The Annals of Statistics
1993, Vol. 21, No 1, 255–285

BOOTSTRAP AND WILD BOOTSTRAP FOR HIGH DIMENSIONAL LINEAR MODELS

BY ENNO MAMMEN

Universität Heidelberg

In this paper two bootstrap procedures are considered for the estimation of the distribution of linear contrasts and of F-test statistics in high dimensional linear models. An asymptotic approach will be chosen where the dimension p of the model may increase for sample size $n \to \infty$. The range of validity will be compared for the normal approximation and for the bootstrap procedures. Furthermore, it will be argued that the rates of convergence are different for the bootstrap procedures in this asymptotic framework. This is in contrast to the usual asymptotic approach where p is fixed.

1. Introduction. In this paper asymptotic results will be presented for the application of Efron's (1979) bootstrap to least squares estimates in linear models where the design vectors are random and the dimension p of the parameter is large. We consider the linear model

$$Y_{i,n} = X_{i,n}^T \beta_n + \varepsilon_{i,n}, \qquad i = 1, \dots, n. \tag{1.1}$$

In this equation $Y_{i,n}$ are n observations, $X_{i,n}$ are the (observed) random design vectors belonging to $\mathbf{R}^p$, β_n is a p-dimensional parameter and $\varepsilon_{i,n}$ are the (unobservable) error variables. We will use an asymptotic approach where everything may depend on n. Therefore dependence on n will not be indicated in the notation. We write $Y_i = Y_{i,n}$, $X_i = X_{i,n}$, $\beta = \beta_n$ and $\varepsilon_i = \varepsilon_{i,n}$. The stochastic structure is described by

$$\begin{aligned} &(X_i, Y_i) \text{ are i.i.d. with finite second moments } EY_i^2 < \infty \\ &\text{and } E\|X_i\|^2 < \infty, \end{aligned} \tag{1.2}$$

$$\beta \text{ minimizes } b \to E\left(Y_i - X_i^T b\right)^2. \tag{1.3}$$

We assume that $EX_i X_i^T$ is nonsingular. Then β is uniquely defined. Furthermore if one additionally assumes that $E(Y_i - X_i^T\beta)^2 > 0$, then the following conditions are always fulfilled after a standardization of the observations (X_i, Y_i) and of the parameter β:

$$EX_i X_i^T = I_p, \tag{1.4}$$

$$0 < \inf_n E\varepsilon_i^2 \le \sup_n E\varepsilon_i^2 < \infty. \tag{1.5}$$

For instance, to get (1.4) define X_i as $(EX_i X_i^T)^{-1/2} X_i$ and β as $(EX_i X_i^T)^{1/2}\beta$.

Received October 1989; revised February 1992.
AMS 1991 *subject classifications*. Primary 62G09; secondary 62F10, 62F12.
Key words and phrases. Bootstrap, wild bootstrap, linear models, dimension asymptotics.

Then using the new X_i and β, for (1.5) replace Y_i and β by cY_i or $c\beta$, respectively, where $c = [E(Y_i - X_i^T\beta)^2]^{-1/2}$.

Note that (1.2) and (1.3) imply that $EX_i\varepsilon_i = EX_i(Y_i - X_i^T\beta) = 0$ holds (because β minimizes the nonrandom quadratic expression $b \to EY_i^2 - 2b^T[EX_iY_i] + b^T[EX_iX_i^T]b$). But these assumptions do *not* imply that

$$\text{given } (X_1, \dots, X_n) \text{ the } \varepsilon_i, i = 1, \dots, n \text{ are conditionally i.i.d. with mean zero.} \tag{1.6}$$

In this paper we treat the case where the assumption of (1.6) is not appropriate. Take for instance the case where the conditional distributions of the error variables ε_i are heteroscedastic. Conditions (1.2) and (1.3) also do not imply that the linear model holds in the sense that $E(\varepsilon_i|X_i = x)$ vanishes for every x. The parameter β defines the "nearest" linear model. For instance this model may be appropriate if we are interested in the best approximation of $x \to E(Y_i|X_i = x)$ in a given finite-dimensional linear function space (for instance polynomial regression). Also in the usual setup where one assumes that $E(\varepsilon_i|X_i) = 0$ holds it may be interesting how bootstrap behaves under violation of this assumption. Therefore we will in general not assume that $E(\varepsilon_i|X_i) = 0$ holds.

We are interested in the estimation of the distribution of the least squares estimator $\hat{\beta}$ of β,

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}, \tag{1.7}$$

where $\mathbf{X}$ is the $n \times p$ data matrix with rows X_i^T and $\mathbf{Y}$ is an $n \times 1$ data vector with elements Y_i.

Two bootstrap procedures have been proposed for the model (1.2) and (1.3). The first one is due to Efron (1982).

BOOTSTRAP. One generates a resample $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$ from $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. Then one defines

$$\hat{\beta}^* = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{Y}^* \tag{1.8}$$

and one uses $\mathscr{L}^*(\sqrt{n}(\hat{\beta}^* - \hat{\beta}))$ as an estimate of $\mathscr{L}(\sqrt{n}(\hat{\beta} - \beta))$, where $\mathscr{L}^*(\dots)$ denotes the conditional distribution $\mathscr{L}(\dots|X_1, \dots, X_n, Y_1, \dots, Y_n)$ and where $\mathbf{X}^*$ is the $n \times p$ data matrix with rows X_i^{*T} and $\mathbf{Y}^*$ is an $n \times 1$ data vector with elements $\mathbf{Y}_i^*$.

The consistency of this procedure has been proved in Freedman (1981) for the case of fixed dimension p. A second possibility of bootstrapping has been considered by Wu (1986) [see also Beran (1986)]. This resampling procedure has been proposed for the case that the additional assumption $E(\varepsilon_i|X_i) = 0$ is appropriate. Nevertheless we will study this procedure also for the case that this assumption is violated. It proceeds as follows.

WILD BOOTSTRAP. First one estimates $\mathscr{L}(\varepsilon_i|X_1, \dots, X_n)$ by an arbitrary distribution $\hat{F}_i$ with

$$E(Z|\hat{F}_i) = 0, \tag{1.9}$$

$$E(Z^2|\hat{F}_i) = \hat{\varepsilon}_i^2, \tag{1.10}$$

$$E(Z^3|\hat{F}_i) = \hat{\varepsilon}_i^3. \tag{1.11}$$

Then one generates independent ε_i^W with $\mathscr{L}(\varepsilon_i^W) = \hat{F}_i$ and one lets

$$Y_i^W = X_i^T\hat{\beta} + \varepsilon_i^W, \qquad i = 1, \dots, n. \tag{1.12}$$

$\hat{\beta}^W$ is the least squares estimator based on the resample

$$\hat{\beta}^W = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}^W. \tag{1.13}$$

We call this bootstrap procedure *wild* bootstrap because n(!) different distributions $\mathscr{L}(\varepsilon_i|X_i)$ are estimated by only n observations. The condition (1.11) has been introduced by Liu (1988) and Härdle and Mammen (1990) to improve the rate of convergence of the bootstrap estimate. Consistency of this bootstrap has been proved by Liu (1988) for a special model, who also gives some heuristic arguments that the bootstrap estimate of the distribution of the studentized least squares estimator converges with rate $1/n$. In Härdle and Mammen (1990) an example (goodness-of-fit test of a parametric versus a nonparametric regression curve) is given for which Efron's bootstrap does not work, but where wild bootstrap is consistent. For other applications of wild bootstrap in nonparametric regression see Härdle and Marron (1991), Cao-Abad (1991) and Cao-Abad and Gonzales-Manteiga (1990).

Here we will consider three constructions of $\hat{F}_i$. If one assumes that $\hat{F}_i$ puts mass only on two points, then $\hat{F}_i = \mathscr{L}(Z_i)$, where $Z_i = -(\sqrt{5} - 1)\hat{\varepsilon}_i/2$ with probability $(\sqrt{5} + 1)/(2\sqrt{5})$ and $Z_i = (\sqrt{5} + 1)\hat{\varepsilon}_i/2$ with probability $1 - (\sqrt{5} + 1)/(2\sqrt{5})$. In the other constructions we put $Z_i = \hat{\varepsilon}_i U_i$, where $\mathscr{L}(U_i)$ does not depend on i and where $EU_i = 0$, $EU_i^2 = EU_i^3 = 1$. We consider two choices of $\mathscr{L}(U_i)$. First one may put $U_i = V_i/\sqrt{2} + (V_i^2 - 1)/2$, where the V_i's are independent $N(0, 1)$-distributed variables. Another choice would be to define U_i as $U_i = (\delta_1 + V_{i,1}/\sqrt{2})(\delta_2 + V_{i,2}/\sqrt{2}) - \delta_1\delta_2$, where the $V_{i,j}$'s are independent $N(0, 1)$-distributed variables and where $\delta_1 = (3/4 + \sqrt{17}/12)^{1/2}$ and $\delta_2 = (3/4 - \sqrt{17}/12)^{1/2}$. In the sequel we will always assume that one of these three constructions of $\hat{F}_i$ has been used. In our simulations we have used the last construction. For further constructions of $\hat{F}_i$ see also Liu (1988).

In this paper we will use an asymptotic approach in which the dimension p of the linear model and of the design vectors $X_1, \dots, X_n$ may depend on the sample size n. We will also allow the dimension p to grow with the sample size n. This feature is important because in many applications models are used where the dimension p is not small compared with n. Then an asymptotic approach in which p is fixed is misleading because the high dimensionality of the model is lost asymptotically. Compare, for instance, bootstrap in our model

(1.1) and (1.2) with bootstrap in a linear model with nonrandom design and i.i.d. errors. The differences between bootstrap in these models can be made clear by an asymptotic approach with increasing p. In our model bootstrap is based on resampling from the $(p+1)$-dimensional tuples (X_i, Y_i). For large p bootstrap does not work always satisfactorily here (see the simulations in Section 3) because it has to mimic a complex stochastic structure of high dimensional distributions. The situation is quite different in linear models with nonrandom design and i.i.d. errors. There bootstrap (based on resampling from the residuals, see Section 4) mimics the relatively simple stochastic structure of one-dimensional i.i.d. error variables and for least squares estimates bootstrap works under weaker conditions than those which are necessary for the classical normal approximation. This has been pointed out in an asymptotic approach with increasing p by Bickel and Freedman (1983): The Mallows distance between the bootstrap estimate and the distribution of the least squares estimate converges to 0 if $p^2/n \to 0$. Furthermore for linear contrasts of the least squares estimate bootstrap works if $p/n \to 0$. The results of Bickel and Freedman (1983) have been generalized to M estimates in Mammen (1989) [for simulations see also Mammen (1992)]. Asymptotics where p may increase have been also proposed for linear models in Huber (1981), Shorack (1982), and Portnoy (1984, 1985) and for loglinear models in Haberman (1977a, b), Ehm (1986, 1991), Portnoy (1988) and Sauermann (1989). In this paper we will use this approach to explain the different performance of bootstrap, wild bootstrap and normal approximation. For a comparison of different resampling procedures in linear models with fixed dimension p see also Liu and Singh (1992).

We will make assumptions on (the tails of) the distributions of X_i which ensure that the random $p \times p$ matrix $n^{-1}\sum_{i=1}^n X_i X_i^T$ is invertible with high probability and that $(n^{-1}\sum_{i=1}^n X_i X_i^T)^{-1} \approx (EX_i X_i^T)^{-1}$. The conditions will exclude the interesting case that some observations become very influential compared with the other observations (leverage points). We have not been able to state asymptotic results in an asymptotic framework which admits increasing dimension p and the occurrence of leverage points. We conjecture that this may be involved already for fixed dimension p.

This paper is organized as follows. In Section 2 we will show that for linear contrasts bootstrap and wild bootstrap work for high dimensional models in the whole range of the validity of asymptotic normality. In the third section the accuracy of these bootstrap procedures and of the normal approximation will be compared by using formal Edgeworth expansions and simulations. In Section 4 we will show consistency of bootstrapping an F-test statistic under conditions which in general do not imply the validity of the approximation by an F-distribution. This will also be done for bootstrap in a linear model with nonrandom design and i.i.d. errors. The proofs of the theorems are given in Section 5.

2. Consistency of bootstrap for linear contrasts. In this section the bootstrap of the distribution of linear contrasts $c^T(\hat{\beta} - \beta)$ will be considered.

It will turn out (see Theorem 1) that bootstrap estimates consistently the unconditional law $\mathscr{L}(\sqrt{n}\, c^T(\hat{\beta} - \beta))$ under weaker assumptions on (the dimension of) the model than those which are necessary for wild bootstrap and normal approximations. As in the introduction we denote $\mathscr{L}(\dots|X_1, \dots, X_n, Y_1, \dots, Y_n)$ by $\mathscr{L}^*(\dots)$. d_∞ is the Kolmogorov distance (sup norm between the distribution functions).

