NBER WORKING PAPER SERIES

HIGHER ORDER PROPERTIES OF THE WILD BOOTSTRAP UNDER MISSPECIFICATION

Patrick M. Kline
Andres Santos

Working Paper 16793
http://www.nber.org/papers/w16793

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2011

Higher Order Properties of the Wild Bootstrap Under Misspecification
Patrick M. Kline and Andres Santos
NBER Working Paper No. 16793
February 2011
JEL No. C12

ABSTRACT

We examine the higher order properties of the wild bootstrap (Wu, 1986) in a linear regression model with stochastic regressors. We find that the ability of the wild bootstrap to provide a higher order refinement is contingent upon whether the errors are mean independent of the regressors or merely uncorrelated. In the latter case, the wild bootstrap may fail to match some of the terms in an Edgeworth expansion of the full sample test statistic, potentially leading to only a partial refinement (Liu and Singh, 1987). To assess the practical implications of this result, we conduct a Monte Carlo study contrasting the performance of the wild bootstrap with the traditional nonparametric bootstrap.

Patrick M. Kline
Department of Economics
UC, Berkeley
508-1 Evans Hall #3880
Berkeley, CA 94720
and NBER
pkline@econ.berkeley.edu

Andres Santos
Department of Economics
UC, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508
a2santos@ucsd.edu

1 Introduction

The wild bootstrap of Wu (1986) and Liu (1988) provides a procedure for conducting inference in the model:

$$Y = X'\beta_0 + \epsilon \ , \tag{1}$$

where $Y \in \mathbf{R}$, $X \in \mathbf{R}^{d_x}$ and ϵ may have a heteroskedastic structure of unknown form. This robustness to arbitrary heteroscedasticity provides a distinct advantage over the residual bootstrap of Freedman (1981) while retaining some of its computational and statistical advantages. This has led to increasing attention among economists who are often concerned with robust inference in small sample environments (Horowitz (1997, 2001), Cameron et al. (2008), Davidson and Flachaire (2008)) and to a variety of recent extensions beyond the basic linear regression model (Cavaliere and Taylor (2008), Gonçalves and Meddahi (2009), Davidson and MacKinnon (2010)). To date, however, the higher order properties of the wild bootstrap have only been studied under the assumption that the errors are mean independent of the regressors. Liu (1988) established that when this condition holds the wild bootstrap provides a refinement over a normal approximation.

In this paper we contribute to the literature by analyzing the higher order properties of the wild bootstrap in instances where the conditional mean function may be misspecified. Concretely, we examine the ability of the wild bootstrap to provide a refinement over the normal approximation when ϵ is uncorrelated with X but not necessarily mean independent of it – a setting pervasive in economics where regressors are stochastic rather than fixed or chosen by the econometrician. It is precisely in such environments that heteroskedasticity is likely to arise (White (1982)) making the higher order properties of the wild bootstrap of particular interest.

We conduct our analysis in two steps. First, we compute the approximate cumulants (Bhattacharya and Ghosh (1978)) of t-statistics under both the full sample and bootstrap distributions with general assumptions on the wild bootstrap weights. We show that both the first and third approximate cumulants may disagree up to order $O_p(n^{-\frac{1}{2}})$ if higher powers of X are correlated with ϵ; a situation that is ruled out under proper specification. This higher order discordance between the approximate cumulants under the full sample and bootstrap distribution implies that if valid Edgeworth expansions exist they would only be equivalent up to order $O_p(n^{-\frac{1}{2}})$ (Hall (1992)). As a result, despite remaining consistent under misspecification, the wild bootstrap may fail to provide a higher order refinement over a normal approximation.

We complement this result by formally establishing the existence of valid one term Edgeworth expansions when the distribution of the wild bootstrap weights is additionally assumed to be strongly nonlattice (Bhattacharya and Rao (1976)). In accord with Liu (1988) we note that one-sided wild bootstrap tests obtain a refinement to order $O_p(n^{-1})$ under proper specification. However, this result is undermined by certain forms of misspecification under which only some, but not all, of the second order terms in the full sample Edgeworth expansion are matched by their bootstrap counterparts. Consequently, the wild bootstrap may provide only a partial refinement over the normal approximation (Liu and Singh (1987)). To assess the practical implications of this result, we conclude by conducting a Monte Carlo study contrasting the performance of the wild bootstrap with the traditional nonparametric bootstrap in the presence of misspecification.

The rest of the paper is organized as follows. Section 2 contains our theoretical results while Section 3 examines the implications of our analysis in a simulation study. We briefly conclude in Section 4 and relegate all proofs to the Appendix.

2 Theoretical Results

While numerous variants of the wild bootstrap exist, we study the original version proposed by Wu (1986) and Liu (1988). Succinctly, given a sample $\{Y_i, X_i\}_{i=1}^n$ and $\hat{\beta}$ the OLS estimator from such sample, the wild bootstrap generates new errors and dependant variables:

$$Y_i^* \equiv X_i'\hat{\beta} + \epsilon_i^* \qquad\qquad \epsilon_i^* \equiv (Y_i - X_i'\hat{\beta})W_i \ , \tag{2}$$

where $\{W_i\}_{i=1}^n$ is an i.i.d. sample independent of the original data $\{Y_i, X_i\}_{i=1}^n$. A bootstrap estimator $\hat{\beta}^*$ can then be computed from the sample $\{Y_i^*, X_i\}_{i=1}^n$ and the distribution of $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$ conditional on $\{Y_i, X_i\}_{i=1}^n$ (but not $\{W_i\}_{i=1}^n$) used to approximate that of $\sqrt{n}(\hat{\beta} - \beta_0)$. While it may not be possible to compute the bootstrap distribution analytically, it is straightforward to simulate it.

We focus our analysis on inference on linear contrasts of β_0, which includes both individual coefficients and predicted values as special cases. In particular, for an arbitrary $c \in \mathbf{R}^{d_x}$ we examine:

$$T_n \equiv \frac{\sqrt{n}}{\hat{\sigma}} c'(\hat{\beta} - \beta_0) \qquad\qquad \hat{\sigma}^2 \equiv c' H_n^{-1} \Sigma_n(\hat{\beta}) H_n^{-1} c \ , \tag{3}$$

where the $d_x \times d_x$ matrices H_n and $\Sigma_n(\beta)$ are defined by:

$$H_n \equiv \frac{1}{n}\sum_{i=1}^n X_i X_i' \qquad\qquad \Sigma_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^n X_i X_i'(Y_i - X_i'\beta)^2 \ . \tag{4}$$

The bootstrap statistic T_n^* is then the analogue to T_n but computed on $\{Y_i^*, X_i\}_{i=1}^n$ instead. Namely,

$$T_n^* \equiv \frac{\sqrt{n}}{\hat{\sigma}^*} c'(\hat{\beta}^* - \hat{\beta}) \qquad\qquad (\hat{\sigma}^*)^2 \equiv c' H_n^{-1} \Sigma_n^*(\hat{\beta}^*) H_n^{-1} c \ , \tag{5}$$

where H_n is as in (4), and $\Sigma_n^*(\beta) \equiv \frac{1}{n}\sum_i X_i X_i'(Y_i^* - X_i'\beta)^2$.

As argued in Mammen (1993), under mild assumptions on the wild bootstrap weights $\{W_i\}_{i=1}^n$, the distribution of T_n^* conditional on $\{Y_i, X_i\}_{i=1}^n$, (but not $\{W_i\}_{i=1}^n$) provides a consistent estimator for the distribution of T_n. Consequently, tests based upon a comparison of the statistic T_n to the quantiles of the bootstrap distribution of T_n^* are asymptotically justified. In what follows, we explore whether such a procedure provides a refinement over employing the quantiles of a standard normal distribution instead.

2.1 Assumptions

We explore the higher order properties of the wild bootstrap under the following assumptions:

Assumption 2.1. *(i)* $\{Y_i, X_i\}_{i=1}^n$ *is an i.i.d. sample, satisfying* (1) *with* $E[X\epsilon] = 0$*; (ii)* (Y, X) *are bounded almost surely; (iii)* $E[XX'] = I$ *and* $\Sigma_0 \equiv E[XX'\epsilon^2]$ *is full rank; (iv) For* $Z \equiv (X'\epsilon, vech(XX')', vech(XX'\epsilon^2)')'$*, and* ξ_Z *its characteristic function,* $\limsup_{\|t\|\to\infty} |\xi_Z(t)| < 1$.[1]

Assumption 2.2. *(i)* $\{W_i\}_{i=1}^n$ *is i.i.d., independent of* $\{Y_i, X_i\}_{i=1}^n$ *with* $E[W] = 0$*,* $E[W^2] = 1$ *and* $E[W^\omega] < \infty$*,* $\omega \geq 9$*; (ii) For* $U \equiv (W, W^2)'$ *,* ξ_U *its characteristic function,* $\limsup_{|t|\to\infty} |\xi_U(t)| < 1$.

Assumption 2.1(i) allows for misspecification of the conditional mean function by requiring $E[X\epsilon] = 0$ rather than $E[\epsilon|X] = 0$. In Assumption 2.1(ii) we impose that (Y, X) be bounded. This specialized (yet widely applicable) setting simplifies the arguments employed in obtaining an Edgeworth expansion for T_n^*. Our finding that the wild bootstrap may fail to provide a higher order refinement under misspecification would not be overturned if Assumption 2.1(ii) were weakened to less stringent moment conditions. Assumption 2.1(ii) additionally imposes that $E[XX'] = I$, which is just a normalization in the present context; see Remark 2.1. The requirements on $\{W_i\}_{i=1}^n$ in Assumption 2.2(i) are standard in the wild bootstrap literature and satisfied by all commonly used choices of wild bootstrap weights.

Assumptions 2.1(i)-(iii) and 2.2(i) suffice for showing that the approximate cumulants of T_n and of T_n^* under the bootstrap distribution may disagree up to order $O_p(n^{-\frac{1}{2}})$ under misspecification. In order to additionally establish the existence of Edgeworth expansions, however, we also impose Assumptions 2.1(iv) and 2.2(ii). These requirements, also known as Cramer's condition, are standard in the Edgeworth expansion literature (Bhattacharya and Rao (1976)). They are satisfied,

[1]For a symmetric matrix A, vech(A) denotes a column vector composed of its unique elements.

for example, if the distributions of Z and U have a component that is absolutely continuous with respect to Lebesgue measure. Unfortunately, this requirement rules out two frequently used wild bootstrap weights: Rademacher random variables and the weighting scheme advocated in Mammen (1993). Thus, while our results on approximate cumulants are applicable to these choices of weights, our results on Edgeworth expansions are not.

