum_{i=1}^{n}x_i^2(e_i^2 - \sigma_i^2) + O_p(n^{-1}).$$

It remains to show that $E_{\mathrm{PB}}D_n^2 = O_p(n^{-2})$ and $E_{\mathrm{PB}}C_n D_n = O_p(n^{-2})$. We write

$$D_n = \frac{\Sigma_{i=1}^{n}\tilde{x}_i\tilde{e}_i}{L_n\Sigma_{i=1}^{n}\tilde{x}_i^2}\left[\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}\tilde{x}_i^2\right]. \tag{2.4}$$

Using the Cauchy–Schwarz inequality and the condition $|x_i| \ge c_1$, we arrive at

$$E_{\mathrm{PB}}D_n^2 \le \frac{c}{n^4}\left\{E_{\mathrm{PB}}\left(\sum_{i=1}^{n}\tilde{x}_i\tilde{e}_i\right)^4 E_{\mathrm{PB}}\left[\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}\tilde{x}_i^2\right]^4\right\}^{1/2} = O_p(n^{-2}),$$

since $E_{\mathrm{PB}}(\tilde{x}_i\tilde{e}_i) = (1/n)\Sigma_{i=1}^{n}x_i e_i = O_p(n^{-1/2})$. As for $E_{\mathrm{PB}}(C_n D_n)$, we first write (2.4) as

$$\begin{aligned}&\frac{\Sigma_{i=1}^{n}\tilde{x}_i\tilde{e}_i}{L_n^2}\left[\sum_{i=1}^{n}\tilde{x}_i^2 - \sum_{i=1}^{n}x_i^2\right]\\ &\quad + \frac{1}{L_n}\left[\frac{1}{\Sigma_{i=1}^{n}\tilde{x}_i^2} - \frac{1}{\Sigma_{i=1}^{n}x_i^2}\right]\left[\sum_{i=1}^{n}\tilde{x}_i\tilde{e}_i\right]\left[\sum_{i=1}^{n}x_i^2 - \sum_{i=1}^{n}\tilde{x}_i^2\right] = \tilde{C}_n + \tilde{D}_n,\end{aligned}$$

say. Direct computations show that $E(C_n\tilde{C}_n) = O_p(n^{-2})$. Finally using Hölder's inequality, we obtain

$$\begin{aligned}\left|E\left(C_n\tilde{D}_n\right)\right| &\le \frac{c}{n^4}E\left(\sum_{i=1}^n \tilde{x}_i^2 - \sum_{i=1}^n x_i^2\right)^2\left(\sum_{i=1}^n \tilde{x}_i\tilde{e}_i\right)\left(\sum_{i=1}^n \tilde{x}_i\tilde{e}_i - \sum_{i=1}^n x_ie_i\right)\\ &\le \frac{c}{n^4}\left[E\left(\sum_{i=1}^n \tilde{x}_i^2 - \sum_{i=1}^n x_i^2\right)^4\right]^{1/2}\left[E\left(\sum_{i=1}^n \tilde{x}_i\tilde{e}_i\right)^4\right]^{1/4}\\ &\quad\times\left[E\left(\sum_{i=1}^n \tilde{x}_i\tilde{e}_i - \sum_{i=1}^n x_ie_i\right)^4\right]^{1/4}\\ &= O_p(n^{-2}),\end{aligned}$$

which implies the desired result.

(iii) The result follows from noting that $V_{\text{ExB}} = \sum_{i=1}^n x_i^2 r_i^2/L_n^2$ (in analogy with the formula $v_n = \sum_{i=1}^n x_i^2\sigma_i^2/L_n^2$ in the heterogeneous case).

(iv) Note that $V_{\text{WB}} = \tilde{S}_n^2/L_n$, where $\tilde{S}_n^2$ is the variance of the empirical population $\{x_ir_i/(L_n/n)^{1/2}\}$, $i = 1,\dots,n$. Thus

$$V_{\text{WB}} = \frac{n}{L_n^2}\frac{1}{n}\sum_{i=1}^n\left(x_ir_i - \frac{1}{n}\sum_{i=1}^n x_ir_i\right)^2$$

and a calculation similar to the proof of (i) of Theorem 1 concludes the proof. □

Two modified jackknife procedures proposed in Hinkley (1977) and Wu (1986) (the so-called weighted jackknife), respectively, are also R-type resampling. As a matter of fact, the effects of the modifications occur only at the second order level, namely $O_p(n^{-1})$.