THEOREM 1. *Choose* $c \in \mathbf{R}^p$ *with* $\|c\| = 1$. *Consider a linear model* (1.1)–(1.3) *which is standardized such that* (1.4) *and* (1.5) *hold. For a fixed* $\delta > 0$ *assume that*

$$p^{1+\delta}/n \to 0, \tag{2.1}$$

$$\sup_{n \ge 1} \sup_{\|d\|=1} E|d^T X_i|^{4K}\left(1 + \varepsilon_i^2\right) < \infty, \tag{2.2}$$

where K is the smallest integer greater than or equal to $2/\delta$,

$$E\left(c^T X_i\right)^2 \varepsilon_i^2 I\left[\left(c^T X_i\right)^2 \varepsilon_i^2 \ge \gamma n\right] \to 0 \quad \textit{for every fixed } \gamma > 0. \tag{2.3}$$

Then

$$d_\infty\left(\mathscr{L}^*\left(\sqrt{n}\, c^T\left(\hat{\beta}^* - \hat{\beta}\right)\right), \mathscr{L}\left(\sqrt{n}\, c^T\left(\hat{\beta} - \beta\right)\right)\right) \to 0 \quad (\textit{in probability}), \tag{2.4}$$

if $E(\varepsilon_i|X_i) = 0$ *or if* $\delta \ge 1/3$,

$$d_\infty\left(\mathscr{L}^*\left(\sqrt{n}\, c^T\left(\hat{\beta}^W - \hat{\beta}\right)\right), \mathscr{L}\left(\sqrt{n}\, c^T\left(\hat{\beta} - \beta\right)\right)\right) \to 0 \quad (\textit{in probability}), \tag{2.5}$$

if $E(\varepsilon_i|X_i) = 0$ *or if* $\delta \ge 1$.

Note that for decreasing δ condition (2.1) becomes weaker and condition (2.2) becomes stronger. We conjecture that the moment conditions in (2.2) may be replaced by weaker but more complicated conditions on the distribution tails of the design vectors. For instance for the case that the elements of X_i are i.i.d. the results of Yin, Bai and Krishnaiah (1988) suggest that it suffices that the fourth moments of the elements of X_i are bounded.

Given the conditions (2.1) and (2.2), the Lindeberg condition (2.3) is necessary for the asymptotic normality of $\sqrt{n}\, c^T(\hat{\beta} - E\hat{\beta})$. This follows from Lemma 3 in Section 5. Therefore Theorem 1 shows that, given (2.1) and (2.2), bootstrap and wild bootstrap work in the whole range of the validity of asymptotic normality of $\sqrt{n}\, c^T(\hat{\beta} - E\hat{\beta})$.

If $E(\varepsilon_i|X_i) \ne 0$ in general, the bias $\sqrt{n}\, c^T(E\hat{\beta} - \beta)$ is of order $p/\sqrt{n}$ (see Lemmas 2 and 3 in Section 5), which may tend to infinity if $\delta < 1$. Then for $\sqrt{n}\, c^T(\hat{\beta} - \beta)$ a mean zero normal approximation and wild bootstrap fail whereas bootstrap works as long as $\delta \ge 1/3$ (i.e., $p^4/n^3 \to 0$). Therefore bootstrap works under weaker model assumptions than wild bootstrap (which is natural by the construction of these procedures). However, in Section 3 we will see that this model robustness of bootstrap must be paid for by a slower rate of convergence than wild bootstrap (for the estimation of the studentized estimator) if the additional model assumption $E(\varepsilon_i|X_i) = 0$ holds.

Under the assumptions of Theorem 1 bootstrap and wild bootstrap estimates also consistently the conditional law $\mathscr{L}(\sqrt{n}\, c^T(\hat{\beta} - \beta)|X_1, \dots, X_n)$ if additionally $E(\varepsilon_i|X_i) = 0$ holds or if $\delta \geq 1$ (i.e., $p^2/n \to 0$). This holds because under these assumptions

$$d_\infty\Big(\mathscr{L}\big(\sqrt{n}\, c^T\big(\hat{\beta} - \beta\big)|X_1, \dots, X_n\big), \qquad \mathscr{L}\big(\sqrt{n}\, c^T\big(\hat{\beta} - \beta\big)\big)\Big) \to 0$$

(in probability). To see this, note that the conditional and the unconditional law converge to the same normal limit $N(0, E(c^T X_i)^2 \varepsilon_i^2)$. For the unconditional law this follows from Lemmas 2 and 4 in Section 5. For the conditional law one argues similarly as in the proofs of Lemmas 4* and 6* in Section 5.

If $E(\varepsilon_i|X_i) = 0$, then for the study of the wild bootstrap another approach may be more appropriate than that outlined in Theorem 1 and in its proof. Consider the case that given a sequence of realizations of $(X_1, \dots, X_n)$ the conditional distribution of $c^T(\hat{\beta} - \beta)$ can be approximated by a normal distribution. This would imply the Lindeberg condition. Then the Lindeberg condition holds also for the conditional law of $c^T(\hat{\beta}^W - \hat{\beta})$ (see Lemma 6*) and under weak additional assumptions the same normal approximation works also for the conditional law of $c^T(\hat{\beta}^W - \hat{\beta})$. This would entail that wild bootstrap approximates $\mathscr{L}(\sqrt{n}\, c^T(\hat{\beta} - \beta)|X_1, \dots, X_n)$ consistently nearly as long as a mean zero normal approximation works.

3. Accuracy of the bootstrap. In this section the cases will be studied more closely where bootstrap and normal approximations work. We conjecture that for high dimensional models wild bootstrap is a more accurate estimate of the distribution of $\sqrt{n}\, c^T(\hat{\beta} - \beta)$ than bootstrap if the bias $\sqrt{n}\, c^T(E\hat{\beta} - \beta)$ is not too large. The accuracy of bootstrap and normal approximations will be compared by a formal Edgeworth expansion and by simulations. For deriving the formal Edgeworth expansion we assume that the linear model holds in the sense of $E(\varepsilon_i|X_i) = 0$. Then w.l.o.g. $E\hat{\beta} = \beta$. We will give no proof of the validity of the Edgeworth expansion. We will show the rates of convergence only for the bootstrap estimates of the first four moments of a linear contrast $c^T(\hat{\beta} - \beta)$ and we will use these results for an heuristic study of bootstrap based on the assumption that the formal second order Edgeworth expansion holds for $c^T(\hat{\beta} - \beta)$, $c^T(\hat{\beta}^W - \hat{\beta})$, $c^T(\hat{\beta}^* - \hat{\beta})$ and their studentized counterparts. Then one gets for them a first order Edgeworth expansion with the following error term: $P(Z \leq x) = \Phi(\tilde{x}) + \gamma(Z)(1 - \tilde{x}^2)\varphi(\tilde{x}) + O(\delta(Z))$, where $\tilde{x} = (x - EZ)\mathrm{var}(Z)^{-1/2}$, $\gamma(Z) = (1/6)E\tilde{Z}^3$ and $\delta(Z) = E\tilde{Z}^4 - 3 + \gamma(Z)^2$ with $\tilde{Z} = (Z - EZ)\mathrm{var}(Z)^{-1/2}$. A higher order analysis for wild bootstrap based on Edgeworth expansions can be found in Liu (1988) for a one dimensional linear model and in Cao-Abad (1991) and Cao-Abad and Gonzales-Manteiga (1990) for nonparametric regression. For a simple approach of a higher order analysis of bootstrap which does not use Edgeworth expansions see also Mammen (1990).

For the analysis of bootstrap we consider the moments of the following expansions of $\sqrt{n}\, c^T(\hat{\beta} - \beta)$ and $\sqrt{n}\, c^T(\hat{\beta}^* - \hat{\beta})$:

$$\hat{\theta}_{c,n} = \sum_{h=0}^{2} c^T A^h \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i, \tag{3.1}$$

$$\hat{\theta}^*_{c,n} = \sum_{h=0}^{2} c^T A^{*h} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i^* \varepsilon_i^*, \tag{3.2}$$

where $A = I - (1/n)\Sigma_{i=1}^n X_i X_i^T$ and $A^* = I - (1/n)\Sigma_{i=1}^n X_i^*(X_i^*)^T$.

THEOREM 2. *Choose $c \in \mathbf{R}^p$ with $\|c\| = 1$. Consider a linear model with (1.1)–(1.5). Assume additionally $E(\varepsilon_i | X_i) = 0$ and*

$$p^{1+\delta}/n \to 0 \quad \text{for a fixed } \delta > 0. \tag{3.3}$$

If the following holds for large enough C_1, C_2 and $C_3 > 0$ (depending on δ)

$$\sup_n E\left(\|X_i\|/\sqrt{p}\right)^{C_1} < +\infty, \tag{3.4}$$

$$\sup_n \sup_{\|d\|=1} E\left(d^T X_i\right)^{C_2} < +\infty, \tag{3.5}$$

$$\sup_n E|\varepsilon_i|^{C_3} < +\infty, \tag{3.6}$$

then

$$\hat{\theta}_{c,n} = \sqrt{n}\, c^T\left(\hat{\beta} - \beta\right) + o_P(p/n), \tag{3.7}$$

$$\hat{\theta}^*_{c,n} = \sqrt{n}\, c^T\left(\hat{\beta}^* - \hat{\beta}\right) + o_P(p/n). \tag{3.8}$$

For the moments of the stochastic expansions $\hat{\theta}_{c,n}$ and $\hat{\theta}^*_{c,n}$ one gets

$$E^* \hat{\theta}^*_{c,n} = O_P(p/n), \tag{3.9}$$

$$\mathrm{var}^*\left(\hat{\theta}^*_{c,n}\right) - \mathrm{var}\left(\hat{\theta}_{c,n}\right) = O_P(n^{-1/2} + p/n), \tag{3.10}$$

$$\mu_3^*\left(\hat{\theta}^*_{c,n}\right) - \mu_3\left(\hat{\theta}_{c,n}\right) = O_P(n^{-1} + p/n^{3/2}), \tag{3.11}$$

$$\mu_4^*\left(\hat{\theta}^*_{c,n}\right) - 3\,\mathrm{var}^*\left(\hat{\theta}^*_{c,n}\right)^2 = O_P(n^{-1}), \tag{3.12}$$

$$\mu_4\left(\hat{\theta}_{c,n}\right) - 3\,\mathrm{var}\left(\hat{\theta}_{c,n}\right)^2 = O(n^{-1}). \tag{3.13}$$

Here for $j = 3$ and $j = 4$ the quantities μ_j, μ_j^* and var* *are defined as $\mu_j(Z) = E((Z - E(Z))^j)$, $\mu_j^*(Z) = E^*((Z - E^*(Z))^j)$ and $\mathrm{var}^*(Z) = E^*(Z - E^*(Z))^2$, where $E^*(Z) = E(Z | X_1, \dots, X_n, Y_1, \dots, Y_n)$.*

Theorem 2 can be used for an heuristic study of the accuracy of the bootstrap. Equation (3.10) suggests that the sup norm between the distribution function of $\sqrt{n}\, c^T(\hat{\beta} - \beta)$ and the bootstrap approximation is of order $O_P(n^{-1/2} + p/n)$. This rate of convergence may be very slow for large p. For many cases it has been proposed to bootstrap the studentized estimator to increase the rate of convergence [see, e.g., Beran (1987)]. However, (3.9) sugests that the rate of convergence of the bootstrap approximation of the studentized linear contrast is of order $O_P(p/n)$. Note that here for large p, $p > \sqrt{n}$, studentization does not increase the rate of convergence.

The situation is quite different for wild bootstrap. Conditionally on $(X_1, \dots, X_n, Y_1, \dots, Y_n)$ the estimate $\sqrt{n}\, c^T(\hat{\beta}^W - \hat{\beta})$ is a sum of conditionally independent zero mean variables. This simplifies the calculation of moments (and would also make a rigorous proof of the Edgeworth expansion easier). Because the moments of the statistic $\sqrt{n}\, c^T(\hat{\beta} - \beta)$ may not exist under our assumptions we consider conditional moments given the following event $H = H_\varepsilon = \{$The maximal absolute eigenvalue of $\mathbf{X}^T\mathbf{X}/n - \mathbf{I}$ is bounded by $(p/n)^\varepsilon\}$ for an ε with $0 < \varepsilon < 1/2$.