Remark 2.1. Since we study T_n for generic vectors $c \in \mathbf{R}^{d_x}$, Assumption 2.1(iii) is just a convenient normalization. Specifically, suppose $E[XX'] = \Sigma_X$ for Σ_X full rank. We may then rewrite (1) as:

$$Y = \tilde{X}'\beta_{I,0} + \epsilon \qquad \tilde{X} = \Sigma_X^{-\frac{1}{2}} X \qquad \beta_{I,0} = \Sigma_X^{\frac{1}{2}}\beta_0 \ . \tag{6}$$

It is then immediate that $E[\tilde{X}\tilde{X}'] = I$. Moreover, since $\sum_i \tilde{X}_i\tilde{X}_i'$ is invertible if and only if H_n is, we obtain that for any $\tilde{c} \in \mathbf{R}^{d_x}$ and $\hat{\beta}_I$ the OLS estimator on $\{Y_i, \tilde{X}_i\}_{i=1}^n$:

$$\sqrt{n}\tilde{c}'(\hat{\beta}_I - \tilde{\beta}_0) = \sqrt{n}\tilde{c}'\Sigma_X^{\frac{1}{2}}(\sum_{i=1}^n X_iX_i')^{-1}\Sigma_X^{\frac{1}{2}}\sum_{i=1}^n \Sigma_X^{-\frac{1}{2}}X_i\epsilon_i = \sqrt{n}c'(\hat{\beta} - \beta_0) \ , \tag{7}$$

where $c = \Sigma_X^{\frac{1}{2}}\tilde{c}$. Similarly, $\tilde{c}'(\frac{1}{n}\sum_i \tilde{X}_i\tilde{X}_i)^{-1}\frac{1}{n}\sum_i \tilde{X}_i\tilde{X}_i'(Y_i - \tilde{X}_i'\hat{\beta}_I)^2(\frac{1}{n}\sum_i \tilde{X}_i\tilde{X}_i')^{-1}\tilde{c} = \hat{\sigma}^2$ for $c = \Sigma_X^{\frac{1}{2}}\tilde{c}$. Hence, since the choice of $c \in \mathbf{R}^{d_x}$ is arbitrary, studying T_n for some c under Assumption 2.1(iii) is equivalent to studying it under the assumption that $E[XX']$ be full rank and $\tilde{c} = \Sigma_X^{-\frac{1}{2}}c$. ∎

Remark 2.2. Assumption 2.1(iv) precludes X from containing a constant term. To accommodate this common case, if the constant is the first element of the vector X, then Assumption 2.1(iv) should hold for $Z \equiv (X'\epsilon, \text{vech}_{-1}(XX')', \text{vech}(XX'\epsilon^2)')'$ where for a vector $v = (v^{(1)}, \ldots, v^{(d)})$ we define $v_{-1} \equiv (v^{(2)}, \ldots, v^{(d)})$. ∎

2.2 Approximate Cumulants

In what follows, for notational simplicity, we denote expectations, probability and law statements conditional on $\{Y_i, X_i\}_{i=1}^n$ (but not $\{W_i\}_{i=1}^n$) by E^*, P^* and L^* respectively. Additionally, we define the following parameters which play a fundamental role in our higher order analysis:

$$\sigma^2 \equiv c'\Sigma_0 c \qquad \kappa \equiv E[(c'X)^3\epsilon^3] \qquad \gamma_0 \equiv E[(c'X)^2X\epsilon] \qquad \gamma_1 \equiv E[(c'X)(X'X)\epsilon] \ . \tag{8}$$

Finally, we let Φ denote the distribution of a standard normal random variable and ϕ its density.

We begin our analysis by obtaining an asymptotic expansion for T_n and T_n^*.

Theorem 2.1. *Suppose Assumption 2.1(i)-(iii) and 2.2(i) hold, and for* $c \in \mathbf{R}^{d_x}$ *define:*

$$L_n \equiv c'\{I + \Delta_n\}\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} X_i\epsilon_i - \frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^{n}(c'X_i)\epsilon_i\{(\hat{\sigma}_R^2 - \sigma^2) - \frac{2}{n}\sum_{i=1}^{n}\gamma_0'X_i\epsilon_i\} \tag{9}$$

$$L_n^* \equiv c'H_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1} X_i\epsilon_i^*\{\frac{1}{\hat{\sigma}} - \frac{1}{2\hat{\sigma}^3}((\hat{\sigma}_s^*)^2 - \hat{\sigma}^2)\} \tag{10}$$

where $\Delta_n \equiv I - H_n$, $\hat{\sigma}_R^2 \equiv c'\Sigma_n(\beta_0)c + 2c'\Delta_n\Sigma_0 c$ *and* $(\hat{\sigma}_s^*)^2 \equiv c'H_n^{-1}\Sigma_n^*(\hat{\beta})H_n^{-1}c$*. It then follows that:*

$$T_n = L_n + o_p(n^{-\frac{1}{2}}) \qquad T_n^* = L_n^* + o_{p^*}(n^{-\frac{1}{2}}) \;\; a.s.$$

Recall that in Assumption 2.1(ii) the covariance $E[XX']$ was normalized to equal the identity matrix. Therefore $\Delta_n \equiv I - H_n$ is the estimation error in the Hessian and the first term in (9) captures the contribution to T_n of not knowing the true value of $E[XX']$. Similarly, the contribution of having to estimate the variance is divided into two parts: (i) $\frac{2}{n}\sum_i \gamma_0'X_i\epsilon_i$ which reflects use of $\hat{\beta}$ rather than β_0 in the sample variance calculations and (ii) $\hat{\sigma}_R^2 - \sigma^2$ which captures the randomness that would be present in estimating σ^2 if β_0 were known. Interestingly, these terms are smaller order under the bootstrap distribution due to the mean independence of ϵ^* and X.

Due to their polynomial form, the moments of L_n and L_n^* are considerably easier to compute than those of T_n and T_n^*. However, the cumulants of L_n and L_n^* provide only an approximation to those of T_n and T_n^* and were for this reason termed "approximate cumulants" by Bhattacharya and Ghosh (1978). Despite their approximate nature, the cumulants of L_n and L_n^* play a crucial role as they may be employed in place of the cumulants of T_n and T_n^* when computing their second order Edgeworth expansions if such expansions are indeed valid. Thus, a discordance between the approximate cumulants is indicative of an analogous difference in the corresponding Edgeworth expansions if such expansions do exist.

Theorem 2.2 shows the approximate cumulants may disagree under misspecification.

Theorem 2.2. *Let* $\mathcal{X}_k(L_n)$ *and* $\mathcal{X}_k^*(L_n^*)$ *denote the* k^{th} *cumulants of* L_n *and* L_n^* *respectively and define* $\hat{\kappa} \equiv \frac{1}{n}\sum_i (c'H_n^{-1}X_i)^3(Y_i - X_i'\hat{\beta})^3$*. If Assumptions 2.1(i)-(iii) and 2.2(i) hold, then:*

$$\mathcal{X}_1(L_n) = -\frac{\kappa}{2\sigma^3\sqrt{n}} - \frac{\gamma_1}{\sigma\sqrt{n}} + \frac{2c'\Sigma_0\gamma_0}{\sigma^3\sqrt{n}} \qquad \mathcal{X}_1^*(L_n^*) = -\frac{E[W^3]\hat{\kappa}}{2\hat{\sigma}^3\sqrt{n}}$$

$$\mathcal{X}_2(L_n) = 1 + O(n^{-1}) \qquad \mathcal{X}_2^*(L_n^*) = 1 + O_{a.s.}(n^{-1})$$

$$\mathcal{X}_3(L_n) = -\frac{2\kappa}{\sigma^3\sqrt{n}} + \frac{6c'\Sigma_0\gamma_0}{\sigma^3\sqrt{n}} + O(n^{-1}) \qquad \mathcal{X}_3^*(L_n^*) = -\frac{2E[W^3]\hat{\kappa}}{\hat{\sigma}^3\sqrt{n}} + O_{a.s.}(n^{-1})\ .$$

Observe first that unless $\kappa = 0$, the wild bootstrap fails to correct the first term in the first and third cumulants if $E[W^3] \neq 1$. This property has already been noted in Liu (1988) who advocates imposing $E[W^3] = 1$ for precisely this reason. However, even with this restriction, two additional terms in the first and third cumulants of L_n remain. These terms capture (i) the correlation between H_n and $\frac{1}{n}\sum_i X_i\epsilon_i$, and (ii) the additional randomness of employing $\hat{\beta}$ rather than β_0 in estimating σ^2. Both these expressions are of smaller order under mean independence but may be present otherwise. Because the wild bootstrap imposes mean independence in the bootstrap distribution it fails to mimic these terms. As a result, a discordance between the full sample and bootstrap approximate cumulants will arise under misspecification if the error term ϵ is correlated with higher powers of X so that γ_0 or γ_1 are nonzero.

2.3 Edgeworth Expansions

Under the additional requirement that the Cramer conditions hold (Assumptions 2.1(iv) and 2.2(ii)) we now establish that the discordance in approximate cumulants indeed translates into an analogous disagreement between Edgeworth expansions.

Theorem 2.3. *Under Assumptions 2.1(i)-(iv) and 2.2(i)-(ii) it follows that uniformly in z:*

$$P(T_n \le z) = \Phi(z) + \frac{\phi(z)\kappa}{6\sigma^3\sqrt{n}}(2z^2+1) - \frac{\phi(z)}{\sigma^3\sqrt{n}}(c'\Sigma_0\gamma_0(z^2+1) - \gamma_1\sigma^2) + o(n^{-\frac{1}{2}}) \tag{11}$$

$$P^*(T_n^* \le z) = \Phi(z) + \frac{\phi(z)\hat{\kappa}E[W^3]}{6\hat{\sigma}^3\sqrt{n}}(2z^2+1) + o(n^{-\frac{1}{2}}) \qquad a.s. \tag{12}$$

As Theorem 2.3 shows, the wild bootstrap provides the usual skewness correction whenever $E[W^3] = 1$. However, when the conditional mean function is misspecified, imposing mean independence in the wild bootstrap sample implies the bootstrap distribution may fail to match all the second order terms in the expansion for T_n. In particular, if ϵ is correlated with higher moments of X, so that γ_0 and γ_1 are not equal to zero, the wild bootstrap will only provide a partial refinement over a normal approximation. The importance of such a refinement is dependent on the degree of misspecification as measured by the magnitude of γ_0 and γ_1. In particular, if the misspecification is local with $\gamma_0, \gamma_1 = O(n^{-\frac{1}{2}})$, then the wild bootstrap does attain the usual higher order refinement.