**3. Extension to general linear regression.** We consider here the extension of representations given in the previous section to the following general linear regression model:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}, \tag{3.1}$$

where the response $\mathbf{Y}$ and the error $\mathbf{e}$ are $n \times 1$ vectors, the parameter $\beta$ is a $k \times 1$ vector and the design $\mathbf{X}$ is a $n \times k$ matrix. On the $i$th element $e_i$ of $\mathbf{e}$, we assume that $E(e_i) = 0$ and $\sigma_i^2 = \text{Var}(e_i) \le c$ for some positive constant $c$ and $e_i$'s are independent.

We first present in detail the case of the classical bootstrap in the E-type and Wu's external bootstrap in the R-type. Outlines of other resampling procedures are also given. First we need some definitions.

Definition. A $k \times k$ matrix sequence $\mathbf{A}_n$ is said to be of order $O_p(b_n)$, denoted by $\mathbf{A}_n = O_p(b_n)$, if each component of $\mathbf{A}_n$ is of order $O_p(b_n)$.

Note that the above definition is equivalent to the one that $\mathbf{l}'\mathbf{A}_n\mathbf{m}$ is $O_p(b_n)$ for any fixed vectors $\mathbf{l}$ and $\mathbf{m}$. It is this latter definition that we use repeatedly in the proofs. For general linear models, all resampling procedures extend naturally, so we omit their description. Let $\mathbf{L}_n = \mathbf{X}'\mathbf{X}$. The least square estimator of $\beta$ is $\hat{\beta} = \mathbf{L}_n^{-1}\mathbf{X}'\mathbf{Y}$. The dispersion matrix of $\hat{\beta}$ is $\mathbf{v}_n = \mathbf{L}_n^{-1}\mathbf{X}'[\mathrm{diag}(\sigma_i^2)]\mathbf{X}\mathbf{L}_n^{-1}$, where $[\mathrm{diag}(c_i)]$ is the $n \times n$ diagonal matrix whose $i$th diagonal element is $c_i$. The counterpart of $\tilde{v}_n$ is now $\tilde{\mathbf{v}}_n = (n^{-1}\sum_{i=1}^n \sigma_i^2)\mathbf{L}_n^{-1}$. We assume throughout this section that

(3.2) all eigenvalues of $\mathbf{L}_n/n$ are inside $[m, M]$, where $0 < m \le M < \infty$.

In particular, condition (3.2) is satisfied if the matrix sequence $\mathbf{L}_n/n$ converges to a fixed positive definite matrix. This is in fact a reasonable and commonly used condition, since $\mathbf{L}_n/n$ is always nonnegative definite.

THEOREM 3 (Classical bootstrap as the E-type). *Let $\mathbf{V}_{\mathrm{B}}$ be the estimator of $\mathbf{v}_n$ based on the classical bootstrap. Then*

$$n(\mathbf{V}_{\mathrm{B}} - \tilde{\mathbf{v}}_n) = \sum_{i=1}^{n} (e_i^2 - \sigma_i^2)\mathbf{L}_n^{-1} + O_p(n^{-1}).$$

THEOREM 4 (External bootstrap as the R-type). *Let $\sum_{i=1}^n \|\mathbf{x}_i\|^6 = O(n)$, where $\|\mathbf{x}_i\|$ is the Euclidian norm of the $i$th row $\mathbf{x}_i'$ of $\mathbf{X}$. Then*

$$n(\mathbf{V}_{\mathrm{ExB}} - \mathbf{v}_n) = n\mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}\left(e_i^2 - \sigma_i^2\right)\right]\mathbf{X}\mathbf{L}_n^{-1} + O_p(n^{-1}).$$

Before giving the proofs of Theorems 3 and 4, we would like to discuss some of their applications. First note that $\tilde{\mathbf{v}}_n$ and $\mathbf{v}_n$ are identical only when the error variables are homogeneous. This immediately implies that the R-type procedures produce consistent estimators of the dispersion matrix of $\hat{\beta}$ in the heterogeneous case, which is not achieved by the E-type procedures.