THEOREM 3. *Choose a fixed ε with $0 < \varepsilon < 1/2$. Then under the assumptions of Theorem 2 with appropriately chosen C_1, C_2 and $C_3 > 0$ one gets for every fixed $\kappa < 0$,*

$$P(H_\varepsilon) = 1 - O(n^{-\kappa}), \tag{3.14}$$

$$\operatorname{var}_H^*\left(n^{1/2}c^T\left(\hat{\beta}^W - \hat{\beta}\right)\right) - \operatorname{var}_H\left(n^{1/2}c^T\left(\hat{\beta} - \beta\right)\right) = O_P(n^{-1/2} + p/n), \tag{3.15}$$

$$\mu_{3,H}^*\left(n^{1/2}c^T\left(\hat{\beta}^W - \hat{\beta}\right)\right) - \mu_{3,H}\left(n^{1/2}c^T\left(\hat{\beta} - \beta\right)\right) = O_P(n^{-1} + pn^{-3/2}), \tag{3.16}$$

$$\mu_{3,H}\left(n^{1/2}c^T\left(\hat{\beta} - \beta\right)\right) = O_P(n^{-1/2}), \tag{3.17}$$

$$\mu_{4,H}^*\left(n^{1/2}c^T\left(\hat{\beta}^W - \hat{\beta}\right)\right) - 3\operatorname{var}_H^*\left(n^{1/2}c^T\left(\hat{\beta}^W - \hat{\beta}\right)\right)^2 = O_P(n^{-1}), \tag{3.18}$$

$$\mu_{4,H}\left(n^{1/2}c^T\left(\hat{\beta} - \beta\right)\right) - 3\operatorname{var}_H\left(n^{1/2}c^T\left(\hat{\beta} - \beta\right)\right)^2 = O_P(n^{-1}). \tag{3.19}$$

Here for $j = 3$ and $j = 4$ the quantities $\mu_{j,H}$, $\mu_{j,H}^*$, var_H and var_H^* are defined as $\mu_{j,H}(Z) = E((Z - E(Z|H))^j|H)$, $\mu_{j,H}^*(Z) = E_H^*((Z - E_H^*(Z))^j)$, $\operatorname{var}_H(Z) = E((Z - E(Z|H))^2|H)$ and $\operatorname{var}_H^*(Z) = E_H^*(Z - E_H^*(Z))^2$, where $E_H^*(Z) = E(Z|H, X_1, \dots, X_n, Y_1, \dots, Y_n)$ and $H = H_\varepsilon$.

Theorem 3 suggests that wild bootstrap of an unstudentized linear contrast has the same rate of convergence as bootstrap. But studentization leads always to an improvement of the rate of convergence of wild bootstrap. Then wild bootstrap produces approximations of order $O_P(n^{-1} + pn^{-3/2})$. For studentization one may use the variance estimate $\hat{\sigma}_c^2 = \operatorname{var}^W(\sqrt{n}\, c^T\hat{\beta}^W) = (1/n)\sum_{i=1}^n (c^T(\mathbf{X}^T\mathbf{X}/n)^{-1}X_i)^2(Y_i - X_i^T\hat{\beta})^2$. This choice of $\hat{\sigma}_c^2$ is asymptotically equivalent to the bootstrap variance estimator (see the proof of Lemma 4* in Section 5). The poor rate of convergence of bootstrap for the studentized linear contrast comes from a bias effect [see (3.9)]. Therefore this rate does not depend on the choice of the variance estimator $\hat{\sigma}_c^2$ (as long as $\hat{\sigma}_c^2$ fulfills certain regularity conditions). Because of (3.15) the accuracy of the normal approximation $N(0, \hat{\sigma}_c^2)$ is $O_P(n^{-1/2} + pn^{-1})$. The different rates of convergence are summarized in Table 1.

TABLE 1
Rates of convergence of the bootstrap procedures and the mean zero normal approximation under the assumption $E(\varepsilon_t|X_t) = 0$

Estimation of	$\mathscr{L}(\sqrt{n}\, c^T(\hat\beta - \beta))$	$\mathscr{L}(\sqrt{n}\, c^T(\hat\beta - \beta) / \hat\sigma_c)$
Normal approximation $N(0, \hat\sigma_c^2)$	$O_P(n^{-1/2} + pn^{-1})$	
Wild bootstrap	$O_P(n^{-1/2} + pn^{-1})$	$O_P(n^{-1} + pn^{-3/2})$
Bootstrap	$O_P(n^{-1/2} + pn^{-1})$	$O_P(pn^{-1})$

We have also compared bootstrap, wild bootstrap and normal approximation in a simulation study. The results are displayed in Figures 2–5. We have considered the following model. The sample size n is 50 and the dimension p is 5. The design vectors are defined as

$$X_{i1} = 1 \quad \text{and} \quad X_{ij} = U_{ij} Z_i / 2, \qquad i = 1, \dots, 50; j = 2, \dots, 5$$

and for different choices of $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ the observations are put as

$$Y_i = \gamma_1 Q_i + \gamma_2 Q_i V_i + \gamma_3 V_i,$$

where $Q_i = (\|X_i\|^2 - 1 - EZ_i^2)$.

The variables $U_{1,2}, \dots, U_{50,5}, V_1, \dots, V_{50}, Z_1, \dots, Z_{50}$ are independent with the following distributions. The $U_{i,j}$'s have a standard normal distribution $N(0,1)$. The distribution of the V_i's is a mixture of normal distributions: $(1/2)N(1/2, (1.2)^2) + (1/2)N(-1/2, (0.7)^2)$. For a plot of the error density see Figure 1. The Z_i's are uniformly distributed on the interval $[1, 3]$. We have

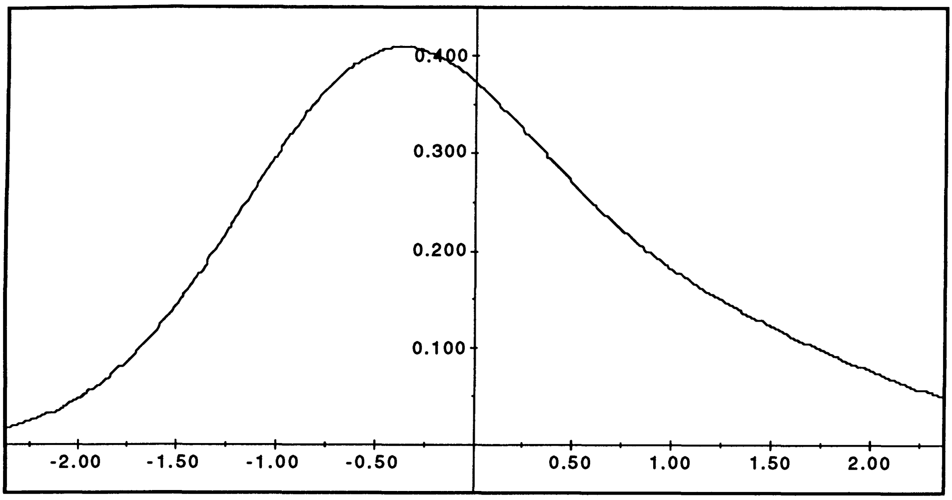

FIG. 1. *Density of the error variables ε_t in the simulations.*

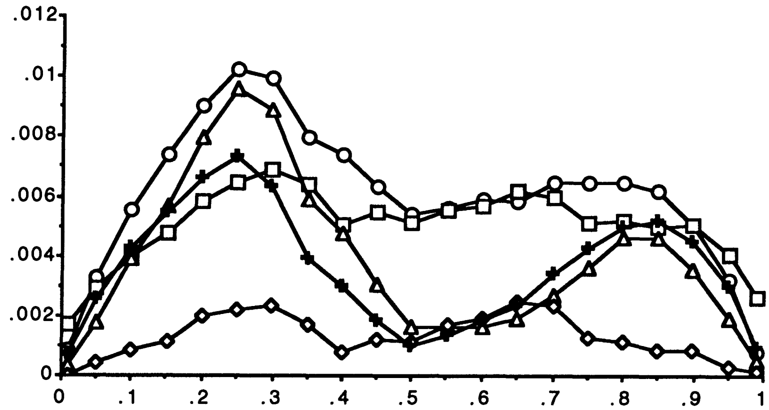

FIG. 2. *Monte Carlo estimates of the expected squared error of estimates of the distribution function of an unstudentized linear contrast and a studentized linear contrast, plotted against the distribution function (Model I, $p = 5$, $c = c_1$): $\triangle$ wild bootstrap estimate; $\circ$ bootstrap estimate; $+$ normal approximation; $\diamond$ wild bootstrap estimate of the studentized linear contrast; $\square$ bootstrap estimate of the studentized linear contrast.*

used the following choices of γ:

$$\gamma = (0, 1, 1) \qquad \text{(Case I)},$$
$$\gamma = (1, 0.25, 1) \qquad \text{(Case II)}.$$

In case I the linear model also holds in the sense of $E(\varepsilon_i|X_i) = 0$ but the error variables ε_i are conditionally heteroschedastic. In case II the conditional expectation $E(\varepsilon_i|X_i)$ does not vanish. We have made 1000 simulations. Every bootstrap resampling and every wild bootstrap resampling uses 1000 replications. We study the estimates for the distribution function of two linear contrasts (studentized and unstudentized) based on bootstrap, wild bootstrap and normal approximation with estimated variance. The following two linear contrasts $c^T(\hat{\beta} - \beta)$ have been simulated: $c = c_1 = (1, 0, \dots, 0)^T$ and $c = c_2 = (0, 1, 0, 0, 0)^T$. In Figures 2–5 the Monte Carlo estimate for the expected squared error of these estimates are plotted against the distribution function of the linear contrasts $c^T(\hat{\beta} - \beta)$ or of the studentized linear contrasts $c^T(\hat{\beta} - \beta)/\hat{\sigma}_c$. In Tables 2–5 the Monte Carlo estimates for the expectation and the standard deviation are given at the 10% and the 90% quantiles.

We draw the following conclusions from our simulations.

1. Consider first the case of the first linear contrast c_1 in case II [$\gamma =$ (1, 0.25, 1)]. In this case $c_1^T(\hat{\beta} - \beta)$ has a large bias: $Ec_1^T(\hat{\beta} - \beta)/(\mathrm{var}(c_1^T\hat{\beta}))^{1/2} = -0.77$. The bootstrap and wild bootstrap estimate (and the mean zero normal approximation) of the distribution function are shifted compared with the real distribution function. The bootstrap estimate gives here the best approximation. This is in accordance with Theorem 1, which shows that bootstrap is more robust against departures of $E(\varepsilon_i|X_i)$ from 0.

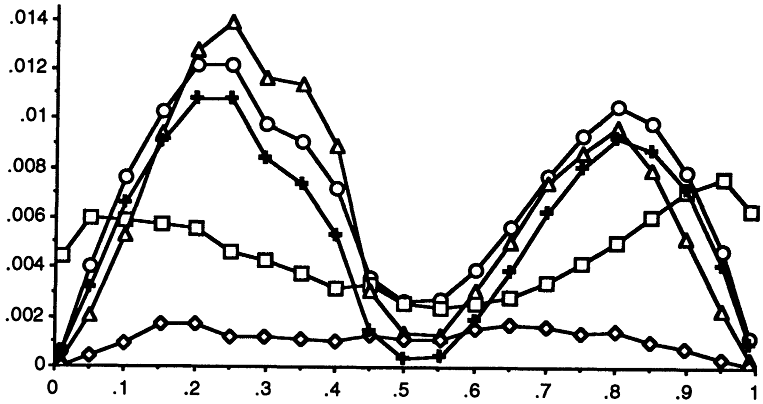

FIG. 3. *Monte Carlo estimates of the expected squared error of estimates of the distribution function of an unstudentized linear contrast and a studentized linear contrast, plotted against the distribution function (Model I, $p = 5$, $c = c_2$): $\triangle$ wild bootstrap estimate; $\circ$ bootstrap estimate; + normal approximation; $\diamond$ wild bootstrap estimate of the studentized linear contrast; $\square$ bootstrap estimate of the studentized linear contrast.*

2. The situation is quite different for the second linear contrast c_2. $c_2^T(\hat{\beta} - \beta)$ has only a small bias $Ec_2^T(\hat{\beta} - \beta)/(\text{var}(c_2^T\hat{\beta}))^{1/2} = -0.009$. Here one has a similar picture as in case I. The location of the real distribution is estimated more accurately by all estimates. The bootstrap estimate has a large variance compared with the wild bootstrap estimate (see Tables 2, 3 and 5). The wild bootstrap estimate of the studentized linear contrast produces here the best approximation.