3 Monte Carlo

We turn now to a study of the effect of misspecification on the finite sample performance of the wild bootstrap through a series of Monte Carlo sampling experiments. To ensure that our theoretical results are relevant, we restrict our attention to cases where: (i) X is continuously distributed

Table 1: REJECTION RATES FOR 0.05 NOMINAL SIZE - ONE SIDED TESTS

	Sample Size $n = 10$. Alternative Hypothesis $H_1 : \beta < 0$								
	Noise Level $\lambda = 0.25$			Noise Level $\lambda = 0.5$			Noise Level $\lambda = 1$		
	Analytical	Wild	Pairs	Analytical	Wild	Pairs	Analytical	Wild	Pairs
$\psi = -0.2$	0.100	0.054	0.073	0.102	0.061	0.077	0.096	0.071	0.073
$\psi = 0.0$	0.094	0.076	0.078	0.094	0.076	0.078	0.094	0.076	0.078
$\psi = 0.2$	0.221	0.186	0.163	0.153	0.130	0.120	0.114	0.092	0.095

	Sample Size $n = 10$. Alternative Hypothesis $H_1 : \beta > 0$								
	Noise Level $\lambda = 0.25$			Noise Level $\lambda = 0.5$			Noise Level $\lambda = 1$		
	Analytical	Wild	Pairs	Analytical	Wild	Pairs	Analytical	Wild	Pairs
$\psi = -0.2$	0.207	0.136	0.115	0.149	0.104	0.082	0.112	0.079	0.061
$\psi = 0.0$	0.078	0.052	0.039	0.078	0.052	0.039	0.078	0.052	0.039
$\psi = 0.2$	0.094	0.047	0.049	0.083	0.055	0.047	0.078	0.050	0.046

	Sample Size $n = 20$. Alternative Hypothesis $H_1 : \beta < 0$								
	Noise Level $\lambda = 0.25$			Noise Level $\lambda = 0.5$			Noise Level $\lambda = 1$		
	Analytical	Wild	Pairs	Analytical	Wild	Pairs	Analytical	Wild	Pairs
$\psi = -0.2$	0.077	0.053	0.060	0.076	0.059	0.060	0.075	0.069	0.087
$\psi = 0.0$	0.068	0.072	0.095	0.068	0.072	0.095	0.068	0.072	0.095
$\psi = 0.2$	0.175	0.155	0.145	0.127	0.109	0.127	0.090	0.078	0.110

	Sample Size $n = 20$. Alternative Hypothesis $H_1 : \beta > 0$								
	Noise Level $\lambda = 0.25$			Noise Level $\lambda = 0.5$			Noise Level $\lambda = 1$		
	Analytical	Wild	Pairs	Analytical	Wild	Pairs	Analytical	Wild	Pairs
$\psi = -0.2$	0.148	0.107	0.093	0.101	0.081	0.081	0.070	0.057	0.054
$\psi = 0.0$	0.048	0.035	0.049	0.048	0.035	0.049	0.048	0.035	0.049
$\psi = 0.2$	0.070	0.044	0.042	0.052	0.043	0.047	0.049	0.039	0.043

and bounded, (ii) ϵ is continuously distributed and bounded and (iii) the bootstrap weights W are continuously distributed with $E[W] = 0$, $E[W^2] = 1$, and $E[W^3] = 1$.

Let $Z \sim TN(\mu, \sigma^2, \tau)$ denote a normal random variable with mean μ and variance σ^2, truncated to lie in the interval $[-\tau, \tau]$. The regressor X was drawn from a mixture of $Z_1 \sim TN(0, 1, 2)$ with probability 0.1 and from $Z_2 \sim TN(1, 4, 4)$ with probability 0.9, recentered and scaled to have mean zero and variance one. We generate the variable Y according to the relationship:

$$Y_i = \psi\{X_i^2 - E[X^3]X_i - 1\} + \lambda\eta \ , \tag{13}$$

where η is the exponential of a $TN(0, 1, 2)$ random variable, recentered to have mean zero, and ψ, λ are scalar parameters that will be changed across different Monte Carlo specifications.

We examine the ability of the wild bootstrap to control size when conducting inference on the

Table 2: Rejection rates for 0.05 nominal size - Two sided tests

	Sample Size $n = 10$. Alternative Hypothesis $H_1 : \beta \neq 0$								
	Noise Level $\lambda = 0.25$			Noise Level $\lambda = 0.5$			Noise Level $\lambda = 1$		
	Analytical	Wild	Pairs	Analytical	Wild	Pairs	Analytical	Wild	Pairs
$\psi = -0.2$	0.242	0.145	0.070	0.181	0.121	0.052	0.136	0.097	0.039
$\psi = 0.0$	0.109	0.087	0.033	0.109	0.087	0.033	0.109	0.087	0.033
$\psi = 0.2$	0.244	0.167	0.062	0.174	0.130	0.051	0.131	0.106	0.042

	Sample Size $n = 20$. Alternative Hypothesis $H_1 : \beta \neq 0$								
	Noise Level $\lambda = 0.25$			Noise Level $\lambda = 0.5$			Noise Level $\lambda = 1$		
	Analytical	Wild	Pairs	Analytical	Wild	Pairs	Analytical	Wild	Pairs
$\psi = -0.2$	0.156	0.110	0.068	0.120	0.094	0.051	0.095	0.071	0.032
$\psi = 0.0$	0.066	0.060	0.028	0.066	0.060	0.028	0.066	0.060	0.028
$\psi = 0.2$	0.176	0.139	0.074	0.109	0.099	0.048	0.079	0.072	0.035

slope coefficient in the following linear regression model:

$$Y_i = \alpha + X_i\beta + \epsilon \ . \tag{14}$$

By construction, the unique parameters (α, β) ensuring that $E[X\epsilon] = 0$ in (14) are $(\alpha, \beta) = 0$. The parameter ψ in (13) therefore governs the extent of misspecification in the regression model, with $\psi = 0$ corresponding to proper specification ($E[Y|X] = 0$). Similarly, the scale parameter λ in (13) controls the level of noise in the linear regression.

Table 1 shows the empirical rejection rates of one-sided tests under different values of the parameters governing misspecification and residual noise. Code for our Monte Carlo experiments is available online. All rejection rates were computed using 200 bootstrap repetitions and 1,000 Monte Carlo replications. We implement the wild bootstrap drawing the weights W from a recentered Gamma distribution with shape parameter 4 and scale parameter 1/2 as suggested by Liu (1988). For comparison with the wild bootstrap, we also examine the ability of the nonparametric ("pairs") bootstrap and analytical t-tests to control size. [2]

The results suggest both the wild and nonparametric bootstraps yield improvements over an analytical t-test for one sided alternaitves. The relative performance of the two bootstraps under misspecification ($\psi \neq 0$) is dependent on the level of misspecification, the direction of the test and the level of noise. Table 2 provides false rejection rates for two-sided tests. Here the ranking of the various techniques is more clear cut with the nonparametric bootstrap performing best

[2]The nonparametric bootstrap computes the distribution of $\sqrt{n}c'(\hat{\beta} - \beta_0)/\hat{\sigma}$ under the empirical measure.

under misspecification and the normal approximation worst. Notably, the improvement of the wild bootstrap over the first order analytical approximation is still substantial, illustrating the practical importance of our theoretical finding of a partial refinement.

4 Conclusion

We find that the wild bootstrap may provide only a partial refinement over a normal approximation when the conditional mean function is misspecified. This suggests that while the wild bootstrap may not work as well as the nonparametric bootstrap in many settings where regression is used, it will likely still generate an improvement over analytical techniques. Our Monte Carlo study, for example, found that the wild bootstrap performed nearly as well as the nonparametric bootstrap in one-sided tests and still provided substantial improvements over normal approximations in two-sided tests. We conclude that in small sample environments where misspecification is of concern, the nonparametric bootstrap possesses a modest advantage over the wild bootstrap.

APPENDIX A - Proofs of Theorems 2.1 and 2.2

The following is a table of the notation and definitions that will be used throughout the appendix.

$\|\cdot\|_F$	On a matrix A, $\|A\|_F$ denotes the Frobenius norm.
$\|\cdot\|_o$	On a matrix A, $\|A\|_o$ denotes the usual operator norm.
$\|\lambda\|$	For a vector λ of positive integers and $\lambda^{(i)}$ its i^{th} coordinate $\|\lambda\| = \sum_i \lambda^{(i)}$.
$D^\lambda f$	For $f : \mathbf{R}^d \to \mathbf{R}$ and $\lambda \in \mathbf{R}$, $D^\lambda f = \frac{\partial^{\|\lambda\|} f}{\partial^{\lambda^{(1)}} \ldots \partial^{\lambda^{(d)}}}$.
e_i	The OLS residual $e_i = (Y_i - X_i\hat{\beta})$.
Φ	The distribution of a standard normal random variable in $\mathbf{R}^d$ (d may be context specific).

Lemma A.1. *Let $\{Z_i\}_{i=1}^n$ be an i.i.d. sample of Z a $k \times p$ random matrix with $\|Z\|_F$ bounded a.s.. Then:*

$$P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \{Z_i - E[Z_i]\}\|_F > M_n) = o(n^{-\frac{1}{2}}) \ ,$$

for any sequence $M_n \uparrow \infty$ such that $\log(n) = o(M_n)$.