In practice, the parameter of interest $\theta$ is often a linear combination of the elements of the vector $\beta$, say $\theta = \sum_{i=1}^k a_i\beta_i$, where $\beta_i$ is the $i$th element of $\beta$. Important special cases are the contrasts in $\beta_i$'s. A natural estimator is $\hat{\theta} = \sum_{i=1}^k a_i\hat{\beta}_i$, where $\hat{\beta}_i$ is the $i$th element of $\hat{\beta}$. Thus the classical bootstrap estimate of $\mathrm{Var}(\hat{\theta})$ is $\mathbf{a}'\mathbf{V}_{\mathrm{B}}\mathbf{a}$ while the external bootstrap estimate of $\mathrm{Var}(\hat{\theta})$ is $\mathbf{a}'\mathbf{V}_{\mathrm{ExB}}\mathbf{a}$, where $\mathbf{a}' = (a_1, \ldots, a_k)$. As a direct consequence of Theorems 3 and 4, we obtain

$$n(\mathbf{a}'\mathbf{V}_{\mathrm{B}}\mathbf{a} - \mathbf{a}'\tilde{\mathbf{v}}_n\mathbf{a}) = \frac{1}{n}\sum_{i=1}^{n} (e_i^2 - \sigma_i^2)\left[\mathbf{a}'\left(\frac{\mathbf{L}_n}{n}\right)^{-1}\mathbf{a}\right] + O_p(n^{-1})$$

and

$$n(\mathbf{a}'\mathbf{V}_{\mathrm{ExB}}\mathbf{a} - \mathbf{a}'\mathbf{v}_n\mathbf{a}) = n\mathbf{a}'\mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}(e_i^2 - \sigma_i^2)\right]\mathbf{X}\mathbf{L}_n^{-1}\mathbf{a}$$
$$= \frac{1}{n}\sum_{i=1}^{n} b_i^2(e_i^2 - \sigma_i^2) + O_p(n^{-1}),$$

where $b_i = n\mathbf{a}'\mathbf{L}_n^{-1}\mathbf{x}_i$. Note that $\sum_{i=1}^n b_i^2 = n^2\mathbf{a}'\mathbf{L}_n^{-1}\mathbf{a}$, since $\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i' = L_n$. Consequently, the asymptotic relative efficiency of type E compared to type R is

$$\text{A.R.E.} = n\sum_{i=1}^{n} b_i^4 \Bigg/ \left(\sum_{i=1}^{n} b_i^2\right)^2.$$

Thus A.R.E. − 1 = *squared coefficient of variation* of the $b_i^2$'s. This is indeed the counterpart of the deficiency factor (2.2) given in the simple linear regression case.

The above discussion on relative efficiency of E-type versus R-type resampling is not limited to linear combinations of $\beta_i$'s only. In fact the same reasoning applies to any smooth function $g(\beta)$ whose estimator is $g(\hat{\beta})$. One can estimate the variance of $g(\hat{\beta})$ by the classical bootstrap through two expressions: (i) $E_{\mathrm{B}}[g(\hat{\beta}_B) - g(\hat{\beta})]$ or (ii) $\nabla_{\hat{\beta}}'\mathbf{V}_B\nabla_{\hat{\beta}}$, where $\nabla_{\hat{\beta}}$ is the vector of partial derivatives of $g(\cdot)$ at $\hat{\beta}$. For both methods the asymptotic variance (a.v.) of the variance estimator is $\nabla_\beta\mathbf{L}_n^{-1}\nabla_\beta(\sum_{i=1}^n\sigma_i^2/n)$. Similar estimators can be given by the external bootstrap and its a.v. is $\nabla_\beta'\mathbf{L}_n^{-1}\mathbf{X}'[\mathrm{diag}(\sigma_i^2)]\mathbf{X}\mathbf{L}_n^{-1}\nabla_\beta$. Thus the comparison of efficiencies in the previous discussion holds for $g(\beta)$ as well.