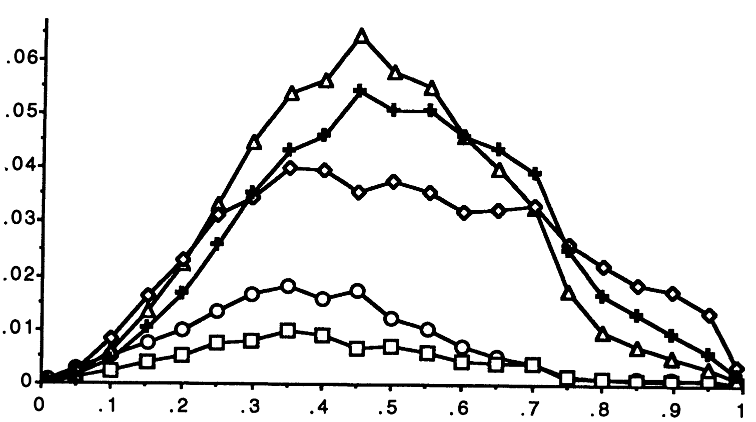

FIG. 4. *Monte Carlo estimates of the expected squared error of estimates of the distribution function of an unstudentized linear contrast and a studentized linear contrast, plotted against the distribution function. (Model II, $p = 5$, $c = c_1$): $\triangle$ wild bootstrap estimate; $\circ$ bootstrap estimate; + normal approximation; $\diamond$ wild bootstrap estimate of the studentized linear contrast, $\square$ bootstrap estimate of the studentized linear contrast.*

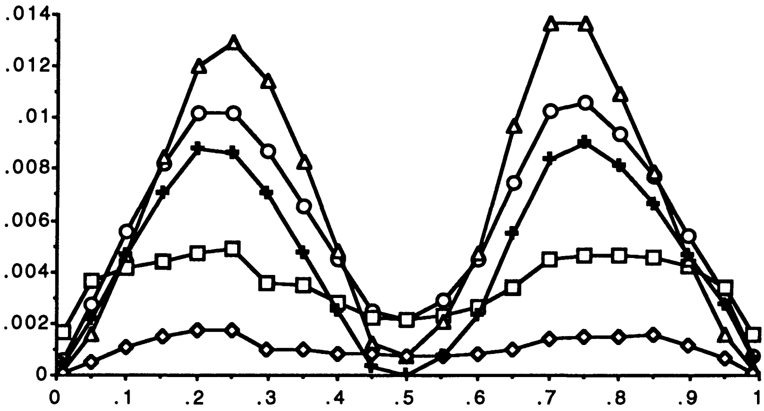

FIG. 5. *Monte Carlo estimates of the expected squared error of estimates of the distribution function of an unstudentized linear contrast and a studentized linear contrast, plotted against the distribution function (Model II, $p = 5$, $c = c_2$): △ wild bootstrap estimate; ○ bootstrap estimate; + normal approximation; ◇ wild bootstrap estimate of the studentized linear contrast; □ bootstrap estimate of the studentized linear contrast.*

TABLE 2
Monte Carlo estimates of the expectation (first row) and standard deviation (second row) of estimates of the distribution function of an unstudentized linear contrast (columns 1 and 3) and a studentized linear contrast (columns 2 and 4) at the 10% and 90% quantiles (Model I, $p = 5$, $c = c_1$)

	10% quantile		90% quantile	
	(1)	(2)	(3)	(4)
Bootstrap	0.083	0.094	0.900	0.866
	0.073	0.063	0.071	0.062
Wild bootstrap	0.068	0.075	0.919	0.877
	0.053	0.014	0.057	0.017
Normal approximation	0.084	—	0.899	—
	0.063	—	0.067	—

TABLE 3
Monte Carlo estimates of the expectation (first row) and standard deviation (second row) of estimates of the distribution function of an unstudentized linear contrast (columns 1 and 3) and a studentized linear contrast (columns 2 and 4) at the 10% and 90% quantiles (Model I, $p = 5$, $c = c_2$)

	10% quantile		90% quantile	
	(1)	(2)	(3)	(4)
Bootstrap	0.080	0.112	0.906	0.870
	0.085	0.076	0.089	0.079
Wild bootstrap	0.060	0.080	0.928	0.896
	0.061	0.024	0.067	0.027
Normal approximation	0.084	—	0.900	—
	0.080	—	0.085	—

TABLE 4
Monte Carlo estimates of the expectation (first row) and standard deviation (second row) of estimates of the distribution function of an unstudentized linear contrast (columns 1 and 3) and a studentized linear contrast (columns 2 and 4) at the 10% and 90% quantiles (Model II, $p = 5$, $c = c_1$)

	10% quantile		90% quantile	
	(1)	(2)	(3)	(4)
Bootstrap	0.089	0.076	0.900	0.876
	0.070	0.041	0.036	0.019
Wild bootstrap	0.028	0.009	0.848	0.769
	0.028	0.003	0.050	0.016
Normal approximation	0.046	—	0.821	—
	0.039	—	0.058	—

3. Consider again case I and the second linear contrast c_2 in case II. The wild bootstrap estimate works *much* better here for the studentized linear contrast than for the unstudentized contrast. This is *not* true for the bootstrap estimate (consider also the tails of the distributions in Figures 2, 3 and 5). Furthermore, in other simulations in Mammen (1992) bootstrap of the unstudentized estimate works slightly better than bootstrap of the studentized estimate. Note that this contradicts the usual higher-order results for fixed dimension p, which say that studentization improves the rate of convergence of bootstrap. However, this is in accordance with our discussion based on Edgeworth expansions, where we have argued that studentization of the linear contrast does not lead to a faster rate of convergence of the bootstrap estimate if the dimension p is large.
4. Note that also in case II for $c = c_2$ wild bootstrap is preferable to bootstrap although in this model the departure from the linear model is drastic: $E(\varepsilon_i|X_i)$ is of larger order than the conditional standard deviation $(\mathrm{var}(\varepsilon_i|X_i))^{1/2}$ of the error variables.

TABLE 5
Monte Carlo estimates of the expectation (first row) and standard deviation (second row) of estimates of the distribution function of an unstudentized linear contrast (columns 1 and 3) and a studentized linear contrast (columns 2 and 4) at the 10% and 90% quantiles (Model II, $p = 5$, $c = c_2$)

	10% quantile		90% quantile	
	(1)	(2)	(3)	(4)
Bootstrap	0.069	0.102	0.928	0.899
	0.068	0.064	0.068	0.065
Wild bootstrap	0.049	0.077	0.949	0.925
	0.044	0.023	0.045	0.023
Normal approximation	0.073	—	0.924	—
	0.063	—	0.064	—

4. Bootstrapping F-tests. Bootstrap may be used to estimate critical values of an F-test statistic. In this section we will show that bootstrap works in high dimensional linear models under weak conditions (see Theorem 3). Note that in our model (1.1)–(1.3) in general the approximation by an F-distribution fails.

For linear subspaces $\mathbf{H}_0 \subset \mathbf{H}_1$ of $\mathbf{R}^p$, $0 \in \mathbf{H}_0$, we consider the testing problem $\beta \in \mathbf{H}_0$ versus $\beta \in \mathbf{H}_1$. The F-test statistic for this testing problem is defined by

$$T = \frac{\|\Pi_1 \mathbf{Y} - \Pi_0 \mathbf{Y}\|^2/(p_1 - p_0)}{\|\mathbf{Y} - \Pi_1 \mathbf{Y}\|^2/(n - p_1)}$$

where (for $i = 1, 2$) $\Pi_i \mathbf{Y}$ is the projection of $\mathbf{Y}$ onto $L_i = \{z | z_j = X_j^T b$ for $b \in \mathbf{H}_i\}$ and p_i is the dimension of L_i. Denote the least squares estimator under the hypotheses $\beta \in \mathbf{H}_i$ by $\hat{\beta}_i$. For the determination of critical values we want to estimate the distribution of T for parameters β in the hypotheses $\mathbf{H}_0$. Furthermore on the alternative $\beta \in \mathbf{H}_1 \setminus \mathbf{H}_0$ the bootstrap estimate of the distribution of T should converge to the distribution of T under a parameter β_0 which lies in the hypothesis $\mathbf{H}_0$. This would guarantee a satisfactory power of T when the bootstrap critical values are used.

In the next theorem we will show that bootstrap and wild bootstrap work after the following slight modifications:

(4.1a) $\{(X_i^*, Y_i^*)\}$ is a resample from $\left\{\left(X_i, Y_i - X_i^T\hat{\beta}_1 + X_i^T\hat{\beta}_0\right)\right\}$.

(4.1b) The conditional variance and the conditional third moment of ε_i^W are $(Y_i - X_i^T\hat{\beta}_1)^2$ and $(Y_i - X_i^T\hat{\beta}_1)^3$ (respectively). The Y_i^W are defined as $Y_i^W = X_i^T\hat{\beta}_0 + \varepsilon_i^W$.

Using these bootstrap observations one can construct test statistics

$$T^* = \frac{\|(\Pi_1 - \Pi_0)(\mathbf{Y}^*)\|^2/(p_1 - p_0)}{\|(I - \Pi_1)(\mathbf{Y}^*)\|^2/(n - p_1)} \quad \text{and}$$

$$T^W = \frac{\|(\Pi_1 - \Pi_0)(\mathbf{Y}^W)\|^2/(p_1 - p_0)}{\|(I - \Pi_1)(\mathbf{Y}^W)\|^2/(n - p_1)}.$$

THEOREM 4. *Assume* (1.1)–(1.5) *for* $\beta \in \mathbf{H}_1$, *with* $E(\varepsilon_i | X_i) = 0$. *Assume additionally that for a fixed* $\delta > 0$,

$$p_1^{1+\delta}/n \to 0, \tag{4.2}$$

$$\sup_{n \geq 1} \sup_{\|d\|=1, d \in H_1} E\left(d^T X_i\right)^{4K}\left(1 + \varepsilon_i^2\right) < \infty, \tag{4.3}$$

where K is the smallest integer greater than or equal to $2/\delta$,

$$p_1\sqrt{p_1 - p_0}/n \to 0, \tag{4.4}$$

$$\sup_{n\geq 1}\sup_d E\left(d^T X_i\right)^4 \varepsilon_i^4 < \infty, \tag{4.5}$$

where the second supremum is taken over $d \in \mathbf{H}_1 \cap \mathbf{H}_0^\perp$ *with* $\|d\| = 1$. *Then the following holds for the resampling plans* (4.1a) *and* (4.1b):

$$d_\infty\left(\mathscr{L}^*(T^*), \mathscr{L}(T^0)\right) \to 0 \quad (\textit{in probability}), \tag{4.6}$$

$$d_\infty\left(\mathscr{L}^*(T^W), \mathscr{L}(T^0)\right) \to 0 \quad (\textit{in probability}), \tag{4.7}$$

where

$$T^0 = \frac{\|(\Pi_1 - \Pi_0)(\mathbf{Y} - \mathbf{X}\beta)\|^2/(p_1 - p_0)}{\|(I - \Pi_1)(\mathbf{Y})\|^2/(n - p_1)}.$$

Note that

$$T^0 = \frac{\|(\Pi_1 - \Pi_0)\boldsymbol{\varepsilon}\|^2/(p_1 - p_0)}{\|(I - \Pi_1)\boldsymbol{\varepsilon}\|^2/(n - p_1)}$$

whether $\mathbf{H}_0$ is true or not and that if $\beta \in \mathbf{H}_0$, then $T = T^0$. Therefore bootstrap and wild bootstrap estimate the distribution of T consistently on the hypotheses $\mathbf{H}_0$. On the alternative $\mathbf{H}_1$ the wild bootstrap and the bootstrap estimate converges to the distribution of T under a distribution with $\beta = 0 \in \mathbf{H}_0$. For the validity of the approximation by an F-distribution in general one has to assume additionally that $E(\varepsilon_i^2|X_i) = E\varepsilon_i^2$, which is not necessary for the bootstrap and the wild bootstrap.

If $p_1 - p_0$ is bounded, then in Theorem 4 it suffices to assume instead of (4.5) that for every $\gamma > 0$ and $d_n \in \mathbf{H}_1 \cap \mathbf{H}_0^\perp$ with $\|d_n\| = 1$: $E(d_n^T X_i)^2\varepsilon_i^2 I[(d_n^T X_i)^2\varepsilon_i^2 > \gamma n] \to 0$. For the case of $p_1 - p_0 \to \infty$, (4.5) can be replaced by the weaker assumption $\sup_{n\geq 1}(p_1 - p_0)^{-2}E\|X_{i,2}\|^4\varepsilon_i^4 < \infty$, where $X_{i,2}$ is the projection of X_i onto $\mathbf{H}_0^\perp \cap \mathbf{H}_1$.

For F-tests we consider here also the homoscedastic model (1.6). Instead of looking at conditional laws given $(X_1, \dots, X_n)$ we assume now that $X_1, \dots, X_n$ are nonrandom. If assumption (1.6) held, one could apply the bootstrap which is based on resampling from the centered residuals $\hat{\varepsilon}_i - \hat{\varepsilon}.$ [Efron (1982)]. Here $\hat{\varepsilon}_i = Y_i - X_i^T\hat{\beta}_1$ are the residuals and $\hat{\varepsilon}. = (1/n)\Sigma_{i=1}^n \hat{\varepsilon}_i$.

Bootstrap (Residual resampling). One generates a resample $\varepsilon_1^\otimes, \dots, \varepsilon_n^\otimes$ from $\{\sqrt{n/(n - p_1)}\,(\hat{\varepsilon}_1 - \hat{\varepsilon}.), \dots, \sqrt{n/(n - p_1)}\,(\hat{\varepsilon}_n - \hat{\varepsilon}.)\}$. Put $\boldsymbol{\varepsilon}^\otimes = (\varepsilon_1^\otimes, \dots, \varepsilon_n^\otimes)^T$. Then the bootstrap observations are generated by

$$\mathbf{Y}^\otimes = \mathbf{X}\hat{\beta}_0 + \boldsymbol{\varepsilon}^\otimes. \tag{4.8}$$

Using the bootstrap observations $(X_1, Y_1^{\otimes}), \dots, (X_n, Y_n^{\otimes})$ one constructs

$$T^{\otimes} = \frac{\left\|(\Pi_1 - \Pi_0)(\mathbf{Y}^{\otimes})\right\|^2/(p_1 - p_0)}{\left\|(I - \Pi_1)(\mathbf{Y}^{\otimes})\right\|^2/(n - p_1)}.$$

Now on the hypotheses $\mathscr{L}(T)$ is estimated by $\mathscr{L}^*(T^{\otimes})$.