Proof: Let $Z^{(l,j)}$ denote the (l,j) entry of Z. To establish the claim of the Lemma, then note that:

$$P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \{Z_i - E[Z_i]\}\|_F > M_n) \leq P(\max_{1\leq l \leq k, 1\leq j\leq p} |\frac{kp}{\sqrt{n}}\sum_{i=1}^n \{Z_i^{(l,j)} - E[Z_i^{(l,j)}]\}| > M_n)$$

$$\leq \sum_{l=1}^k \sum_{j=1}^p P(|\frac{kp}{\sqrt{n}}\sum_{i=1}^n \{Z_i^{(l,j)} - E[Z_i^{(l,j)}]\}| > M_n) \ . \quad (15)$$

Since $|Z_i^{(l,j)} - E[Z_i^{(l,j)}]| \leq K$ a.s. for some $K > 0$ and $1 \leq l \leq k$, $1 \leq j \leq p$, Bernstein's inequality implies:

$$P(|\frac{kp}{\sqrt{n}}\sum_{i=1}^n \{Z_i^{(l,j)} - E[Z_i^{(l,j)}]\}| > M_n) \leq 2\exp\{-\frac{M_n}{2kpK}\} \ , \quad (16)$$

since $M_n \uparrow \infty$. Results (15), (16) and $\log(n) = o(M_n)$ then establish the Lemma. ■

Lemma A.2. *Let $\Delta_n \equiv I - H_n$, $\hat{\sigma}_R^2 \equiv c'\Sigma_n(\beta_0)c + 2c'\Delta_n\Sigma_0 c$ and Assumptions 2.1(i)-(iii) hold. Then:*

(i) $P(\|\frac{1}{\sqrt{n}}\sum_i X_i\epsilon_i\| > M_n) = o(n^{-\frac{1}{2}})$ *for any sequence $M_n \uparrow \infty$ with $\log(n) = o(M_n)$.*

(ii) $P(\|H_n^{-1} - \sum_{j=0}^k \Delta_n^j\|_o > n^{-\alpha}) = o(n^{-\frac{1}{2}})$ *for any $\alpha \in [0, \frac{k+1}{2})$.*

(iii) $P(\|\hat{\beta} - \beta_0\| > n^{-\alpha}) = o(n^{-\frac{1}{2}})$ *for any $\alpha \in [0, \frac{1}{2})$.*

(iv) $P(|\hat{\sigma}^2 - \hat{\sigma}_R^2 + \frac{2}{n}\sum_i \gamma_0' X_i\epsilon_i| > n^{-\alpha}) = o(n^{-\frac{1}{2}})$ *for any $\alpha \in [0, \frac{1}{2})$.*

Proof: Since $\|X\epsilon\|$ is bounded a.s. by Assumption 2.1(ii), the first claim follows by Lemma A.1. For the second claim, notice Lemma A.1 implies that for any $M_n \uparrow \infty$ such that $\log(n) = o(M_n)$ we must have:

$$P(\|\Delta_n\|_F \geq \frac{M_n}{\sqrt{n}}) = o(n^{-\frac{1}{2}}) \ . \quad (17)$$

Moreover, notice that if $\|\Delta_n\|_F < 1$, then $H_n^{-1} = \sum_{j=0}^\infty \Delta_n^j$. Hence, we obtain:

$$P(\|H_n^{-1} - \sum_{j=0}^k \Delta_n^j\|_o > n^{-\alpha}) \leq P(\|\sum_{j\geq k+1} \Delta_n^j\|_o > n^{-\alpha} \text{ and } \|\Delta_n\|_F < 1) + P(\|\Delta_n\|_F \geq 1)$$

$$\leq P(\sum_{j\geq k+1} \xi(\Delta_n^j) > n^{-\alpha} \text{ and } \|\Delta_n\|_F < 1) + o(n^{-\frac{1}{2}}) \leq P(\frac{\xi^{k+1}(\Delta_n)}{1 - \xi(\Delta_n)} > n^{-\alpha}) + o(n^{-\frac{1}{2}}) \ , \quad (18)$$

where $\xi(\Delta_n^j)$ is the largest eigenvalue of Δ_n^j and we have exploited $\|\Delta_n^j\|_o = \xi(\Delta_n^j)$ and $\xi(\Delta_n^j) = \xi^j(\Delta_n)$. for the second and third inequalities. Moreover, since $\xi(\Delta_n) = \|\Delta_n\|_o \leq \|\Delta_n\|_F$, result (17) implies that $P(|\xi(\Delta_n)| \geq 1/2) = o(n^{-\frac{1}{2}})$. Therefore, from (18) we are able to conclude that:

$$P(\|H_n^{-1} - \sum_{j=0}^{k} \Delta_n^j\|_o > n^{-\alpha}) \leq P(2\xi^{k+1}(\Delta_n) > n^{-\alpha}) + o(n^{-\frac{1}{2}}) \leq P(2\|\Delta_n\|_F^{k+1} > n^{-\alpha}) + o(n^{-\frac{1}{2}}) \ . \quad (19)$$

To conclude, exploit (19) and set $M_n = n^{\frac{1}{2} - \frac{\alpha}{k+1}}$ in (17) to obtain $P(2\|\Delta_n\|_F > n^{-\frac{\alpha}{k+1}}) = o(n^{-\frac{1}{2}})$.

Next, note that Corollary III.2.6 in Bhatia (1997) implies $|\xi(H_n^{-1}) - 1| = |\xi(H_n^{-1}) - \xi(I)| \leq \|H_n^{-1} - I\|_F$. By part (ii) of the Lemma, it follows that $P(\|H_n^{-1}\|_o > 2) = o(n^{-\frac{1}{2}})$. Hence, we obtain:

$$P(\|\hat{\beta} - \beta_0\| > n^{-\alpha})$$
$$\leq P(\|\frac{2}{n}\sum_{i=1}^{n} X_i\epsilon_i\| > n^{-\alpha}) + P(\|H_n^{-1}\|_o > 2) = P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\epsilon_i\| > \frac{n^{\frac{1}{2}-\alpha}}{2}) + o(n^{-\frac{1}{2}}) \ . \quad (20)$$

The third claim of the Lemma is then established by (20), part (i) and $\alpha < 1/2$.

In order to establish the final claim of the Lemma, first observe that by direct calculation we obtain:

$$P(\|\Sigma_n(\hat{\beta}) - \Sigma_n(\beta_0)\|_F > n^{-\frac{\alpha}{2}}) = P(\|\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\{(X_i'(\hat{\beta} - \beta_0))^2 - 2\epsilon_i X_i'(\hat{\beta} - \beta_0)\}\|_F > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \quad (21)$$

where the final result is implied by part (iii), (X, ϵ) bounded a.s. by Assumption 2.1(ii) and $\alpha < 1$. Similarly, by Lemma A.1, for any sequence $M_n \uparrow \infty$ such that $\log(n) = o(M_n)$ we also have:

$$P(\|\Sigma_n(\beta_0) - \Sigma_0\|_F > \frac{M_n}{\sqrt{n}}) = o(n^{-\frac{1}{2}}) \ . \quad (22)$$

Let $K > 0$ be such that $\|\Sigma_0\|_o < K$ and note that since (21)-(22) imply $P(\|\Sigma_n(\hat{\beta}) - \Sigma_0\|_o > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}})$, it follows that $P(\|\Sigma_n(\hat{\beta})\|_o > K) = o(n^{-\frac{1}{2}})$. Hence, we conclude from part (ii) of the Lemma that:

$$P(|c'(H_n^{-1} - I)\Sigma_n(\hat{\beta})(H_n^{-1} - I)c| > n^{-\alpha})$$
$$\leq P(K\|c\|^2\|H_n^{-1} - I\|_o^2 > n^{-\alpha}) + P(\|\Sigma_n(\hat{\beta})\|_o > K) = o(n^{-\frac{1}{2}}) \ . \quad (23)$$

Similarly, exploiting again that $P(\|\Sigma_n(\hat{\beta})\|_o > K) = o(n^{-\frac{1}{2}})$ and part (ii) of the Lemma we also obtain:

$$P(|c'(H_n^{-1} - I - \Delta_n)\Sigma_n(\hat{\beta})c| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ . \quad (24)$$

Moreover, since $\alpha < 1$, exploiting (17), (21) and (22) we also conclude:

$$P(|c'\Delta_n(\Sigma_n(\hat{\beta}) - \Sigma_0)c| > n^{-\alpha}) \leq P(\|c\|^2\|\Delta_n\|_o\|\Sigma_n(\hat{\beta}) - \Sigma_0\|_o > n^{-\alpha})$$
$$\leq P(\|c\|\|\Delta_n\|_F > n^{-\frac{\alpha}{2}}) + P(\|c\|\|\Sigma_n(\hat{\beta}) - \Sigma_0\|_F > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \ . \quad (25)$$

Since (X, ϵ) is bounded, Lemma A.1 implies that $P(\|\frac{1}{n}\sum_i (c'X_i)^2\epsilon_i X_i - \gamma_0\| > \frac{M_n}{\sqrt{n}}) = o(n^{-\frac{1}{2}})$ for any $M_n \uparrow \infty$ with $\log(n) = o(M_n)$. Hence, using manipulations as in (25) we can conclude that:

$$P(\|\frac{1}{n}\sum_{i=1}^{n}\{\epsilon_i(c'X_i)^2X_i' - \gamma_0'\}(\hat{\beta} - \beta_0)\| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ . \quad (26)$$

Next, exploit parts (i) and (ii) of the Lemma and argue as in (25) to additionally conclude that:

$$P(|\gamma_0'(\hat{\beta} - \beta_0) - \frac{1}{n}\sum_{i=1}^{n}\gamma_0' X_i\epsilon_i| > n^{-\alpha}) \leq P(\|\gamma_0\|\|H_n^{-1} - I\|_o\|\frac{1}{n}\sum_{i=1}^{n} X_i\epsilon_i\| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ . \quad (27)$$

Hence, by results (26), (27), X bounded a.s. and part (iii) of the Lemma we establish that:

$$P(|c'\Sigma_n(\hat{\beta})c - c'\Sigma_n(\beta_0)c + \frac{2}{n}\sum_{i=1}^{n}\gamma_0'X_i\epsilon_i| > n^{-\alpha})$$
$$= P(|\frac{2}{n}\sum_{i=1}^{n}\gamma_0'X_i\epsilon_i - \frac{2}{n}\sum_{i=1}^{n}(c'X_i)^2\epsilon_i X_i'(\hat{\beta}-\beta_0) + \frac{1}{n}\sum_{i=1}^{n}(c'X_i)^2(X_i'(\hat{\beta}-\beta_0))^2| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ . \quad (28)$$

To conclude, note that by direct manipulations we obtain that:

$$\hat{\sigma}^2 = c'(H_n^{-1} - I)\Sigma_n(\hat{\beta})(H_n^{-1} - I)c + c'\Sigma_n(\hat{\beta})c + 2c'(H_n^{-1} - I)\Sigma_n(\hat{\beta})c \ , \quad (29)$$

and hence the final claim of the Lemma follows from (23), (24), (25) and (28). ■

Lemma A.3. *Let Assumptions 2.1(i)-(iii) hold and L_n be as in (9). Then for any $\alpha \in [0,1)$:*

$$\limsup_{n\to\infty} P(|T_n - L_n| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ .$$

Proof: By a Taylor expansion we obtain for some $\bar{\sigma}^2$ a convex combination of $\hat{\sigma}^2$ and σ^2 that:

$$T_n - L_n = c'\{H_n^{-1} - I - \Delta_n\}\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}X_i\epsilon_i + \frac{(\sigma-\hat{\sigma})}{\hat{\sigma}\sigma}c'\{H_n^{-1} - I\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\epsilon_i$$
$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}c'X_i\epsilon_i\{-\frac{1}{2\sigma^3}(\hat{\sigma}^2 - \hat{\sigma}_R^2 + \frac{2}{n}\sum_{i=1}^{n}\gamma_0'X_i\epsilon_i) + \frac{3}{4\bar{\sigma}^5}(\hat{\sigma}^2-\sigma^2)^2\} \ . \quad (30)$$