Is there a multivariate version of this efficiency comparison between E-type and R-type procedures? Suppose the parameter of interest is $g(\beta) = [g_1(\beta), g_2(\beta)]$ whose estimate is $g(\hat{\beta})$. Let $\mathbf{D}_n$ denote the $3 \times 1$ vector consisting of the components of the dispersion matrix of $g(\hat{\beta})$. Let $\hat{\mathbf{D}}_{\mathrm{B}}$ and $\hat{\mathbf{D}}_{\mathrm{ExB}}$ be the estimates of $\mathbf{D}_n$ based on the classical bootstrap and the external bootstrap, respectively. One can use Theorems 3 and 4 to conclude that in the homogeneous case $\mathrm{ad}(\hat{\mathbf{D}}_{\mathrm{ExB}}) - \mathrm{ad}(\hat{\mathbf{D}}_{\mathrm{B}})$ is nonnegative definite. Here $\mathrm{ad}(\cdot)$ denotes the asymptotic dispersion of $\sqrt{n}(\cdot - \mathbf{D}_n)$. In the final analysis, this amounts to comparing two weighted means—one with equal weights and the other unequal. Notably this multivariate comparison holds when $g(\beta) = (\beta_1, \dots, \beta_k)$ in which case $\mathbf{D}_n$ is a $k(k+1)/2 \times 1$ vector and $\mathrm{ad}(\mathbf{D}_n)$ is a $[(k(k+1)/2)] \times [(k(k+1)/2)]$ matrix.

The comparison of the performances of R-type and E-type variance estimators in the homogeneous case is best summarized by the following result on the difference of their asymptotic dispersion matrices.

THEOREM 5. *If we regard $\mathbf{v}_n$, $\mathbf{V}_{\mathrm{E}}$ and $\mathbf{V}_{\mathrm{R}}$ as vectors of length $[k(k+1)/2]$, where $\mathbf{V}_{\mathrm{E}}$ and $\mathbf{V}_{\mathrm{R}}$, respectively, stand for the* E*-type and* R*-type estimators of $\mathbf{v}_n$, then*

$$\mathrm{ad}(\mathbf{V}_{\mathrm{R}}) - \mathrm{ad}(\mathbf{V}_{\mathrm{E}}) \quad \textit{is nonnegative definite}.$$

Here ad(·) stands for the $[k(k+1)/2] \times [k(k+1)/2]$ asymptotic dispersion matrix of ·. The asymptotic dispersion matrix means the dispersion matrix of the limiting normal distribution for $\sqrt{n}(\cdot - \mathbf{v}_n)$.

PROOF OF THEOREM 3. Let $Y_i = \mathbf{x}'_i\beta + e_i$, denote the $i$th row of the model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, $i = 1, \dots, n$. Note that $\mathbf{V}_{\mathrm{B}} = \mathbf{L}_n^{-1}[\Sigma(r_i - \bar{r}_n)^2/n]$. Essentially one needs to prove that

$$\sum (r_i - \bar{r}_n)^2/n - \sum e_i^2/n = O_p(n^{-1}).$$

Since $r_i = y_i - \mathbf{x}'_i\hat{\beta} = e_i - \mathbf{x}'_i(\hat{\beta} - \beta)$,

$$\sum r_i^2/n - \sum e_i^2/n = (\hat{\beta} - \beta)'\mathbf{L}_n(\hat{\beta} - \beta)/n + 2(\hat{\beta} - \beta)'\sum \mathbf{x}_i e_i/n.$$

Use (i) + (ii) to denote the above sum. Then (i) is of $O_p(n^{-1})$ using condition (3.2) and Lemma 1 given below. For (ii), we look at the dispersion matrix of $\Sigma\,\mathbf{x}_i e_i/n$, which is $\mathbf{X}'[\mathrm{diag}(\sigma_i^2)]\mathbf{X}/n^2$. It remains to be seen that $\mathbf{X}'[\mathrm{diag}(\sigma_i^2)]\mathbf{X}/n$ has its eigenvalues bounded. For any $k \times 1$ vector $\mathbf{l}$, we have $\mathbf{l}'\{\mathbf{X}'/\sqrt{n}[\mathrm{diag}(\sigma_i^2)]\mathbf{X}/\sqrt{n}\}\mathbf{l} \le c\Sigma_{i=1}^k m_i^2$, where $\mathbf{m}'(= (m_1, \dots, m_k)) = \mathbf{l}'\mathbf{X}'$. Since condition (3.2) implies that $\mathbf{m}'\mathbf{m} \le c_1\mathbf{l}'\mathbf{l}$, the claim is established.