This is a slight modification of the procedure in Efron (1982) where it has been proposed to resample from $\{\hat{\varepsilon}_1 - \hat{\varepsilon}_., \dots, \hat{\varepsilon}_n - \hat{\varepsilon}_.\}$. In the following theorem we use the following Mallows distances $d_q(\mu, \nu) = \inf\{(E\|X - Y\|^q)^{1/q}: \mathscr{L}(X) = \mu, \mathscr{L}(Y) = \nu\}$ and $d(\mu, \nu) = \inf\{(E[\inf(\|X - Y\|^2, 1)])^{1/2}: \mathscr{L}(X) = \mu, \mathscr{L}(Y) = \nu\}$.

THEOREM 5. *Assume* (4.8) *and*

$$Y_i = X_i^T\beta + \varepsilon_i, \tag{4.9}$$

where $\beta \in \mathbf{H}_1$ *and* ε_i *are i.i.d. with* $E\varepsilon_i = 0$ *and where* $\sum_{i=1}^n X_i X_i^T$ *is nonsingular. We make the following assumptions*:

(4.10) $d_4(\mu, \hat{\mu}) \to 0$ *(in probability), where* $\mu = \mathscr{L}(\varepsilon_i)$ *and* $\hat{\mu}$ *is the empirical measure* $\hat{\mu} = (1/n)\sum_{i=1}^n \delta_{\varepsilon_i}$,

$$p_1/n \to 0, \tag{4.11}$$

$$0 < \inf_n E\varepsilon_i^4 \le \sup_n E\varepsilon_i^4 < \infty. \tag{4.12}$$

Then

$$d\left(\mathscr{L}^*\left((p_1 - p_0)^{1/2}T^{\otimes}\right), \mathscr{L}\left((p_1 - p_0)^{1/2}T^0\right)\right) \to 0 \ (\textit{in probability}), \tag{4.13}$$

$$\begin{aligned} &d_2\Big(\mathscr{L}^*\Big(\left\|(\Pi_1 - \Pi_0)(\mathbf{Y}^{\otimes})\right\|^2/(p_1 - p_0)^{1/2}\Big),\\ &\mathscr{L}\Big(\left\|(\Pi_1 - \Pi_0)(\mathbf{Y} - \mathbf{X}b)\right\|^2/(p_1 - p_0)^{1/2}\Big) \to 0 \ (\textit{in probability}). \end{aligned} \tag{4.14}$$

It can easily be seen that condition (4.10) holds if the distribution of ε_i does not depend on n and $E\varepsilon_i^4$ is bounded. Note that in our notation everything may depend on n.

Statement (4.13) makes sense because $(p_1 - p_0)^{1/2}$ is the right norming factor: Note that $\|(I - \Pi_1)(\mathbf{Y})\|^2/(n - p_1) = E\varepsilon_i^2 + O_P(n^{-1/2})$ and $\|(\Pi_1 - \Pi_0)(\mathbf{Y} - \mathbf{X}\beta)\|^2/(p_1 - p_0) = E\varepsilon_i^2 + O_P((p_1 - p_0)^{-1/2})$ (see the proof of Theorem 5).

If one would resample from $\{\hat{\varepsilon}_1 - \hat{\varepsilon}_., \dots, \hat{\varepsilon}_n - \hat{\varepsilon}_.\}$ instead of $\{\sqrt{n/(n - p_1)}\,(\hat{\varepsilon}_1 - \hat{\varepsilon}_.), \dots, \sqrt{n/(n - p_1)}\,(\hat{\varepsilon}_n - \hat{\varepsilon}_.)\}$, then for the consistency of bootstrap in (4.14) one has additionally to assume $\sqrt{p_1 - p_0}\,p_1/n \to 0$.

This result may be compared with the necessary conditions for the validity of the approximation of $\mathscr{L}(T)$ by an F-distribution. It can be seen that—given (4.11) and (4.12)—for bounded $p_1 - p_0$ it is necessary and sufficient that the maximal diagonal element of $(\Pi_1 - \Pi_0)$ converges to 0. Note that this condition is not necessary for the bootstrap.

5. Proofs.

PROOF OF THEOREM 1. In the proof of Theorem 1 we will show that (2.1)–(2.3) entail

$$\sqrt{n}\, c^T\left(\hat{\beta} - E\hat{\beta}_K\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (c^T X_i)\varepsilon_i + o_p(1),$$

$$\sqrt{n}\, c^T\left(\hat{\beta}^* - E^*\hat{\beta}_K^*\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (c^T X_i^*)\varepsilon_i^* + o_p(1),$$

$$\sqrt{n}\, c^T\left(\hat{\beta}^W - E^*\hat{\beta}_K^W\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (c^T X_i)\varepsilon_i^W + o_p(1).$$

Here $\hat{\beta}_K$, $\hat{\beta}_K^*$ and $\hat{\beta}_K^W$ are higher order stochastic expansions of $\hat{\beta}$, $\hat{\beta}^*$ and $\hat{\beta}^W$ (respectively). $E^*(\ )$ denotes the conditional expectation $E(\ |\mathbf{X}, \mathbf{Y})$ and $\varepsilon_i^* = Y_i^* - X_i^{*T}\hat{\beta}$. By application of the Lindeberg condition (2.3) we will then show that these expressions have the same normal limit. The convergence of the expectations will be ensured by $E(\varepsilon_i | X_i) = 0$ or $\delta \geq 1/3$ ($\delta \geq 1$, respectively). Note that (1.1)–(1.3) imply $EX_i\varepsilon_i = 0$. The proof of Theorem 1 is divided into several lemmas.

LEMMA 0. *Assume* (2.2). *Then*

$$E\left(\|X_i\|/\sqrt{p}\right)^{4K}\left(1 + \varepsilon_i^2\right) \text{ is uniformly bounded for } n \geq 1. \tag{5.1}$$

PROOF. Choose an orthogonal basis $\{e_j\colon j = 1, \dots, p\}$ of $\mathbf{R}^p$. Then

$$\begin{aligned} E\left(\|X_i\|^2/p\right)^{2K}\left(1 + \varepsilon_i^2\right) &= E\left(p^{-1} \sum_{j=1}^{p} \left(e_j^T X_i\right)^2\right)^{2K} \left(1 + \varepsilon_i^2\right) \\ &= p^{-2K} \sum_{j_1, \dots, j_{2K}=1}^{p} E\left(e_{j_1}^T X_i\right)^2 \cdots \left(e_{j_{2K}}^T X_i\right)^2\left(1 + \varepsilon_i^2\right) \\ &\leq \text{const.} \end{aligned}$$

□

LEMMA 1. *Under the assumptions of Theorem* 1,

$$\lambda_{\text{amax}}\left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i^T - I\right) = O_P\left(p^{1/2+\delta/4} n^{-1/2}\right) \tag{5.2}$$

(λ_{amax} *denotes the maximal absolute eigenvalue*).

PROOF. Put $A = I - (1/n)\Sigma_{i=1}^{n} X_i X_i^T$. We will show

$$E\Lambda = O(p^{K+1}n^{-K}) \tag{5.3}$$

for $\Lambda = \text{trace}(A^{2K}) \geq \lambda_{\max}(A^2)^K$. This implies (5.2).

PROOF OF (5.3). First note that

$$E\Lambda = n^{-2K}\sum_J \text{trace}\, E\big(X_{i_1}X_{i_1}^T - I\big) \cdot \cdots \cdot \big(X_{i_{2K}}X_{i_{2K}}^T - I\big), \tag{5.4}$$

where $J = \{(i_1, \dots, i_{2K})$: for every j there exists $h \neq j$ with $i_h = i_j\}$.

After evaluating the matrix product in (5.4) $E\Lambda$ is a sum with summands of the following type:

$$\begin{aligned} n^{-2K}S &= n^{-2K}\,\text{trace}\, X_{i_1}X_{i_1}^T \cdot \cdots \cdot X_{i_s}X_{i_s}^T \\ &= n^{-2K}\big(X_{i_1}^T X_{i_2}\big) \cdot \cdots \cdot \big(X_{i_s}^T X_{i_1}\big), \end{aligned} \tag{5.5}$$

where $0 \leq s \leq 2K$. Put $r = \#\{i_1, \dots, i_s\}$. To every summand S one can construct a graph with r nodes and s edges. Every node corresponds to an index i_j and two nodes i_h and i_j are connected by an edge if $(X_{i_h}^T X_{i_j})$ appears in S. The order of a node i is the number of edges touching i [where an edge going from i to i (a loop) counts twice]. For $(r, s) = (K, 2K)$ the order of every node is 4. The graphs are always connected and the order of every node is even.

Now the number of *different* graphs corresponding to the sum in (5.4) is bounded. Furthermore one can show that the number of summands of type (5.5) corresponding to the *same* graph with r nodes and s edges is bounded by $K^K n^K$ if $s \leq 2r$ and by $K^K n^{r+[(2K-s)/2]}$ if $s \geq 2r$ ($[x]$ denotes $\sup(n \in \mathbf{N}$: $n \leq x)$). For $s \leq 2r$ this follows immediately because the maximal number of different indices in J is K and because for every index j there exist less than K possibilities to choose $h \neq j$ with $i_h = i_j$. For $s \geq 2r$ note that after replacement of *two* factors $(X_{i_j}X_{i_j}^T - I)$ and $(X_{i_h}X_{i_h}^T - I)$ by I (at most) *one* index may disappear in $\text{trace}(X_{i_1}X_{i_1}^T - I) \cdot \cdots \cdot (X_{i_{2K}}X_{i_{2K}}^T - I)$. Therefore if $s \geq 2r$ the term S corresponds to a summand in (5.4) with an index $(i_1, \dots, i_{2K}) \in J$ such that $\#\{i_1, \dots, i_{2K}\} \leq r + [(2K - s)/2]$.

We will show

$$E(S) = O(p^{s+1-r}). \tag{5.6}$$

This implies the statement of Lemma 1 because of the following:

$$\begin{aligned} n^{-2K}p^{s+1-r}n^{K} &\leq p^{K+1}n^{-K} \quad \text{if } s \leq 2r, \\ n^{-2K}p^{s+1-r}n^{r+[(2K-s)/2]} &\leq p^{K+1}n^{-K}\big(p^{s-r-K}/n^{s/2-r}\big) \\ &\leq p^{K+1}n^{-K} \quad \text{if } s \geq 2r. \end{aligned} \tag{5.7}$$

PROOF OF (5.6). Define

$Q = \{i_1, \dots, i_s\}$,

$\tilde{Q} = \{i \in Q$: There are exactly two edges connecting i with another node$\}$.

First consider the case $\tilde{Q} = \varnothing$. From (2.2) one can follow

$$ES = O\left(p^{\#\text{loops}} p^{(1/2)(s - \#\text{loops})}\right) = O\left(p^{(1/2)(s + \#\text{loops})}\right), \tag{5.8}$$

where #loops denotes the number of loops in the graph. Because of $\tilde{Q} = \varnothing$ and because the order of every node is even there are at least four edges for every node i connecting i with another node. Then there must be at least $2r$ edges which are no loops. This implies #loops $\le s - 2r$ and $ES = O(p^{s-r}) = O(p^{s-r+1})$.

Suppose now that $\tilde{Q} \neq \varnothing$. Then there exists an $i_0 \in \tilde{Q}$. Denote the neighbors of i_0 by i_1 and i_2 (we do not exclude the case $i_1 = i_2$). Now note that

$$E\left(\left(X_{i_0}^T X_{i_1}\right)\left(X_{i_0}^T X_{i_2}\right)\left(X_{i_0}^T X_{i_0}\right)^h \middle| X_{i_1}, X_{i_2}\right) = \left(X_{i_1}^T M X_{i_2}\right) p^h, \tag{5.9}$$

where M is a $p \times p$ matrix with bounded maximal absolute eigenvalue (uniformly in i_0 and n). Without loss of generality assume $M = I$. Then

$$E(S | X_i: i \neq i_0) = p^h \tilde{S}, \tag{5.10}$$

where $\tilde{S}$ is of type (5.5). The graph which corresponds to $\tilde{S}$ has $r - 1$ nodes and $s - h - 1$ edges. It can be generated from the graph of S by removing the node i_0 and all edges which touch i_0 and by adding an edge which goes from i_1 to i_2. Because of (5.10) it suffices to prove (5.6) for $r = 1$. □

LEMMA 2. *Under the assumptions of Theorem* 1,

$$\operatorname{var}\left(c^T A^h \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \varepsilon_i\right) = O\left(\left(\frac{p}{n}\right)^h\right), \tag{5.11}$$

$$E\left(c^T A^h \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \varepsilon_i\right) = O\left(\frac{p^k}{n^{k-1/2}}\right), \tag{5.12}$$

for $1 \le h \le K - 1$ *and* $k = [h/2] + 1$.