To study the right hand side of (30), first observe that Lemma A.2(i) and A.2(ii) imply that:

$$P(|c'\{H_n^{-1} - I - \Delta_n\}\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}X_i\epsilon_i| > n^{-\alpha})$$
$$\leq P(\|c\|\|H_n^{-1} - I - \Delta_n\|_o > \frac{1}{n^{\alpha}\log^2(n)}) + P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\epsilon_i\| > \log^2(n)) = o(n^{-\frac{1}{2}}) \ . \quad (31)$$

Moreover, by identical manipulations but exploiting Lemma A.2(i) and A.2(iv) we can similarly conclude:

$$P(|\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^{n}c'X_i\epsilon_i\{\hat{\sigma}^2 - \hat{\sigma}_R^2 + \frac{2}{n}\sum_{i=1}^{n}\gamma_0'X_i\epsilon_i\}| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ . \quad (32)$$

Next, notice that (X,ϵ) bounded a.s. and Lemma A.1 further imply that:

$$P(|c'(\Sigma_n(\beta_0) - \Sigma_0)c| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \qquad P(|\frac{1}{n}\sum_{i=1}^{n}\gamma_0'X_i\epsilon_i| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \ . \quad (33)$$

Therefore, we obtain from (29) together with (23) and (28) that since $\alpha < 1$ we must have:

$$P(|\hat{\sigma}^2 - \sigma^2| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \ . \quad (34)$$

This implies that $P(|\hat{\sigma} - \sigma| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}})$ and since $\bar{\sigma}$ is a convex combination of σ^2 and $\hat{\sigma}^2$ that $P(\bar{\sigma} > \epsilon) = o(n^{-\frac{1}{2}})$ for any $\epsilon < \sigma$. Hence, exploiting (34) and manipulations as in (31) we can conclude:

$$P(|\frac{(\hat{\sigma}^2-\sigma^2)^2}{\bar{\sigma}^5\sqrt{n}}\sum_{i=1}^{n}c'X_i\epsilon_i| > n^{-\alpha}) \leq P((\hat{\sigma}^2-\sigma^2)^2 > \frac{\epsilon^5}{n^{\alpha}\log^2(n)}) + o(n^{-\frac{1}{2}}) = o(n^{-\frac{1}{2}}) \ . \quad (35)$$

Similarly, for $\epsilon < \sigma$ we can exploit $P(\hat{\sigma} > \epsilon) = o(n^{-\frac{1}{2}})$ and Lemma A.2(i) to obtain:

$$P(|\frac{(\sigma-\hat{\sigma})}{\hat{\sigma}\sigma}c'\{H_n^{-1} - I\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\epsilon_i| > n^{-\alpha}) \leq P(\frac{|\sigma-\hat{\sigma}|\|c\|}{\epsilon^2}\|H_n^{-1} - I\|_o > \frac{1}{n^{\alpha}\log^2(n)}) + o(n^{-\frac{1}{2}})$$
$$\leq P(|\sigma-\hat{\sigma}| > \frac{\epsilon^2}{\|c\|n^{\frac{\alpha}{2}}\log(n)}) + P(\|H_n^{-1} - I\|_o > \frac{1}{n^{\frac{\alpha}{2}}\log(n)}) + o(n^{-\frac{1}{2}}) = o(n^{-\frac{1}{2}}) \ . \quad (36)$$

where the final result follows from Lemma A.2(ii), equation (34) and $\alpha < 1$. The Lemma is then established due to the decomposition in (30) and results (31), (32), (35) and (36). ■

Lemma A.4. *Let $\{A_{in}\}_{i=1}^n$ be a triangular array of $k \times p$ matrices, $\{c_n\}_{i=1}^n$ be a sequence of scalars with $\{A_{in}\}_{i=1}^n$ and $\{c_n\}_{i=1}^n$ measurable functions of $\{Y_i, X_i\}_{i=1}^n$. Suppose Assumptions 2.1(i) and 2.2(i) hold and*

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \|A_{in}\|_F^\omega < \infty \qquad c_n^{-1} = o(n^\alpha) \qquad a.s. \tag{37}$$

for some $\alpha \in [0, \frac{\omega-1}{2\omega})$. Then, for any $g : \mathbf{R} \to \mathbf{R}$ such that $E[g^\omega(W)] < \infty$, it follows that:

$$P^*(\|\frac{1}{n}\sum_{i=1}^n A_{in}\{g(W_i) - E[g(W_i)]\}\|_F > c_n) = o(n^{-\frac{1}{2}}) \qquad a.s. \ .$$

Proof: Let $A_{in}^{(l,j)}$ denote the (l, j) entry of A_{in} and proceed as in equation (15) to conclude that:

$$P^*(\|\frac{1}{n}\sum_{i=1}^n A_{in}\{g(W_i) - E[g(W_i)]\}\|_F > c_n) \leq \sum_{l=1}^k\sum_{j=1}^p P^*(|\frac{kp}{n}\sum_{i=1}^n A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\}| > c_n) \ . \tag{38}$$

Next, apply Markov's inequality and the Marcinkiewicz and Rosenthal inequalities (Lemmas 1.4.13 and 1.5.9 in de la Pena and Gine (1999)) to obtain for some constants C_1 and C_2 that:

$$\sqrt{n}P^*(|\frac{1}{n}\sum_{i=1}^n A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\}| > c_n) \leq \frac{\sqrt{n}}{c_n^\omega}E^*[|\frac{1}{n}\sum_{i=1}^n A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\}|^\omega]$$
$$\leq \frac{\sqrt{n}C_1}{c_n^\omega}E^*[(\frac{1}{n^2}\sum_{i=1}^n (A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\})^2)^{\frac{\omega}{2}}] \leq \frac{\sqrt{n}C_2}{c_n^\omega n^{\frac{\omega}{2}}}(\frac{1}{n}\sum_{i=1}^n (A_{in}^{(l,j)})^2 \mathrm{Var}(g(W_i)))^{\frac{\omega}{2}} \ , \tag{39}$$

where in the final result we have used (37) and $\omega \geq 2$. The claim of the Lemma then follows by (37), (38), (39) and $\alpha \in [0, \frac{\omega-1}{2\omega})$ by hypothesis. ■

Lemma A.5. *Let $(\hat\sigma_s^*)^2 \equiv c' H_n^{-1}\Sigma_n^*(\hat\beta)H_n^{-1}c$ and $\{c_n\}_{i=1}^n$ be measurable scalar-valued functions of $\{Y_i, X_i\}_{i=1}^n$. Let Assumptions 2.1(i)-(ii), 2.2(i) hold and $c_n^{-1} = o(n^\alpha)$ a.s. for some $\alpha \in [0, \frac{\omega-1}{2\omega})$. Then:*

(i) $P^*(\|\hat\beta^* - \hat\beta\| > c_n) = o(n^{-\frac{1}{2}})$ *almost surely.*

(ii) $P^*(|(\hat\sigma^*)^2 - (\hat\sigma_s^*)^2| > c_n^2) = o(n^{-\frac{1}{2}})$ *almost surely.*

(iii) $P^*(|(\hat\sigma_s^*)^2 - \sigma^2| > \epsilon) = o(n^{-\frac{1}{2}})$ *almost surely for any $\epsilon > 0$.*

Proof: Since $\hat\beta \overset{a.s.}{\to} \beta$, (Y, X) are bounded by Assumption 2.1(ii) and $\|H_n^{-1}\|_o \overset{a.s.}{\to} 1$, Lemma A.4 implies:

$$P^*(\|\hat\beta^* - \hat\beta\| > c_n) \leq P^*(\|H_n^{-1}\|_o\|\frac{1}{n}\sum_{i=1}^n X_i(Y_i - X_i\hat\beta)W_i\| > c_n) = o(n^{-\frac{1}{2}}) \qquad a.s. \tag{40}$$

For the second claim of the Lemma, proceed by standard manipulations to obtain the inequalities:

$$P^*(|(\hat\sigma^*)^2 - (\hat\sigma_s^*)^2| > c_n^2)$$
$$= P^*(|c'H_n^{-1}\{\frac{1}{n}\sum_{i=1}^n X_iX_i'(X_i'(\hat\beta^* - \hat\beta))^2 - \frac{2}{n}\sum_{i=1}^n X_iX_i'\epsilon_i^*X_i'(\hat\beta^* - \hat\beta)\}H_n^{-1}c| > c_n^2)$$
$$\leq P^*(\|c\|^2\|H_n^{-1}\|_o^2\{\|\frac{1}{n}\sum_{i=1}^n X_iX_i'(X_i'(\hat\beta^* - \hat\beta))^2\|_o + \|\frac{2}{n}\sum_{i=1}^n X_iX_i'\epsilon_i^*X_i'(\hat\beta^* - \hat\beta)\|_o\} > c_n^2) \ . \tag{41}$$

Since X is bounded a.s., we then obtain from part (i) of the Lemma that for some $K > 0$ we must have:

$$P^*(\|c\|^2\|H_n^{-1}\|_o^2\|\frac{1}{n}\sum_{i=1}^n X_iX_i'(X_i'(\hat\beta^* - \hat\beta))^2\|_o > c_n^2)$$
$$\leq P^*(\|c\|^2\|H_n^{-1}\|_o^2 K\|\hat\beta^* - \hat\beta\|^2 > c_n^2) = o(n^{-\frac{1}{2}}) \qquad a.s. \quad (42)$$

Let $X^{(k)}$ denote the k^{th} coordinate of the vector X. Using $\|\cdot\|_o \leq \|\cdot\|_F$, we can then conclude that:

$$P^*(\|c\|^2\|H_n^{-1}\|_o^2\|\frac{2}{n}\sum_{i=1}^n X_iX_i'\epsilon_i^*X_i'(\hat\beta^* - \hat\beta)\|_o > c_n^2)$$
$$\leq P^*(\|c\|^2\|H_n^{-1}\|_o^2\{\max_{1\leq j\leq d_x, 1\leq k\leq d_x}|\frac{2d_x^2}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(k)}\epsilon_i^*X_i'(\hat\beta^* - \hat\beta)|\} > c_n^2)$$
$$\leq \sum_{j=1}^{d_x}\sum_{k=1}^{d_x} P^*(\|c\|^2\|H_n^{-1}\|_o^2\|\frac{2d_x^2}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(k)}X_i\epsilon_i^*\|\|\hat\beta^* - \hat\beta\| > c_n^2)\ . \quad (43)$$

Moreover, for any (j,k) we can then conclude from Lemma A.4 and part (i) of this Lemma that:

$$P^*(\|c\|^2\|H_n^{-1}\|_o^2\|\frac{2d_x^2}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(k)}X_i\epsilon_i^*\|\|\hat\beta^* - \hat\beta\| > c_n^2)$$
$$\leq P^*(\|\frac{2d_x^2}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(k)}X_i\epsilon_i^*\| > c_n) + P^*(\|c\|^2\|H_n^{-1}\|_o^2\|\hat\beta^* - \hat\beta\| > c_n) = o(n^{-\frac{1}{2}})\ , \quad (44)$$

almost surely. The second claim of the Lemma then follows from (41)-(44).