We still have to show that $\bar{r}_n = O_p(n^{1/2})$. Let us write

$$\bar{r}_n = \bar{e}_n - \sum_{i=1}^{n} \mathbf{x}'_i(\hat{\beta} - \beta)/n.$$

Since $\Sigma\|\mathbf{x}_i\|^2 = O(n)$, the result follows by using Lemma 1. □

LEMMA 1. *Under condition* (3.2), $\|(\hat{\beta} - \beta)\| = O_p(n^{-1/2})$.

Since the dispersion matrix of $\hat{\beta}$ is $\mathbf{L}_n^{-1}\mathbf{X}'[\mathrm{diag}(\sigma_i^2)]\mathbf{X}\mathbf{L}_n^{-1}$, the variance of $\mathbf{l}'\hat{\beta}$ can be written as $\mathbf{m}'\mathbf{X}'[\mathrm{diag}(\sigma_i^2)]\mathbf{X}\mathbf{m}$, where $\mathbf{m}' = \mathbf{l}'\mathbf{L}_n^{-1}$. Note that $\mathbf{X}'[\mathrm{diag}(\sigma_i^2)]\mathbf{X}/n$ has bounded eigenvalues and $n^2\mathbf{m}'\mathbf{m} = \mathbf{l}'(n\mathbf{L}_n^{-1})(n\mathbf{L}_n^{-1})\mathbf{l} \le c\mathbf{l}'\mathbf{l}$ by condition (3.2). Therefore the variance of $\mathbf{l}'\hat{\beta}$ is of order $O(n^{-1})$.

PROOF OF THEOREM 4. Clearly,

$$\mathbf{V}_{\mathrm{ExB}} = \mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}(r_i^2)\right]\mathbf{X}\mathbf{L}_n^{-1}.$$

Therefore our task is to establish the following:

$$n\mathbf{l}'\mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}(r_i^2 - e_i^2)\right]\mathbf{X}\mathbf{L}_n^{-1}\mathbf{m} = O_p(n^{-1}),$$

where $\mathbf{l}$ and $\mathbf{m}$ are two fixed but arbitrary $k \times 1$ vectors. Since $r_i^2 = e_i^2 + [\mathbf{x}'_i(\hat{\beta} - \beta)]^2 - 2e_i\mathbf{x}'_i(\hat{\beta} - \beta)$, it suffices to prove

$$\text{(3.3)} \qquad \text{(i)} = n\mathbf{l}'\mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}\left(\left(\mathbf{x}'_i(\hat{\beta} - \beta)\right)^2\right)\right]\mathbf{X}\mathbf{L}_n^{-1}\mathbf{m} = O_p(n^{-1})$$

and

$$\text{(3.4)} \qquad \text{(ii)} = n\mathbf{l}'\mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}\left(\left(2e_i\mathbf{x}'_i(\hat{\beta} - \beta)\right)\right)\right]\mathbf{X}\mathbf{L}_n^{-1}\mathbf{m} = O_p(n^{-1}).$$

Note that

$$(3.5) \qquad \text{(i)} = \frac{1}{n} \sum a_i b_i \left[\mathbf{x}_i'(\hat{\beta} - \beta)\right]^2,$$

where $\mathbf{a} = (a_1, \ldots, a_n)' = n\mathbf{l}'\mathbf{L}_n^{-1}\mathbf{X}'$ and $\mathbf{b} = (b_1, \ldots, b_n)' = n\mathbf{m}'\mathbf{L}_n^{-1}\mathbf{X}'$. The Cauchy–Schwarz inequality is then applied to show that

$$(3.6) \qquad a_i^2 \le c_1 \|\mathbf{l}\|^2 \|\mathbf{x}_i\|^2$$

and

$$(3.7) \qquad b_i^2 \le c_2 \|\mathbf{m}\|^2 \|\mathbf{x}_i\|^2,$$

where $c_1$ and $c_2$ are two constants determined by the smallest eigenvalue of $\mathbf{L}_n$. Using (3.6) and applying Cauchy–Schwarz inequality to (3.5), we thus obtain (3.3). As for (ii), we rewrite it as

$$\begin{aligned}
\text{(ii)} &= n\mathbf{l}'\mathbf{L}_n^{-1}\mathbf{X}'\left[\operatorname{diag}\left(\left(e_i \mathbf{x}_i'(\hat{\beta} - \beta)\right)\right)\right]\mathbf{X}\mathbf{L}_n^{-1}\mathbf{m} \\
&= (1/n) \sum a_i b_i e_i \mathbf{x}_i'(\hat{\beta} - \beta) \\
&= (\hat{\beta} - \beta)'\left(\sum a_i b_i e_i \mathbf{x}_i / n\right).
\end{aligned}$$