PROOF. The statements of the lemma follow by lengthy computations. Note that

$$c^T A^h \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \varepsilon_i = \sum_{1 \le i_1, \dots, i_h \le n} c^T\left(X_{i_1} X_{i_1}^T - I\right) \cdots \left(X_{i_h} X_{i_h}^T - I\right) X_{i_h} \varepsilon_{i_h}.$$

Remember that (1.1)–(1.3) imply $EX_i \varepsilon_i = 0$. □

LEMMA 3. *Under the assumptions of Theorem* 1,

$$\sqrt{n}\, c^T\left(\hat{\beta} - E\hat{\beta}_K\right) = c^T \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \varepsilon_i + o_p(1), \tag{5.13}$$

where $\hat{\beta}_K = \beta + \Sigma_{0 \le h \le K-1} A^h (1/\sqrt{n}) \Sigma_{i=1}^n X_i \varepsilon_i$.

PROOF. Use $((1/n)\Sigma_{i=1}^n X_i X_i^T)^{-1} = (I - A)^{-1} = I + A + A^2 + \cdots$. Because of Lemma 1,

$$\begin{aligned} \left| c^T \sum_{h \ge K} A^h \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \varepsilon_i \right| &\le \lambda_{\mathrm{amax}}(A^K) \frac{1}{1 - \lambda_{\mathrm{amax}}(A)} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \varepsilon_i \right\| \\ &= O_p\left(\left(p^{1/2+\delta/4} n^{-1/2}\right)^K p^{1/2} \right) \\ &= O_p\left(\left(p^{1+\delta} n^{-1}\right)^{K/2} \right) = o_p(1). \end{aligned} \tag{5.14}$$

Because of Lemma 2,

$$\left| c^T \sum_{1 \le h \le K-1} A^h \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \varepsilon_i - EA^h \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \varepsilon_i \right| = o_p(1). \qquad \square$$

LEMMA 4. *Under the assumptions of Theorem* 1,

$$d_\infty\left(\mathscr{L}\left(\sqrt{n}\, c^T\left(\hat{\beta} - E\hat{\beta}_K\right) \right), N\left(0, E\left(c^T X_i\right)^2 \varepsilon_i^2\right) \right) \to 0. \tag{5.15}$$

PROOF. Use condition (2.3) and Lemma 3. $\square$

Now we state the bootstrap counterparts of Lemmas 1 to 4.

LEMMA 1*. *Under the conditions of Theorem* 1,

$$\lambda_{\mathrm{amax}}(A^*) = O_p\left(\frac{p^{1/2+\delta/4}}{n^{1/2}} \right), \tag{5.16}$$

where $A^* = I - (1/n)\Sigma_{i=1}^n X_i^* (X_i^*)^T$.

PROOF. Similarly as in the proof of Lemma 1 we consider $\Lambda^* = \mathrm{trace}(A^*)^{2K}$. Note that $E\Lambda^*$ is a sum of summands of type (5.5). This sum can be treated as in the proof of Lemma 1. $\square$

LEMMA 2*. *Under the assumptions of Theorem* 1,

$$\left\| c^T A^{*h} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^* \varepsilon_i^* - E^* c^T A^{*h} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^* \varepsilon_i^* \right\| = o_p(1), \tag{5.17}$$

$$\text{for } 0 \le h \le K-1.$$

PROOF. See the proof of Lemma 2. □

LEMMA 3*. *Under the assumptions of Theorem* 1,

$$\sqrt{n}\, c^T\left(\hat{\beta}^* - E\hat{\beta}_K^*\right) = c^T \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^* \varepsilon_i^* + o_p(1), \tag{5.18}$$

where $\hat{\beta}_K^* = \hat{\beta} + \sum_{0 \le h \le K-1} A^{*h}(1/\sqrt{n})\sum_{i=1}^n X_i^* \varepsilon_i^*$.

PROOF. Note that because of Lemma 1* (see the proof of Lemma 3),

$$\begin{aligned} c^T\hat{\beta}^* &= c^T(I - A^*)^{-1} n^{-1} \sum_{i=1}^n X_i^* Y_i^* = c^T\hat{\beta} + c^T(I - A^*)^{-1} n^{-1} \sum_{i=1}^n X_i^* \varepsilon_i^* \\ &= c^T\hat{\beta}_K^* + o_P(n^{-1/2}). \end{aligned}$$

With Lemma 2* and $E^* X_i^* \varepsilon_i^* = n^{-1}\sum_{i=1}^n X_i \hat{\varepsilon}_i = 0$ this implies (5.18). □

LEMMA 4*. *Under the assumptions of Theorem* 1,

$$d_\infty\left(\mathscr{L}^*\left(\sqrt{n}\, c^T\left(\hat{\beta}^* - E\hat{\beta}_K^*\right), N\left(0, E\left(c^T X_i\right)^2 \varepsilon_i^2\right)\right) \to 0 \quad (\textit{in probability}). \tag{5.19}$$

PROOF. For every i there exists a j such that $(X_i^*, Y_i^*) = (X_j, Y_j)$. Then $\varepsilon_i^* = Y_i^* - X_i^{*T}\hat{\beta} = Y_j - X_j^T\hat{\beta} = \varepsilon_j - X_j^T(\hat{\beta} - \beta)$. Put $\varepsilon_{i,1}^* = \varepsilon_j$ and $\varepsilon_{i,2}^* = -X_j^T(\hat{\beta} - \beta)$. We apply Lemma 3* and write $n^{-1/2}\sum_{i=1}^n (c^T X_i^*)\varepsilon_i^* = T_1 + T_2$, where $T_1 = (1/\sqrt{n})\sum_{i=1}^n (c^T X_i^*)\varepsilon_{i,1}^*$ and $T_2 = (1/\sqrt{n})\sum_{i=1}^n (c^T X_i^*)\varepsilon_{i,2}^*$. First we show

$$T_2 - E^* T_2 \to 0 \quad (\text{in probability}). \tag{5.20}$$

PROOF OF (5.20). First note that

$$\begin{aligned} \operatorname{var}^*(T_2) = E^*(T_2 - E^* T_2)^2 &\le \frac{1}{n} \sum_{i=1}^n \left(c^T X_i\right)^2 \left(X_i^T(\hat{\beta} - \beta)\right)^2 \\ &= (\hat{\beta} - \beta)^T \frac{1}{n} \sum_{i=1}^n Z_i Z_i^T (\hat{\beta} - \beta) \end{aligned}$$

for $Z_i = (c^T X_i) X_i$. Using $\|\hat{\beta} - \beta\|^2 = O_P(p/n)$ and using that $E(\|Z_i\|/\sqrt{p})^{2K}$

and $E(d^TZ_i)^{2K}$ are uniformly bounded for $\|d\| = 1$ and $n \geq 1$ [see (2.2) and Lemma 0] one can show with the same arguments as in the proof of Lemma 1 that $\lambda_{\text{amax}}(n^{-1}\Sigma_{i=1}^n Z_iZ_i^T - EZ_iZ_i^T) = o_P(1)$. This implies $\text{var}^*(T_2) = O_P(p/n)$ and (5.20). □

Now note that $\Sigma_{i=1}^n (c^TX_i)(Y_i - X_i^T\hat{\beta}) = 0$ implies that $E^*T_2 = -n^{-1/2}\Sigma_{i=1}^n (c^TX_i)(X_i^T(\hat{\beta} - \beta)) = -n^{-1/2}\Sigma_{i=1}^n (c^TX_i)\varepsilon_i = -E^*T_1$. Therefore it remains to show the asymptotic normality of $T_1 - E^*T_1$. This can be shown by proving the Lindeberg condition: $\exists\ \gamma_n \to 0$ such that $n^{-1}\Sigma_{i=1}^n [(c^TX_i)\varepsilon_i - \Delta]^2 I([(c^TX_i)\varepsilon_i - \Delta]^2 > \gamma_n n) \to 0$ (in probability), where $\Delta = n^{-1}\Sigma_{i=1}^n (c^TX_i)\varepsilon_i$. For the proof of the Lindeberg condition first note that $\Delta = o_p(1)$ and that $n^{-1}\Sigma_{i=1}^n (c^TX_i)^2\varepsilon_i^2 I((c^TX_i)^2\varepsilon_i^2 > \gamma_n n) \to 0$ (in probability) if γ_n goes slowly enough to 0 [because of (2.3)]. This implies for γ_n converging slowly enough to 0:

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^n \big[(c^TX_i)\varepsilon_i - \Delta\big]^2 I\Big(\big[(c^TX_i)\varepsilon_i - \Delta\big]^2 > \gamma_n n\Big) \\
&\quad \leq n^{-1}\sum_{i=1}^n \big[(c^TX_i)\varepsilon_i\big]^2 I\Big(\big[(c^TX_i)\varepsilon_i - \Delta\big]^2 > \gamma_n n\Big) + o_P(1) \\
&\quad \leq n^{-1}\sum_{i=1}^n \big[(c^TX_i)\varepsilon_i\big]^2 \Big\{I\Big[\big|(c^TX_i)\varepsilon_i\big| \geq 0.5\gamma_n^{1/2}n^{1/2}\Big] \\
&\qquad\qquad + I\Big[\big|(c^TX_i)\varepsilon_i\big| \leq 0.5\gamma_n^{1/2}n^{1/2}, \\
&\qquad\qquad\qquad |\Delta| \geq 0.5\gamma_n^{1/2}n^{1/2}\Big]\Big\} + o_P(1) \\
&\quad \leq n^{-1}\sum_{i=1}^n 0.25\gamma_n n I\big[|\Delta| \geq 0.5\gamma_n^{1/2}n^{1/2}\big] + o_P(1) \\
&\quad \leq \Delta^2 + o_P(1) = o_P(1).
\end{aligned}
$$

□

For the proof of the first statement of Theorem 1 it remains to show the following lemma:

LEMMA 5. *Under the assumptions of Theorem* 1

$$
E^*\Big(c^T\big(\hat{\beta}_K^* - \hat{\beta}\big)\Big) - E\Big(c^T\big(\hat{\beta}_K - \beta\big)\Big) = o_P\big(1/\sqrt{n}\,\big) \tag{5.21}
$$

if $\delta \geq 1/3$ *or* $E(\varepsilon_i|X_i) = 0$.

PROOF. Assume first $E(\varepsilon_i|X_i) = 0$. Then $E(c^T\hat{\beta}_K) = c^T\beta$. Furthermore by straightforward calculations one can show $E^*(c^T\hat{\beta}_K^*) = c^T\hat{\beta} + o_p(n^{-1/2})$ (see also the proof of Theorem 2). This shows (5.21).

Assume now $\delta \geq 1/3$. Then $p^2n^{-3/2} = o(1)$ and (see Lemma 2)

$$\sqrt{n}\,Ec^T\big(\hat{\beta}_K - \beta\big) = Ec^T(I + A)n^{-1/2}\sum_{i=1}^{n} X_i\varepsilon_i + O(p^2n^{-3/2})$$

$$= -n^{-1/2}E\big(c^TX_1\big)\|X_1\|^2\varepsilon_1 + o(1).$$

By similar calculations one gets

$$\sqrt{n}\,E^*c^T\big(\hat{\beta}_K^* - \hat{\beta}\big) = -n^{-3/2}\sum_{i=1}^{n}\big(c^TX_i\big)\|X_i\|^2\varepsilon_i + O_P(p^2n^{-3/2})$$

$$= \sqrt{n}\,Ec^T\big(\hat{\beta}_K - \beta\big) + o_P(1),$$

but this implies (5.21). □

We will now prove the second statement of Theorem 1—the consistency of wild bootstrap.

LEMMA 6. *Under the conditions of Theorem* 1,

$$\sqrt{n}\,c^T\big(\hat{\beta}^W - \hat{\beta}\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(c^TX_i\big)\varepsilon_i^W + o_p(1) \tag{5.22}$$

if $E(\varepsilon_i|X_i) = 0$ *or* $\delta \geq 1$.

PROOF. The proof goes along the lines of the proofs of Lemmas 1–3. Note that $\sqrt{n}\,E^*c^T(\hat{\beta}^W - \hat{\beta}) = o_p(1)$ (see Lemma 2). □

LEMMA 6*. *Under the conditions of Theorem* 1,

$$d_\infty\Big(\mathscr{L}^*\Big(\sqrt{n}\,c^T\big(\hat{\beta}^W - \hat{\beta}\big)\Big), N\Big(0, E\big(c^TX_i\big)^2\varepsilon_i^2\Big)\Big) \to 0 \quad (\textit{in probability}) \tag{5.23}$$

if $E(\varepsilon_i|X_i) = 0$ *or* $\delta \geq 1$.