To conclude, exploit that $\|H_n^{-1}\|_o \stackrel{a.s.}{\to} 1$ and $\hat\sigma^2 \stackrel{a.s.}{\to} \sigma$ together with Lemma A.4 to obtain:

$$P^*(|(\hat\sigma_s^*)^2 - \sigma^2| > \epsilon) \leq P^*(|(\hat\sigma_s^*)^2 - \hat\sigma^2| > \epsilon - |\hat\sigma^2 - \sigma^2|)$$
$$\leq P^*(\|\frac{1}{n}\sum_{i=1}^n X_iX_i'(Y_i - X_i\hat\beta)^2(W_i^2 - 1)\|_F > \frac{\epsilon - |\hat\sigma^2 - \sigma^2|}{\|c\|^2\|H_n^{-1}\|_o^2}) = o(n^{-\frac{1}{2}}) \qquad a.s.\ , \quad (45)$$

which establishes the third and final claim of the Lemma. ■

Lemma A.6. *Let Assumptions 2.1(i)-(ii), 2.2(i), and for* $c \in \mathbf{R}^{d_x}$ *define the following random variables:*

$$T_{s,n}^* \equiv \frac{\sqrt{n}c'}{\hat\sigma_s^*}(\hat\beta^* - \beta) \qquad (\hat\sigma_s^*)^2 \equiv c'H_n^{-1}\Sigma_n^*(\hat\beta)H_n^{-1}c\ . \quad (46)$$

It then follows that $P^*(|T_n^* - T_{s,n}^*| > n^{-\alpha}) = o(n^{-\frac{1}{2}})$ *almost surely for any* $\alpha \in [0, \frac{2\omega-3}{2\omega})$.

Proof: Let $\epsilon < \sigma^2$ and note that parts (ii) and (iii) of Lemma A.5 imply $P^*(\hat\sigma^*\hat\sigma_s^* < \epsilon) = o(n^{-\frac{1}{2}})$ almost surely. For any $\gamma \in [0, \frac{\omega-1}{2\omega})$, part (i) of Lemma A.5 then establishes that:

$$P^*(|T_n^* - T_{s,n}^*| > n^{-\alpha}) \leq P^*(\frac{\sqrt{n}|\hat\sigma^* - \hat\sigma_s^*|}{\hat\sigma^*\hat\sigma_s^*} \times \|c\|\|\hat\beta^* - \hat\beta\| > n^{-\alpha})$$
$$\leq P^*(\sqrt{n}|\hat\sigma^* - \hat\sigma_s^*| > \frac{\epsilon}{n^{\alpha-\gamma}}) + P^*(\|\hat\beta^* - \hat\beta\| > \frac{1}{n^\gamma\|c\|}) + P^*(\hat\sigma^*\hat\sigma_s^* < \epsilon)$$
$$= P^*(\sqrt{n}|\hat\sigma^* - \hat\sigma_s^*| > \frac{\epsilon}{n^{\alpha-\gamma}}) + o(n^{-\frac{1}{2}}) \qquad a.s.\ . \quad (47)$$

Since for any $\alpha \in [0, \frac{2\omega-3}{2\omega})$ we may pick $\gamma \in [0, \frac{\omega-1}{2\omega})$ so that $\alpha - \gamma + \frac{1}{2} \in [0, \frac{\omega-1}{\omega})$, the claim of the Lemma then follows from result (47) and part (ii) of Lemma A.5. ■

Lemma A.7. *Let Assumptions 2.1(i)-(iii), 2.2(i) hold, $e_i \equiv (Y_i - X_i'\hat{\beta})$ and $\hat{\kappa} \equiv \frac{1}{n}\sum_i (c'H_n^{-1}X_i)^3 e_i^3$. Then:*

$$E[L_n] = -\frac{\kappa}{2\sigma^3\sqrt{n}} - \frac{\gamma_1}{\sigma\sqrt{n}} + \frac{2c'\Sigma_0\gamma_0}{\sigma^3\sqrt{n}} \qquad\qquad E^*[L_n^*] = -\frac{E[W^3]\hat{\kappa}}{2\hat{\sigma}^3\sqrt{n}} \ .$$

Proof: We first derive an expression for $E[L_n]$. Note that $E[XX'] = I$ and $E[X\epsilon] = 0$ imply:

$$E[c'\Delta_n \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i] = c'E[\frac{1}{n}\sum_{i=1}^n (I - X_iX_i')\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i] = -\frac{1}{\sigma\sqrt{n}}E[(c'X)X'X\epsilon] \tag{48}$$

due to the i.i.d. assumption. Similarly, exploiting the i.i.d. assumption and $E[(c'X)\epsilon] = E[\Delta_n] = 0$ yields:

$$\begin{aligned} E[\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i(\hat{\sigma}_R^2 - \sigma^2)] &= E[\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i\{c'(\Sigma_n(\beta_0) - \Sigma_0)c + 2c'\Delta_n\Sigma_0 c\}] \\ &= \frac{1}{2\sigma^3\sqrt{n}}\{E[(c'X)^3\epsilon^3] - 2E[\epsilon(c'X)^2X']\Sigma_0 c\} \ . \end{aligned} \tag{49}$$

The expression for $E[L_n]$ can then be obtained from (48), (49) and by analogous arguments concluding:

$$E[\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i \times \frac{2}{n}\sum_{i=1}^n \gamma_0' X_i\epsilon_i] = \frac{c'\Sigma_0\gamma_0}{\sigma^3\sqrt{n}} \ . \tag{50}$$

In order to compute $E^*[L_n^*]$, observe that $W \perp (Y, X)$ and $E[W^2] = 1$ implies that:

$$E^*[L_n^*] = -\frac{1}{2\hat{\sigma}^3}E^*[\frac{c'H_n^{-1}}{\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i^*\frac{1}{n}\sum_{i=1}^n c'H_n^{-1}X_iX_i'H_n^{-1}ce_i^2(W_i^2 - 1)] = -\frac{E[W^3]\hat{\kappa}}{2\hat{\sigma}^3\sqrt{n}} \ , \tag{51}$$

which establishes the second claim of the Lemma. ■

Lemma A.8. *Under Assumptions 2.1(i)-(iii) and 2.2(i), the second moments of L_n and L_n^* satisfy:*

$$E[L_n^2] = 1 + O(n^{-1}) \qquad\qquad E^*[(L_n^*)^2] = 1 + O_{a.s.}(n^{-1}) \ .$$

Proof: To calculate $E[L_n^2]$, first note that $E[XX'] = I$, $E[X\epsilon] = 0$ and direct calculations yield:

$$\begin{aligned} E[(c'\Delta_n\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n X_i\epsilon_i)^2] &= E[(\frac{c'}{n}\sum_{i=1}^n (I - X_iX_i')\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n X_i\epsilon_i)^2] = \frac{1}{\sigma^2 n^2}E[(c'(I - X_iX_i')(\sum_{k=1}^n X_k\epsilon_k))^2] \\ &+ \frac{(n-1)}{\sigma^2 n^2}E[\{c'(I - X_iX_i)\sum_{k=1}^n X_k\epsilon_k\}\{c'(I - X_jX_j')\sum_{k=1}^n X_k\epsilon_k\}] = O(n^{-1}) \ . \end{aligned} \tag{52}$$

Similarly, exploiting the i.i.d. assumption together with $E[X\epsilon] = 0$ and $E[I - XX'] = 0$ we obtain:

$$\begin{aligned} E[(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n c'X_i\epsilon_i)(c'\Delta_n\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n X_i\epsilon_i)] &= \frac{1}{n^2\sigma^2}E[(\sum_{i=1}^n c'X_i\epsilon_i)(c'\sum_{i=1}^n (I - X_iX_i'))(\sum_{i=1}^n X_i\epsilon_i)] \\ &= \frac{1}{n\sigma^2}E[(c'X\epsilon)(c'X\epsilon - c'XX'X\epsilon)] = O(n^{-1}) \ . \end{aligned} \tag{53}$$

Exploiting identical arguments to (52) on the squares of the remaining terms of L_n and the Cauchy-Schwarz inequality and arguments identical to those in (53) to address cross terms arising from expanding the square, it is then straightforward to establish that:

$$E[L_n^2] = E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'X_i\epsilon_i)^2] + O(n^{-1}) = \frac{c'E[XX'\epsilon^2]c}{\sigma^2} + O(n^{-1}) = 1 + O(n^{-1}) \ . \tag{54}$$

For notational simplicity, let $a_{in} \equiv c' H_n^{-1} X_i$ and set $e_i \equiv (Y_i - X_i'\hat{\beta})$. To compute $E^*[(L_n^*)^2]$, first note that the i.i.d. assumption together with $E^*[(\epsilon_i^*)^4] = e_i^4 E[W_i^4]$, $E^*[(\epsilon_i^*)^2] = e_i^2$ and $E^*[\epsilon_i^*] = 0$ imply that:

$$\frac{1}{\hat{\sigma}^4 n^2} E^*[(\sum_{i=1}^n a_{in}\epsilon_i^*)^2 (\sum_{i=1}^n a_{in}^2\{(\epsilon_i^*)^2 - e_i^2\})] = \frac{1}{\hat{\sigma}^4 n^2}\sum_{i=1}^n a_{in}^4 e_i^4 (E[W^4] - 1) = O_{a.s}(n^{-1}) \ . \tag{55}$$