According to the result of Lemma 1, we only need to prove that $\sum a_i b_i e_i \mathbf{x}_i / n$ is $O_p(n^{-1/2})$. This is shown if we prove that for any $k \times 1$ vector $\mathbf{t}$, $\mathbf{t}'(\sum a_i b_i e_i \mathbf{x}_i / n) = O_p(n^{-1/2})$. To achieve this, we compute the variance of the term on the left and establish that it is of the order $O(n^{-1})$. This result holds under the assumed condition $\sum \|\mathbf{x}_i\|^6 = O(n)$ and (3.2), since

$$n^{-2} \sum a_i b_i (\mathbf{t}'\mathbf{x}_i)^2 \sigma_i^2 \le c \|\mathbf{l}\|^2 \|\mathbf{m}\|^2 \|\mathbf{t}\|^2 n^{-2} \sum \|\mathbf{x}_i\|^6 = O(n^{-1}). \qquad \square$$

For the standard delete-1 jackknife, we state the representation and sketch the idea behind its proof.

THEOREM 6 (Standard delete-1 jackknife as the R-type). *Assume that* $\sum \|\mathbf{x}_i\|^6 \le cn$ *for all* $i = 1, \ldots, n$ *and* $\max \|\mathbf{x}_i\| = o(n)$. *Then*

$$n(\mathbf{V}_{\mathrm{J}} - \mathbf{v}_n) = n\mathbf{L}_n^{-1}\mathbf{X}'\left[\operatorname{diag}\left(e_i^2 - \sigma_i^2\right)\right]\mathbf{X}\mathbf{L}_n^{-1} + O_p(n^{-1}).$$

PROOF. To visualize how one gets the leading term in the representation above, we first consider the pseudovalue $\mathbf{J}_i$'s. Let the regression model after deleting the $i$th row be $\mathbf{Y}_{(i)} = \mathbf{X}_{(i)}\beta + \mathbf{e}_{(i)}$, with $\mathbf{L}_{n(i)} = \mathbf{X}_{(i)}'\mathbf{X}_{(i)}$ and $\hat{\beta}_{(i)} = L_{n(i)}^{-1}\mathbf{X}_{(i)}'\mathbf{Y}_{(i)}$. We would like to remark here that under the assumed conditions, $\mathbf{L}_{n(i)}$ (for each $i$) has all the properties of $\mathbf{L}_n$ after certain $n$ onward. In practice, one could use a $g$-inverse (which coincides with the regular inverse if the latter exists) to ensure the existence of the inverses needed for delete-1 (even delete-$d$) jackknife procedure. Note that

$$\begin{aligned}\mathbf{J}_i - \hat{\beta} &= n\hat{\beta} - (n-1)\hat{\beta}_{(i)} - \hat{\beta}\\ &= n\mathbf{L}_n^{-1}\mathbf{X}\mathbf{e} - (n-1)\mathbf{L}_{n(i)}^{-1}\mathbf{X}_{(i)}\mathbf{e}_{(i)} - (\hat{\beta} - \beta)\\ &\simeq n\left[\mathbf{L}_n^{-1}\mathbf{X}\mathbf{e} - \mathbf{L}_n^{-1}\mathbf{X}_{(i)}\mathbf{e}_{(i)}\right]\\ &= n\mathbf{L}_n^{-1}\mathbf{x}_i e_i.\end{aligned}$$

Thus

$$\mathbf{V}_{\mathrm{J}} = \frac{1}{n(n-1)} \sum (\mathbf{J}_i - \hat{\beta})(\mathbf{J}_i - \hat{\beta})' \simeq \mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}(e_i^2)\right]\mathbf{X}\mathbf{L}_n^{-1},$$

which gives rise to the leading term in the representation.