PROOF. In the first part of the proof we will show that

$$\operatorname{var}^*\Big(\sqrt{n}\,c^T\big(\hat{\beta}^W - \hat{\beta}\big)\Big) \to E\big(c^TX_i\big)^2\varepsilon_i^2. \tag{5.24}$$

In the second part we will prove the asymptotic normality of

$$\mathscr{L}^*\left(\sqrt{n}\, c^T\left(\hat{\beta}^W - \hat{\beta}\right)\right).$$

First note that

$$\begin{aligned} \operatorname{var}^*\left(\sqrt{n}\, c^T\left(\hat{\beta}^W - \hat{\beta}\right)\right) &= \frac{1}{n} \sum_{i=1}^{n} E^*\left(c^T X_i\right)^2\left(\varepsilon_i^W\right)^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(c^T X_i\right)^2\left(\varepsilon_i - X_i^T\left(\hat{\beta} - \beta\right)\right)^2. \end{aligned}$$

Now

$$(5.25) \quad \frac{1}{n} \sum_{i=1}^{n} \left(c^T X_i\right)^2\left(X_i^T\left(\hat{\beta} - \beta\right)\right)^2 = o_p(1) \quad \text{(see the proof of Lemma 4}^*\text{)}.$$

Furthermore we will show

$$(5.26) \qquad \frac{1}{n} \sum_{i=1}^{n} \left(c^T X_i\right)^2 \varepsilon_i^2 \to E\left(c^T X_i\right)^2 \varepsilon_i^2.$$

But (5.25) and (5.26) imply (5.24).

PROOF OF (5.26). First choose $\gamma_n \to 0$ such that

$$E\left(c^T X_i\right)^2 \varepsilon_i^2 I\left(\left(c^T X_i\right)^2 \varepsilon_i^2 \geq \gamma_n n\right) \to 0.$$

Choose $c_n \to 0$, $c_n > \gamma_n$ such that

$$nP\left(\left(c^T X_i\right)^2 \varepsilon_i^2 > c_n n\right) \leq E\left(c^T X_i\right)^2 \varepsilon_i^2 / c_n I\left(\left(c^T X_i\right)^2 \varepsilon_i^2 \geq \gamma_n n\right) \to 0.$$

Then $\sup_{1 \leq i \leq n} (c^T X_i)^2 \varepsilon_i^2 = o_p(c_n n)$ and $(1/n)\sum_{i=1}^{n} (c^T X_i)^2 \varepsilon_i^2 - ((c^T X_i)^2 \varepsilon_i^2 \wedge c_n n) \to 0$ (in probability) ($a \wedge b$ denotes the infimum of a and b).

Now

$$\operatorname{var}\left(\frac{1}{n} \sum_{i=1}^{n} \left(\left(c^T X_i\right)^2 \varepsilon_i^2 \wedge c_n n\right)\right) \leq \frac{1}{n} E\left(c^T X_i\right)^2 \varepsilon_i^2 c_n n \to 0.$$

This implies (5.26). □

For the proof of asymptotic normality of $\mathscr{L}^*(\sqrt{n}\, c^T(\hat{\beta}^W - \hat{\beta}))$ it suffices to show, for $\gamma > 0$,

$$\frac{1}{n} \sum_{i=1}^{n} E^*\left(c^T X_i\right)^2\left(\varepsilon_i^W\right)^2 I\left(\left(c^T X_i\right)^2\left(\varepsilon_i^W\right)^2 \geq \gamma n\right) \to 0 \quad \text{(in probability)}.$$

But this follows similarly as in the proof of Lemma 4* from

$$\frac{1}{n}\sum_{i=1}^{n}(c^TX_i)^2(\varepsilon_i)^2 I\Big((c^TX_i)^2\varepsilon_i^2 \geq \gamma n\Big) = o_p(1) \quad \text{for } \gamma > 0$$

because of

$$|\varepsilon_i^W|^2 \leq \text{const. } \hat{\varepsilon}_i^2 \quad \text{or} \quad |\varepsilon_i^W|^2 \leq \text{const. } U_i^2\hat{\varepsilon}_i^2,$$

where $U_1, \dots, U_n$ are i.i.d. variables with distribution as described in the introduction. □

PROOF OF THEOREM 2. Equation (3.7) follows by Lemma 2. Equation (3.8) follows similarly. Note that we have additionally assumed that $E(\varepsilon_i|X_i) = 0$ holds. The expansions for the moments in (3.9)–(3.13) follow by lengthy calculations. For instance, to see (3.9), one has to consider

$$E^*\big(\hat{\theta}_{c,n}^*\big) = \sum_{h=0}^{2} E^*\left(c^TA^{*h}n^{-1/2}\sum_{i=1}^{n}X_i^*\varepsilon_i^*\right). \tag{5.27}$$

For the first term of the right-hand side of (5.27) one gets, using $\sum_{i=1}^n X_i\hat{\varepsilon}_i = 0$,

$$E^*\left(c^Tn^{-1/2}\sum_{i=1}^{n}X_i^*\varepsilon_i^*\right) = c^Tn^{-1/2}\sum_{i=1}^{n}X_i\hat{\varepsilon}_i = 0.$$

For the second term one gets, using again $\sum_{i=1}^n X_i\hat{\varepsilon}_i = 0$,

$$\begin{aligned}
E^*&\left(c^TA^*n^{-1/2}\sum_{i=1}^{n}X_i^*\varepsilon_i^*\right)\\
&= E^*\left(c^Tn^{-3/2}\sum_{i,j=1}^{n}\big(I - X_j^*X_j^{*T}\big)X_i^*\varepsilon_i^*\right)\\
&= E^*\left(c^Tn^{-3/2}\sum_{i=1}^{n}\big(I - X_i^*X_i^{*T}\big)X_i^*\varepsilon_i^*\right)\\
&= -c^Tn^{-3/2}\sum_{i=1}^{n}\|X_i\|^2X_i\hat{\varepsilon}_i\\
&= -n^{-3/2}\sum_{i=1}^{n}\|X_i\|^2c^TX_i\varepsilon_i\\
&\quad + n^{-5/2}\sum_{i,k=1}^{n}\|X_i\|^2c^TX_iX_i^T\left(n^{-1}\sum_{j=1}^{n}X_jX_j^T\right)^{-1}X_k\varepsilon_k.
\end{aligned}$$

These terms are both of order $O_P(p/n)$. This can easily be seen for the first term by calculating the second moment. For the second term, one plugs in the expansion $(\Sigma_{j=1}^n X_j X_j^T)^{-1} = I + A + A^2 + \cdots$. Then to bound the rest term one uses $\|n^{-2}\Sigma_{i=1}^n \|X_i\|^2 c^T X_i X_i\| = O_P(p/n)$, $\|n^{-1/2}\Sigma_{k=1}^n X_k \varepsilon_k\| = O_P(p^{1/2})$ and Lemma 1.

Similarly one shows that the third term of the right-hand side of (5.27) is of order $O_P(p/n)$. □

PROOF OF THEOREM 3. Equation (3.14) follows from (5.3) with appropriately chosen K. For the proof of the expansions (3.15)–(3.19) one calculates and compares the conditional moments of $n^{1/2}c^T(\hat{\beta} - \beta)$ (given $X_1, \dots, X_n$ and H_ε) and the conditional moments of $n^{1/2}c^T(\hat{\beta}^W - \hat{\beta})$ (given $Y_1, \dots, Y_n$, $X_1, \dots, X_n$ and H_ε). The calculation of these moments is easier than in the proof of Theorem 2 because $n^{1/2}c^T(\hat{\beta} - \beta)$ and $n^{1/2}c^T(\hat{\beta}^W - \hat{\beta})$ are sums of conditionally independent zero mean variables. □

PROOF OF THEOREM 4. For simplicity we assume $\beta \in \mathbf{H}_0$. In the next lemma we treat the denominator of T^0, T^* and T^W.

LEMMA 7. *Assume the conditions of Theorem 4. Then*

$$\|(I - \Pi_1)\mathbf{Y}\|^2/(n - p_1) = \Delta_1 + O_P\big(p_1/n + n^{-1/2}\big), \tag{5.28}$$

$$\|(I - \Pi_1)\mathbf{Y}^W\|^2/(n - p_1) = \Delta_1 + O_P\big(p_1/n + n^{-1/2}\big), \tag{5.29}$$

$$\|(I - \Pi_1)\mathbf{Y}^*\|^2/(n - p_1) = \Delta_1 + O_P\big(p_1/n + n^{-1/2}\big), \tag{5.30}$$

where $\Delta_1 = E\varepsilon_1^2$.

PROOF OF LEMMA 7. For the proof of (5.28) note that $\|(I - \Pi_1)\mathbf{Y}\|^2 = \|(I - \Pi_1)\boldsymbol{\varepsilon}\|^2 = \|\boldsymbol{\varepsilon}\|^2 - \|\Pi_1\boldsymbol{\varepsilon}\|^2$. Write $X_{i,1}$ for the projection of X_i onto $\mathbf{H}_1$ and denote by J: $\mathbf{H}_1 \to \mathbf{H}_1$ the inverse of the linear map which is defined on $\mathbf{H}_1$ by the matrix $(1/n)\Sigma_{i=1}^n X_{i,1}X_{i,1}^T)$. Then one gets

$$\|\Pi_1\boldsymbol{\varepsilon}\|^2 = n^{-2} \sum_{i,j,k=1}^{n} \big(X_{i,1}^T J X_{j,1}\varepsilon_j\big)\big(X_{i,1}^T J X_{k,1}\varepsilon_k\big) = n^{-1} \sum_{i,j=1}^{n} \big(X_{i,1}^T J X_{j,1}\big)\varepsilon_j \varepsilon_k.$$

Now using a stochastic expansion of the random linear map J (see Lemma 1) one can show $\|\Pi_1\boldsymbol{\varepsilon}\|^2 = n^{-1}\Sigma_{i,j=1}^n (X_{i,1}^T X_{j,1})\varepsilon_j\varepsilon_k + o_P(p_1) = O_P(p_1)$. $\|\boldsymbol{\varepsilon}\|^2 = nE\varepsilon_1^2 + O_P(n^{1/2})$ implies (5.28). Equations (5.29) and (5.30) follow similarly.

□

Now we treat the numerator of T^0, T^* and T^W.

LEMMA 8. *Assume $\beta \in \mathbf{H}_0$ and the conditions of Theorem 4. Then*

$$\|(\Pi_1 - \Pi_0)\mathbf{Y}\|^2/\sqrt{p_1 - p_0} = \Delta_2 + Z + o_P(1), \tag{5.31}$$

$$\|(\Pi_1 - \Pi_0)\mathbf{Y}^{\mathbf{W}}\|^2/\sqrt{p_1 - p_0} = \Delta_2 + Z^W + o_P(1), \tag{5.32}$$

$$\|(\Pi_1 - \Pi_0)\mathbf{Y}^*\|^2/\sqrt{p_1 - p_0} = \Delta_2 + Z^* + o_P(1), \tag{5.33}$$

where $\Delta_2 = E\|X_{1,2}\|^2\varepsilon_1^2(p_1 - p_0)^{-1/2}$ and

$$Z = n^{-1}(p_1 - p_0)^{-1/2} \sum_{i,j=1}^{n} \varepsilon_i\big(X_{i,2}^T X_{j,2}\big)\varepsilon_j,$$

$$Z^W = n^{-1}(p_1 - p_0)^{-1/2} \sum_{i,j=1}^{n} \varepsilon_i^W\big(X_{i,2}^T X_{j,2}\big)\varepsilon_j^W,$$

$$Z^* = n^{-1}(p_1 - p_0)^{-1/2} \sum_{i,j=1}^{n} \varepsilon_i^*\big(X_{i,2}^{*T} X_{j,2}^*\big)\varepsilon_j^*.$$

Here $X_{i,2}$ and $X_{i,2}^*$ are the projections of X_i and X_i^* (resp.) onto $\mathbf{H}_0^\perp \cap \mathbf{H}_1$.

The proof of Lemma 8 is similar to the proof of Lemma 7. The following lemma implies the statement of Theorem 4 for $p_1 - p_0 \to \infty$. Note that $\Delta_2(\Delta_1 + O_P(p_1/n + n^{-1/2}))^{-1} = \Delta_2\Delta_1^{-1} + O_P(1)$ follows from (4.4) and $\Delta_2 = O((p_1 - p_0)^{1/2})$. In the case of bounded $p_1 - p_0$ Theorem 4 follows because $n^{-1/2}\sum_{i=1}^n X_{i,2}\varepsilon_i$, $n^{-1/2}\sum_{i=1}^n X_{i,2}^*\varepsilon_i^*$ and $n^{-1/2}\sum_{i=1}^n X_{i,2}\varepsilon_i^W$ have the same asymptotic normal limit.

LEMMA 9. *Assume $p_1 - p_0 \to \infty$, $\beta \in \mathbf{H}_0$ and the conditions of Theorem 4. Then*

$$d_\infty\big(P, N(0,\sigma^2)\big) \to 0 \quad (\textit{in probability}),$$

where $P = \mathscr{L}(Z)$, $\mathscr{L}^*(Z^W)$ or $\mathscr{L}^*(Z^*)$ and $\sigma^2 = E(p_1 - p_0)^{-1}\varepsilon_1^2\varepsilon_2^2(X_{1,2}^T X_{2,2})^2$.