Next, also note that by direct calculations, $\{W_i\}_{i=1}^n$ being i.i.d. and $E^*[(\epsilon_i^*)^3] = e_i^3 E[W^3]$ we may establish:

$$\begin{aligned}
&\frac{1}{4\hat{\sigma}^6 n^3} E^*[(\sum_{i=1}^n a_{in}\epsilon_i^*)^2 (\sum_{i=1}^n a_{in}^2\{(\epsilon_i^*)^2 - e_i^2\})^2] \\
&= \frac{1}{4\hat{\sigma}^6 n^3}\{\sum_{i=1}^n E^*[a_{in}^2(\epsilon_i^*)^2(\sum_{k=1}^n a_{kn}^2\{(\epsilon_k^*)^2 - e_k^2\})^2] + \sum_{i=1}^n\sum_{j\neq i} E^*[(a_{in}\epsilon_i^*)(a_{jn}\epsilon_j^*)(\sum_{k=1}^n a_{kn}^2\{(\epsilon_k^*)^2 - e_k^2\})^2]\} \\
&= \frac{1}{4\hat{\sigma}^6 n^3}\{\sum_{i=1}^n\sum_{k=1}^n a_{in}^2 a_{kn}^4 E^*[(\epsilon_i^*)^2\{(\epsilon_k^*)^2 - e_k^2\}^2] + 2\sum_{i=1}^n\sum_{j\neq i} a_{in}^3 e_i^3 a_{jn}^3 e_j^3 (E[W^3])^2\} \ .
\end{aligned} \tag{56}$$

Therefore, expanding the square, noting that $\frac{1}{n}\sum_i a_{in}^2 e_i^2 = \hat{\sigma}^2$ and exploiting (55) and (56):

$$E^*[(L_n^*)^2] = \frac{1}{n\hat{\sigma}^2} E^*[(\sum_{i=1}^n a_{in}\epsilon_i^*)^2] + O_{a.s.}(n^{-1}) = 1 + O_{a.s.}(n^{-1}) \ , \tag{57}$$

which establishes the second and final claim of the Lemma. ■

Lemma A.9. *Let Assumptions 2.1(i)-(iii), 2.2(i) hold* $e_i \equiv (Y_i - X_i'\hat{\beta})$ *and* $\hat{\kappa} \equiv \frac{1}{n}\sum_i (c' H_n^{-1} X_i)^3 e_i^3$. *Then:*

$$E[L_n^3] = -\frac{7\kappa}{2\sigma^3\sqrt{n}} - \frac{3\gamma_1}{\sigma\sqrt{n}} + \frac{12 c'\Sigma_0\gamma_0}{\sigma^3\sqrt{n}} + O(n^{-1}) \qquad E^*[(L_n^*)^3] = -\frac{7E[W^3]\hat{\kappa}}{2\hat{\sigma}^3\sqrt{n}} + O_{a.s.}(n^{-1}) \ . \tag{58}$$

Proof: The calculations are cumbersome and for brevity we provide only the essential steps. Define:

$$\Gamma_n \equiv c'\Delta_n \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i - \frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i\{(\hat{\sigma}_R^2 - \sigma^2) - \frac{2}{n}\sum_{i=1}^n \gamma_0' X_i\epsilon_i\} \ . \tag{59}$$

Notice that $L_n = \frac{1}{\sigma\sqrt{n}} c'\sum_i X_i\epsilon_i + \Gamma_n$. Under Assumption 2.1(ii), it can be shown that $E[\Gamma_n^3] = O(n^{-\frac{3}{2}})$ and similarly that $E[(\frac{1}{\sqrt{n}}\sum_i c'X_i\epsilon_i)^3] = O(n^{-\frac{1}{2}})$. Therefore, by direct calculation and Holder's inequality:

$$\begin{aligned}
E[L_n^3] &= E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i)^3] + 3E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i)^2\Gamma_n] + 3E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i)\Gamma_n^2] + E[\Gamma_n^3] \\
&= E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i)^3] + 3E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i)^2\Gamma_n] + O(n^{-1}) \ .
\end{aligned} \tag{60}$$

Hence, we can establish the first claim of the Lemma by analyzing the remaining terms in (60). Note that

$$E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i)^3] = \frac{1}{\sigma^3\sqrt{n}} E[(c'X)^3\epsilon^3] \ , \tag{61}$$

by the i.i.d. assumption and $E[X\epsilon] = 0$. Similarly, by direct calculation we can also obtain the expression:

$$\begin{aligned}
&E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (c'X_i)\epsilon_i)^2 \frac{c'\Delta_n}{\sqrt{n}\sigma}\sum_{i=1}^n X_i\epsilon_i] \\
&\quad = \frac{1}{\sigma^3 n^{\frac{5}{2}}} E[\{\sum_{i=1}^n (c'X_i)^2\epsilon_i^2 + \sum_{i=1}^n (c'X_i)\epsilon_i \sum_{j\neq i}(c'X_j)\epsilon_j\}\sum_{k=1}^n c'(I - X_k X_k')\sum_{l=1}^n X_l\epsilon_l] \\
&\quad = -\frac{c'\Sigma_0 c}{\sigma^3\sqrt{n}} E[(c'X)(X'X)\epsilon] - \frac{2}{\sigma^3\sqrt{n}} E[(c'X)(\gamma_0' X)\epsilon^2] + O(n^{-\frac{3}{2}}) \ .
\end{aligned} \tag{62}$$

By analogous arguments we can compute the remaining terms in $E[(\frac{1}{\sigma\sqrt{n}}\sum_i c'X_i\epsilon_i)^2\Gamma_n]$ and obtain:

$$\frac{1}{2\sigma^5}E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n(c'X_i)\epsilon_i)^3 c'\{\Sigma_n(\beta_0)-\Sigma_0\}c] = \frac{3c'\Sigma_0 c}{2\sigma^5\sqrt{n}}E[(c'X)^3\epsilon^3] + O(n^{-\frac{3}{2}}) \quad (63)$$

$$\frac{1}{\sigma^5}E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n(c'X_i)\epsilon_i)^3\{c'\Delta_n\Sigma_0 c\}] = -\frac{3c'\Sigma_0 c}{\sigma^5\sqrt{n}}\gamma_0'\Sigma_0 c + O(n^{-\frac{3}{2}}) \quad (64)$$

$$\frac{1}{\sigma^5}E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n(c'X_i)\epsilon_i)^3\{\frac{1}{n}\sum_{i=1}^n\gamma_0'X_i\epsilon_i\}] = \frac{3c'\Sigma_0 c}{\sigma^5\sqrt{n}}c'\Sigma_0\gamma_0 + O(n^{-\frac{3}{2}})\ . \quad (65)$$

The first claim of the Lemma then follows by combining the results from (60)-(65).

Letting $a_{in} = c'H_n^{-1}X_i$ and employing Assumption 2.1(ii), it can then be shown that:

$$E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3(\frac{1}{2\hat\sigma^3}\{(\hat\sigma_s^*)^2-\hat\sigma^2\})^2] = O_{a.s.}(n^{-\frac{3}{2}}) \quad (66)$$

$$E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3(\frac{1}{2\hat\sigma^3}\{(\hat\sigma_s^*)^2-\hat\sigma^2\})^3] = O_{a.s.}(n^{-\frac{3}{2}})\ . \quad (67)$$

Therefore, expanding the cube and exploiting that $W \perp (Y,X)$ and $E^*[(\epsilon_i^*)^k] = E[W^k]e_i^k$, it follows that:

$$\begin{aligned} E^*[(L_n^*)^3] &= E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3\{\frac{1}{\hat\sigma^3} - \frac{3((\hat\sigma_s^*)^2-\hat\sigma^2)}{2\hat\sigma^5} + \frac{3((\hat\sigma_s^*)^2-\hat\sigma^2)^2}{4\hat\sigma^7} - \frac{((\hat\sigma_s^*)^2-\hat\sigma^2)^3}{8\hat\sigma^9}\}] \\ &= \frac{E[W_i^3]}{\hat\sigma^3\sqrt{n}}\times\frac{1}{n}\sum_{i=1}^n a_{in}^3 e_i^3 - \frac{3}{2\hat\sigma^5}E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3\{(\hat\sigma_s^*)^2-\hat\sigma^2\}] + O_{a.s}(n^{-\frac{3}{2}})\ . \end{aligned} \quad (68)$$

Moreover, also note that by analogous arguments and direct calculations we further obtain:

$$\begin{aligned} &E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3\{\frac{3}{2\hat\sigma^5 n}\sum_{i=1}^n a_{in}^2\{(\epsilon_i^*)^2-e_i^2\}\}] \\ &= \frac{3}{2\hat\sigma^5 n^{\frac{3}{2}}}\times\frac{1}{n}\sum_{i=1}^n a_{in}^5E^*[(\epsilon_i^*)^3\{(\epsilon_i^*)^2-e_i^2\}] + \frac{9}{2\hat\sigma^5 n^{\frac{5}{2}}}E^*[\{\sum_{i=1}^n a_{in}(\epsilon_i^*)\sum_{j\neq i}a_{jn}^2(\epsilon_j^*)^2\}\sum_{k=1}^n a_{kn}^2\{(\epsilon_k^*)^2-e_i^2\}] \\ &= \frac{9}{2\hat\sigma^5\sqrt{n}}\times\frac{1}{n}\sum_{i=1}^n a_{in}^2e_i^2\times\frac{E[W^3]}{n}\sum_{i=1}^n a_{in}^3e_i^3 + O_{a.s.}(n^{-\frac{3}{2}})\ . \end{aligned} \quad (69)$$

The second claim of the Lemma is then established by (68) and (69). ■

Proof of Theorem 2.1: The first claim of the Theorem is an immediate consequence of Lemma A.3. For the second claim, note that in lieu of Lemma A.6, it suffices to show that $T_{n,s}^* = L_n^* + o_{p^*}(n^{-\frac{1}{2}})$ a.s.. For notational simplicity, let $a_{in} = c'H_n^{-1}X_i(Y_i - X_i'\hat\beta)$ and apply Markov's inequality to conclude that:

$$\begin{aligned} P^*(|(\hat\sigma_s^*)^2-\hat\sigma^2| > \frac{C}{\sqrt{n}}) &= P^*(|\frac{1}{n}\sum_{i=1}^n a_{in}^2(W_i^2-1)| > \frac{C}{\sqrt{n}}) \\ &\leq \frac{n}{C^2}E^*[(\frac{1}{n}\sum_{i=1}^n a_{in}^2(W_i^2-1))^2] = \frac{1}{C^2 n}\sum_{i=1}^n a_{in}^4E[(W_i^2-1)^2]\ . \end{aligned} \quad (70)$$

However, under our moment assumptions, $\frac{1}{n}\sum_i a_{in}^4E[(W_i^2-1)^2] \overset{a.s.}{\to} E[(c'X)^4\epsilon_i^4]E[(W^2-1)^2] < \infty$, and therefore from (70) it follows that $(\hat\sigma_s^*)^2 = \hat\sigma^2 + O_{p^*}(n^{-\frac{1}{2}})$ almost surely. The second claim of the Lemma then follows from a second order Taylor expansion. ■

Proof of Theorem 2.2: Follows immediately from Lemmas A.7, A.8, A.9 and direct calculation. ■