The rigorous proof of the remainder term to be $O_p(n^{-1})$ hinges on the following simple identity:

$$\mathbf{L}_n^{-1} - \mathbf{L}_{n(i)}^{-1} = -\mathbf{L}_n^{-1}\mathbf{x}_i\mathbf{x}_i'\mathbf{L}_{n(i)}^{-1}. \tag{3.8}$$

This can be shown replacing $\mathbf{x}_i\mathbf{x}_i'$ by $(\mathbf{L}_n - \mathbf{L}_{n(i)})$. The detailed argument is omitted since it is fairly straightforward. □

REMARK 1. The identity (3.8) extends naturally to the case of delete-$d$ jackknife with $\mathbf{x}_i\mathbf{x}_i'$ replaced by the sum of $\mathbf{x}_i\mathbf{x}_i'$ over all deleted rows. This helps verify that the delete-$d$ jackknife is also type R.

Now we turn to the paired bootstrap. Let $(\tilde{\mathbf{x}}_i, \tilde{Y}_i)$, $i = 1, \dots, n$, be a random sample drawn from $(\mathbf{x}_i, Y_i)$, $i = 1, \dots, n$. Let $\tilde{\mathbf{X}}$ be the design matrix from the bootstrap sample, that is, $\tilde{\mathbf{X}}' = (\tilde{x}_1, \dots, \tilde{x}_n)$. We also define $\tilde{\mathbf{L}}_n$ to be $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$. Since $\tilde{\mathbf{L}}_n$ could be singular, one may have difficulty in defining $\hat{\beta}_{\mathrm{PB}}$. It may be interesting to find out if the paired bootstrap has an R-type representation for any choice of $g$-inverse of $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$, and if there is an optimal choice of $g$-inverse for this purpose. To deal with this singularity, we propose a simple truncation technique which preserves the spirit of the paired bootstrap and retains its R-type characteristic. Define

$$\hat{\beta}_{\mathrm{PB}} = \begin{cases} \tilde{\mathbf{L}}_n\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}, & \text{if } (\det \tilde{\mathbf{L}}_n) \geq (\det \mathbf{L}_n)/2,\\ \hat{\beta}, & \text{otherwise,}\end{cases}$$

where det($*$) stands for the determinant of $*$. The constant 1/2 used in the above definition can be any positive number between 0 and 1. Thus $\mathbf{V}_{\mathrm{PB}} = E_{\mathrm{PB}}(\hat{\beta}_{\mathrm{PB}} - \hat{\beta})(\hat{\beta}_{\mathrm{PB}} - \hat{\beta})'$. The expectation is taken with respect to paired bootstrap resampling.

THEOREM 7. *Assume that* $\mathbf{x}_i'\mathbf{x}_i \leq c_0$ *for some constant* $c_0$ *and for all* $i$. *Then*

$$n(\mathbf{V}_{\mathrm{PB}} - \mathbf{v}_n) = n\mathbf{L}_n^{-1}\mathbf{X}'\left[\mathrm{diag}(e_i^2 - \sigma_i^2)\right]\mathbf{X}\mathbf{L}_n^{-1} + O_p(n^{-1}).$$

The key element in the proof of this assertion is a probability bound stated in Lemma 2.

LEMMA 2. *There exist constants* $c_1$ *and* $c_2$ *such that*

$$P_{\mathrm{PB}}\big(\det \tilde{\mathbf{L}}_n \le (\det \mathbf{L}_n)/2\big) \le c_1 \exp(-c_2 n),$$

*where* $P_{\mathrm{PB}}$ *is the paired bootstrap probability and* $c_1$ *and* $c_2$ *depend upon* $c_0$ *only.*

In order to establish this exponential bound, we rewrite the event $(\det \tilde{\mathbf{L}}_n \le (\det \mathbf{L}_n)/2)$ as $(h(\bar{Z}_n) - h(E_{\mathrm{PB}}\bar{Z}_n) \le -d_n)$, where $d_n$ is a sequence of positive numbers bounded away from zero, $\bar{Z}_n$ a multivariate sample mean and $h(\cdot)$ is a smooth real-valued function with all derivatives. Now, using the multivariate mean value theorem, one turns $P_{\mathrm{PB}}(h(\bar{Z}_n) - h(E_{\mathrm{PB}}\bar{Z}_n) \le -d_n)$ into a finite sum of large deviation probabilities. Finally, the inequality is established with the help of an exponential bound similar to Lemma 3.1 of Singh (1981).