PROOF OF LEMMA 9. First one shows $\tilde{Z} - Z = o_P(1)$, where $\tilde{Z} = n^{-1}(p_1 - p_0)^{-1/2}\sum_{i \neq j} \varepsilon_i(X_{i,2}^T X_{j,2})\varepsilon_j$. According to Theorem 2.1 in de Jong (1987) for the proof of asymptotic normality of Z it suffices to show

$$E\tilde{Z}^4/(E\tilde{Z}^2)^2 \to 3. \tag{5.34}$$

The proof of (5.34) is straightforward. The proof of the asymptotic normality of Z^W and Z^* goes similarly. $\square$

PROOF OF THEOREM 5. Suppose $\sum_{i=1}^n X_i X_i^T = I$ (w.l.o.g.) and define $X_{i,1}$ and $X_{i,2}$ as the projections of X_i onto $\mathbf{H}_1$ or onto $\mathbf{H}_0^\perp \cap \mathbf{H}_1$ (respectively). Then $\hat{\beta}_1 = \sum_i X_{i,1}\varepsilon_i$ and $\sum_i \|X_{i,1}\|^2 = \operatorname{spur} \sum_i X_{i,1}X_{i,1}^T = p_1$. Furthermore

$$\begin{aligned}
\frac{1}{n-p_1}\|(I-\Pi_1)(\mathbf{Y})\|^2 &= \frac{1}{n-p_1}\|\boldsymbol{\varepsilon}\|^2 - \frac{1}{n-p_1}\|\Pi_1(\boldsymbol{\varepsilon})\|^2 \\
&= \frac{1}{n-p_1}\sum_{i=1}^n \varepsilon_i^2 - \frac{1}{n-p_1}\sum_{i,j=1}^n \varepsilon_i X_{i,1}^T X_{j,1}\varepsilon_j \qquad (5.35)\\
&= E\varepsilon_i^2 + O_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

because of

$$\begin{aligned}
\operatorname{var}\left(\frac{1}{n-p_1}\sum_{i\neq j}\varepsilon_i\left(X_{i,1}^T X_{j,1}\right)\varepsilon_j\right) &\le \left(\frac{1}{n-p_1}\right)^2 2\sum_{i\neq j} E\varepsilon_i^2\left(X_{i,1}^T X_{j,1}\right)^2\varepsilon_j^2 \\
&\le (n-p_1)^{-2}\left(E\varepsilon_i^2\right)^2 2\sum_{i=1}^n \|X_{i,1}\|^2 \\
&= O\left(p_1(n-p_1)^{-2}\right) = o(n^{-1})
\end{aligned}$$

and

$$\begin{aligned}
&\operatorname{var}\left(\frac{1}{n-p_1}(n-p_1)^{-2}\sum_{i=1}^n\left(1-\|X_{i,1}\|^2\right)^2\right) \\
&\quad = O\left((n-p_1)^{-2}\sum_{i=1}^n\left(1-\|X_{i,1}\|^2\right)^2\right) = O\left(\frac{1}{n}\right).
\end{aligned}$$

(Note that $\sum_{i=1}^n \|X_{i,1}\|^4 \le \sum_{i=1}^n \|X_{i,1}\|^2 = p_1$ because of $\|X_{i,1}\| \le 1$.)

Similarly one gets

$$\begin{aligned}
&\frac{1}{n-p_1}\|(I-\Pi_1)(\mathbf{Y}^\otimes)\|^2 \\
&= \frac{1}{n-p_1}\|\boldsymbol{\varepsilon}^\otimes\|^2 - \frac{1}{n-p_1}\|\Pi_1(\boldsymbol{\varepsilon}^\otimes)\|^2 \\
&= E^*\left(\varepsilon_i^\otimes\right)^2 + O_P\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^n \frac{n}{n-p_1}\left(\hat{\varepsilon}_i - \hat{\varepsilon}_.\right)^2 + O_P\left(\frac{1}{\sqrt{n}}\right) \qquad (5.36)\\
&= \frac{1}{n-p_1}\sum_{i=1}^n \hat{\varepsilon}_i^2 + O_P\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n-p_1}\|(I-\Pi_1)(\boldsymbol{\varepsilon})\|^2 + O_P\left(\frac{1}{\sqrt{n}}\right) \\
&= E\varepsilon_i^2 + O_P\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

where in the fourth equation the following has been used:

$$\begin{aligned}
\operatorname{var}(\hat{\varepsilon}.) &= \operatorname{var}\left(n^{-1}\sum_{j=1}^{n}\varepsilon_j - n^{-1}\sum_{i,j=1}^{n}\left(X_{i,1}^T X_{j,1}\right)\varepsilon_j\right)\\
&= E\left(\varepsilon_i^2\right)n^{-2}\sum_{j=1}^{n}\left(1-\sum_{i=1}^{n}\left(X_{i,1}^T X_{j,1}\right)\right)^2\\
&\le 2E\left(\varepsilon_i^2\right)\left[n^{-2}\sum_{i,j,k=1}^{n}\left(X_{i,1}^T X_{j,1}\right)\left(X_{j,1}^T X_{k,1}\right)+n^{-1}\right]\\
&= 2E\left(\varepsilon_i^2\right)\left[n^{-2}\sum_{i,k=1}^{n}\left(X_{i,1}^T X_{k,1}\right)+n^{-1}\right]\\
&\le 2E\left(\varepsilon_i^2\right)\left[n^{-2}\left(\sum_{i,k=1}^{n}\left(X_{i,1}^T X_{k,1}\right)^2 n^2\right)^{1/2}+n^{-1}\right]\\
&= 2E\left(\varepsilon_i^2\right)\left[n^{-2}\left(\sum_{i=1}^{n}\|X_{i,1}\|^2 n^2\right)^{1/2}+n^{-1}\right]\\
&= 2E\left(\varepsilon_i^2\right)\left(\sqrt{p_1}+1\right)/n = o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

We show now that

$$d_4\left(\hat{\mu}, \mathscr{L}^*\left(\varepsilon_i^{\otimes}\right)\right) \to 0 \quad \text{(in probability)}. \tag{5.37}$$

This follows from $n^{-1}\Sigma_{j=1}^{n}(\hat{\varepsilon}. + X_j^T\hat{\beta}_1)^4 \to 0$. Note that

$$\frac{1}{n}\sum_{j=1}^{n}\left(X_j^T\beta_1\right)^4 \to 0 \quad \text{(in probability)}, \tag{5.38}$$

$$\hat{\varepsilon}. \to 0 \quad \text{(in probability)}. \tag{5.39}$$

PROOF OF (5.38).

$$\begin{aligned}
En^{-1}\sum_{j=1}^{n}\left(X_j^T\hat{\beta}_1\right)^4 &= En^{-1}\sum_{j=1}^{n}\left(\sum_{i=1}^{n}X_{j,1}^T X_{i,1}\varepsilon_i\right)^4\\
&\le \text{const.}\, n^{-1}\sum_{i,j,k=1}^{n}\left(X_{j,1}^T X_{i,1}\right)^2\left(X_{j,1}^T X_{k,1}\right)^2\\
&\le \text{const.}\, n^{-1}\sum_{j=1}^{n}\|X_{j,1}\|^4\\
&\le \text{const.}\, n^{-1}\sum_{j=1}^{n}\|X_{j,1}\|^2 \to 0.
\end{aligned}$$

Now because of (4.11) and (5.37) there exist (conditionally) independent $\eta_i, \eta_i^{\otimes}$ with $\mathscr{L}^*(\eta_i) = \mathscr{L}(\varepsilon_i)$, $\mathscr{L}^*(\eta_i^{\otimes}) = \mathscr{L}^*(\varepsilon_i^{\otimes})$ and $E^*(\eta_i^2 - (\eta_i^{\otimes})^2)^2 \to 0$ (in probability) [here $E^*(\cdots) = E(\cdots|\mathbf{X}, \mathbf{Y})$]. Define $\mathbf{Z} = \mathbf{X}\beta + \boldsymbol{\eta}$ and $\mathbf{Z}^{\otimes} = \mathbf{X}\hat{\beta}_0 + \boldsymbol{\eta}^{\otimes}$. Then

$$
\begin{aligned}
&E^*\left[\frac{1}{\sqrt{p_1 - p_0}}\|(\Pi_1 - \Pi_0)(\mathbf{Z} - \mathbf{X}\beta)\|^2 - \frac{1}{\sqrt{p_1 - p_0}}\|(\Pi_1 - \Pi_0)(\mathbf{Z}^{\otimes})\|^2\right]^2 \\
&\quad = E^*\left[\frac{1}{\sqrt{p_1 - p_0}}\|(\Pi_1 - \Pi_0)(\boldsymbol{\eta})\|^2 - \frac{1}{\sqrt{p_1 - p_0}}\|(\Pi_1 - \Pi_0)(\boldsymbol{\eta}^{\otimes})\|^2\right]^2 \\
&\quad = E^*\frac{1}{p_1 - p_0}\left\{\sum_{i=1}^n \|X_{i,2}\|^2\left(\eta_i^2 - (\eta_i^{\otimes})^2\right) + \sum_{i \neq j}\left(X_{i,2}^T X_{j,2}\right)\left(\eta_i\eta_j - \eta_i^{\otimes}\eta_j^{\otimes}\right)\right\}^2 \qquad (5.40)\\
&\quad \le 2(p_1 - p_0)^{-1}\left\{E^*\left[\sum_{i=1}^n \cdots\right]^2 + E^*\left[\sum_{i\neq j}\cdots\right]^2\right\} \le T_1 + T_2 + T_3,
\end{aligned}
$$

where

$$
\begin{aligned}
T_1 &= \frac{2}{p_1 - p_0}\sum_{i=1}^n E^*\left[(\cdots)^2\right] = \frac{2}{p_1 - p_0}\sum_{i=1}^n E^*\|X_{i,2}\|^4\left(\eta_i^2 - (\eta_i^{\otimes})^2\right)^2 \\
&\le 2E^*\left(\left(\eta_i^2 - (\eta_i^{\otimes})^2\right)\right)^2 = o_P(1), \\
T_2 &= \frac{2}{p_1 - p_0}\left\{E^*\left[\sum_{i=1}^n \cdots\right]\right\}^2 \le \frac{2}{p_1 - p_0}(p_1 - p_0)^2\left\{E^*\eta_i^2 - E^*\left(\eta_i^{\otimes}\right)^2\right\}^2 \\
&= 2(p_1 - p_0)\left\{E\varepsilon_i^2 - E^*\left(\varepsilon_i^{\otimes}\right)^2\right\}^2 = O_P\left(\frac{p_1 - p_0}{n}\right) = o_P(1), \\
T_3 &= \frac{2}{p_1 - p_0}E^*\left[\sum_{i\neq j}\cdots\right]^2 \\
&\le 2(p_1 - p_0)^{-1}\sum_{i,j=1}^n \left(X_{j,2}^T X_{i,2}\right)^2 E^*\left(\eta_1\eta_2 - \eta_1^{\otimes}\eta_2^{\otimes}\right)^2 \\
&= 2E^*\left(\eta_1\eta_2 - \eta_1^{\otimes}\eta_2^{\otimes}\right)^2 = o_P(1).
\end{aligned}
$$

Now (5.35), (5.36) and (5.40) imply the statements of the theorem. □

REFERENCES

BERAN, R. (1986). Comment on "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14** 1295–1298.

BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457–468.

BICKEL, P. J. and FREEDMAN, D. A. (1983). Bootstrapping regression models with many parameters. In *A Festschrift for Erich L. Lehmann in Honor of his Sixty-fifth Birthday* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 28–48. Wadsworth, Belmont, CA.

CAO-ABAD, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Ann. Statist.* **19** 2226–2231.

CAO-ABAD, R. and GONZALES-MANTEIGA, W. (1990). Bootstrap methods in regression smoothing: an alternative procedure to the wild resampling plan. Preprint.

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.

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.

EHM, W. (1986). On maximum likelihood estimation in high-dimensional log-linear type models. I. The independent case. Preprint SFB 123, Univ. Heidelberg.

EHM, W. (1991). Statistical problems with many parameters: critical quantities for approximate normality and posterior density based inference. Habilitationsschrift, Univ. Heidelberg.

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

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

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

HÄRDLE, W. and MAMMEN, E. (1990). Comparing nonparametric versus parametric regression fits. Preprint SFB 123, Univ. Heidelberg.

HÄRDLE, W. and MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796.

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

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

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

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

MAMMEN, E. (1990). Higher-order accuracy of bootstrap for smooth functionals. Preprint SFB 123, Univ. Heidelberg.

MAMMEN, E. (1992). *When Does Bootstrap Work: Asymptotic Results and Simulations. Lecture Notes in Statist.* **77.** Springer, Berlin.

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

PORTNOY, S. (1985). Asymptotic behaviour of M-estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. [Correction: **19** (1991) 2282.]

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366.

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

SHORACK, G. (1982). Bootstrapping robust regression. *Comm. Statist. A* **11** 961–972.

WU, C. F. J. (1986). Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1295.

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

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT HEIDELBERG
IM NEUENHEIMER FELD 294
W-6900 HEIDELBERG
GERMANY