The rest of the proof of Theorem 7 is in the same spirit as its counterpart in (ii) of Theorem 2.

Finally, we comment briefly on the possible extension of the weighted jackknife and the weighted bootstrap studied in Section 2. The weighted jackknife and bootstrap studied in Section 2 permit only limited extension. Suppose one is interested in a particular linear combination of $\beta_i$'s such as $\mathbf{l}'\beta = \sum_{i=1}^{k} l_i\beta_i$ [or more generally a smooth function $g(\beta)$]. The corresponding estimator $\mathbf{l}'\hat{\beta}$ can be expressed in the form of $\sum_{i=1}^{n} a_i Y_i$. Thus instead of resampling from the centered residuals $(r_1 - \bar{r}_n, \dots, r_n - \bar{r}_n)$, the weighted bootstrap resamples from the centered weighted residuals $((\omega_1 r_1)^*, \dots, (\omega_n r_n)^*)$, where $(\omega_i r_i)^* = \omega_i r_i - \sum \omega_j r_j/n$. Here $\omega_i = a_i/[\sum a_j^2/n]^{1/2}$. The jackknife pseudovalue $\mathbf{J}_i$ for $\mathbf{l}'\hat{\beta}$ has leading terms $\mathbf{l}'\mathbf{L}_n^{-1}\mathbf{x}_i e_i = b_i e_i$ (say). Then, the weighted jackknife estimator for $\mathrm{Var}(\mathbf{l}'\hat{\beta})$ is

$$\frac{\sum_{i=1}^{n} b_i^2}{n} \frac{1}{n(n-1)} \sum_{i=1}^{n} \frac{\big(\mathbf{J}_i - \mathbf{l}'\hat{\beta}\big)^2}{b_i^2}.$$

REMARK 2. One may wonder if the dichotomy of Sections 2 and 3 is prevalent in other regression models (e.g., nonlinear regression models) and in other estimation procedures (e.g., $L_1$ regression and various robust regression methods such as $M$-estimators). Of course, as far as jackknife is concerned, one would need delete-$d_n$ jackknife with $d_n \to \infty$ to produce a consistent estimator in the case of $L_1$ regression. Does the delete-$d_n$ jackknife with $d_n \to \infty$ retain its R-type characteristic throughout the spectrum $1 \le d_n \le n/2$ or does it switch to the other type at some point? Does the performance of delete-$d$ jackknife improve in any sense (say, second order properties) in some range of $d$ as $d$ increases?

**Acknowledgments.** We would like to thank Rudy Beran, Michael Falk, Cunhui Zhang and the referees for their useful comments and suggestions.

## REFERENCES

BERAN, R. (1982). Estimated sampling distributions: The bootstrap and competitors. *Ann. Statist.* **10** 212–225.

BERAN, R. (1984). Jackknife approximation to bootstrap estimates. *Ann. Statist.* **12** 101–118.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

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

EFRON, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **82** 171–200.

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

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–953.

HÄRDLE, W. and MAMMEN, E. (1988). Comparing nonparametric versus parametric regression fits. Univ. Heidelberg. Unpublished manuscript.

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

HINKLEY, D. V. (1988). Bootstrap methods. *J. Roy. Statist. Soc. Ser. B* **50** 321–337.

LIU, R. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708.

MAMMEN, E. (1989). Bootstrap and wild bootstrap for high-dimensional linear model. Univ. Heidelberg. Unpublished manuscript.

MILLER, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1–15.

QUENOUILLE, M. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SHAO, J. and WU, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.* **15** 1563–1579.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

SINGH, K. and BABU, G. J. (1990). On the asymptotic optimality of the bootstrap. *Scand. J. Statist.* **17** 1–10.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187–196. Univ. California Press, Berkeley.

TUKEY, J. (1958). Bias and confidence in not quite large samples (abstract). *Ann. Math. Statist.* **29** 614.

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

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER, BUSCH CAMPUS
NEW BRUNSWICK, NEW JERSEY 08903