APPENDIX B - Proofs of Theorem 2.3

Lemma B.1. *Let Assumption 2.1(i)-(iv) hold and L_n be as in* (9) *with $c \neq 0$. Then, uniformly in $z \in \mathbf{R}$:*

$$P(L_n \leq z) = \Phi(z) + \frac{\phi(z)\kappa}{6\sigma^3\sqrt{n}}(2z^2+1) - \frac{\phi(z)}{\sigma^3\sqrt{n}}(c'\Sigma_0\gamma_0(z^2+1) - \gamma_1\sigma^2) + o(n^{-\frac{1}{2}}) \ .$$

Proof: Letting $Z \equiv (X'\epsilon, \text{vech}(XX')', \text{vech}(XX'\epsilon^2)')'$, it is clear that L_n is a smooth functional of $\frac{1}{n}\sum_i Z_i$ and that Z satisfies Cramer's condition by Assumption 2.1(iv). The claim of the Lemma then follows from Theorem 2.2 in Hall (1992) and Theorem 2.2. ■

Lemma B.2. *Let $\{a_{in}\}_{i=1}^n$ be a triangular array of measurable scalar valued functions of $\{Y_i, X_i\}_{i=1}^n$ and define $V_{in} \equiv (a_{in}W_i, a_{in}^2(W_i^2-1))'$, $\Omega_n \equiv \frac{1}{n}\sum_i E^*[V_{in}V_{in}']$ and $S_n \equiv \frac{1}{\sqrt{n}}\sum_i \Omega_n^{-\frac{1}{2}}V_{in}$. Suppose Assumptions 2.2(i)-(ii) hold and (i) $\Omega_n \overset{a.s.}{\to} \Omega$ with Ω full rank, (ii) $\limsup_{n\to\infty}\max_{1\leq i\leq n}|a_{in}| < \infty$ a.s. and (iii) For $K_n(\epsilon) \equiv \#\{i : \min\{|a_{in}|, a_{in}^2\} \geq \epsilon\}$, there a.s. exists an ϵ_0 such that $K_n(\epsilon_0)/\log(n) \uparrow \infty$. Then:*

$$P^*(S_n \in B) = \sum_{j=0}^{1}\int_B dP_j(-\Phi : \{\mathcal{X}_k^*(S_n)\}) + o(n^{-\frac{1}{2}}) \qquad a.s.$$

uniformly over all Borel sets B with $\Phi((\partial B)^\epsilon) \leq C\epsilon$ for some constant C, $(\partial B)^\epsilon$ the ϵ enlargement of ∂B, $\mathcal{X}_k^*(S_n)$ the k^{th} cumulant of S_n under P^* and P_j the Cramer-Edgeworth measures.

Proof: We proceed by verifying the conditions of Theorem 3.4 in Skovgaard (1986). For $t \in \mathbf{R}^2$, define:

$$\rho_n(t) \equiv \frac{1}{3!\|t\|^3}|\mathcal{X}_3^*(t'S_n)| = \frac{1}{3!\|t\|^3}|E^*[(t'S_n)^3]| \ , \tag{71}$$

since $E[W] = 0$, $E[W^2] = 1$ and $W \perp (Y, X)$. Hence, by Cauchy-Schwartz and convexity we obtain:

$$\rho_n(t) \leq \frac{1}{n^{\frac{3}{2}}\|t\|^3}\sum_{i=1}^n E^*[|t'\Omega_n^{-\frac{1}{2}}V_{in}|^3] \leq \frac{\|\Omega_n^{-\frac{1}{2}}\|_o^3}{n^{\frac{3}{2}}}\sum_{i=1}^n E^*[\|V_{in}\|^3]$$
$$\leq \frac{4\|\Omega_n^{-\frac{1}{2}}\|_o^3}{n^{\frac{3}{2}}}\sum_{i=1}^n\{E^*[|a_{in}|^3|W_i|^3] + E^*[a_{in}^6|W_i^2-1|^3]\} \ . \tag{72}$$

Note that $\Omega_n \overset{a.s.}{\to} \Omega$ with Ω full rank by hypothesis, implies $\|\Omega_n^{-\frac{1}{2}}\|_o \overset{a.s.}{\to} \|\Omega^{-\frac{1}{2}}\|_o < \infty$. Moreover, since $\{a_{in}\}_{i=1}^n$ is not random with respect to P^*, we obtain from condition (ii) and result (72) that almost surely:

$$\limsup_{n\to\infty}\{\sup_{t\in\mathbf{R}^2}\sqrt{n}\rho_n(t)\} \leq \limsup_{n\to\infty}\{4\|\Omega_n^{-\frac{1}{2}}\|_o^3(E[|W|^3] + E[|W^2-1|^3]) \times \max_{1\leq i\leq n}\{|a_{in}|^3 + a_{in}^6\}\} < \infty \ . \tag{73}$$

Therefore, we conclude that almost surely there exists a sequence $\{r_n\}$ satisfying the following:

$$\sup_{t\in\mathbf{R}^2}\rho_n(t) \leq \frac{1}{r_n} \qquad\qquad r_n \asymp \sqrt{n} \ , \tag{74}$$

which verifies conditions (I) and (II) of Theorem 3.4 in Skovgaard (1986).

Next, let $\xi_n^*(t) \equiv E^*[\exp(it'S_n)]$. We aim to show that almost surely there exists a $\delta > 0$ such that:

$$\limsup_{n\to\infty}\{\sup_{0<h<\delta r_n, t\in\mathbf{R}^2}|\frac{d^4}{dh^4}\log(\xi_n^*(\frac{th}{\|t\|}))| \times r_n^2\} < \infty \ . \tag{75}$$

Towards this end, define $\xi_{in}^*(t) \equiv E^*[\exp(it'\Omega_n^{-\frac{1}{2}}V_{in}/\sqrt{n})]$. By Corollary 8.2 in Bhattacharya and Rao (1976), $\{a_{in}\}_{i=1}^n$ being nonrandom with respect to P^* and direct calculation it then follows that:

$$|\xi_{in}^*(t) - 1| \leq \frac{\|t\|^2}{2n}E^*[\|\Omega_n^{-\frac{1}{2}}V_{in}\|^2] \leq \frac{\|t\|^2\|\Omega_n^{-\frac{1}{2}}\|_o^2}{2n}E[W^2 + (W^2-1)^2] \times \max_{1\leq i\leq n}(a_{in}^2 + a_{in}^4) \ . \tag{76}$$

Condition (ii), $\|\Omega_n^{-\frac{1}{2}}\|_o \stackrel{a.s.}{\to} \|\Omega^{-\frac{1}{2}}\|_o < \infty$ and $r_n \asymp \sqrt{n}$ then imply that almost surely there is a $\delta > 0$ with:

$$\limsup_{n\to\infty}\{\sup_{\|t\|\le\delta r_n}|\xi_{in}^*(t)-1|\} \le \frac{\delta E[W^2+(W^2-1)^2]}{2} \times \limsup_{n\to\infty}\{\frac{r_n^2\|\Omega_n^{-\frac{1}{2}}\|_o^2}{n}\{\max_{1\le i\le n}(a_{in}^2+a_{in}^4)\}\} < \frac{1}{2} . \quad (77)$$

Since $\xi_n^*(t) = \prod_i \xi_{in}^*(t)$ by the i.i.d. assumption and $W \perp (Y, X)$ we obtain by direct calculation:

$$\limsup_{n\to\infty}\{\sup_{0<h<\delta r_n, t\in\mathbf{R}^2}|\frac{d^4}{dh^4}\log(\xi_n^*(\frac{th}{\|t\|}))| \times r_n^2\} \le \limsup_{n\to\infty}\{\sup_{0<h<\delta r_n, t\in\mathbf{R}^2} r_n^2\sum_{i=1}^n|\frac{d^4}{dh^4}\log(\xi_{in}^*(\frac{th}{\|t\|}))|\}$$

$$\le \limsup_{n\to\infty}\{\sup_{\|t\|\le\delta r_n} r_n^2\sum_{i=1}^n\sum_{|\lambda|=4}|D^\lambda\log(\xi_{in}^*(t))|\} \le \limsup_{n\to\infty}\{16r_n^2\sum_{i=1}^n E^*[\|\frac{\Omega_n^{-\frac{1}{2}}V_{in}}{\sqrt{n}}\|^4]\} , \quad (78)$$

where the final inequality holds by Lemma 9.4 in Bhattacharya and Rao (1976) and result (77) implying $|\xi_{in}^*(t)-1| < \frac{1}{2}$ for all $\|t\| \le \delta r_n$ and all $1 \le i \le n$ for n large enough. Moreover,

$$\limsup_{n\to\infty}\{r_n^2\sum_{i=1}^n E^*[\|\frac{\Omega_n^{-\frac{1}{2}}V_{in}}{\sqrt{n}}\|^4]\} \le \limsup_{n\to\infty}\{\frac{r_n^2}{n}\times\frac{2\|\Omega_n^{-\frac{1}{2}}\|_o^4}{n}\sum_{i=1}^n\{a_{in}^4E[W^4]+a_{in}^8E[(W^2-1)^4]\}\} < \infty \quad (79)$$

almost surely, by condition (i), (ii) and (74). It follows from (78) and (79) that (75) holds almost surely, which verifies condition (IV) of Theorem 3.4 in Skovgaard (1986).

To conclude, we aim to show that almost surely for any $\delta > 0$ it follows that:

$$\limsup_{n\to\infty}\{r_n^6 \times \sup_{\delta r_n\le\|t\|}|\xi_n^*(t)|\} < \infty . \quad (80)$$

Let ξ_U denote the characteristic function of $U \equiv (W, W^2-1)'$, $\eta(\epsilon) \equiv \sup_{\|t\|\ge\epsilon}|\xi_U(t)|$ and define:

$$A_{in} \equiv \begin{pmatrix} a_{in} & 0 \\ 0 & a_{in}^2 \end{pmatrix} . \quad (81)$$

Since $\Omega_n^{-\frac{1}{2}}$, A_{in} are not random with respect to P^* and $W \perp (Y, X)$ we then obtain by direct calculation:

$$\sup_{\delta r_n\le\|t\|}|\xi_n^*(t)| = \sup_{\delta r_n\le\|t\|}\prod_{i=1}^n|\xi_{in}^*(t)| = \sup_{\delta r_n\le\|t\|}\prod_{i=1}^n|\xi_U(\frac{A_{in}\Omega_n^{-\frac{1}{2}}}{\sqrt{n}}t)| \le \{\eta(\epsilon)\}^{\#\{i:\|A_{in}\Omega_n^{-\frac{1}{2}}t\|\ge\epsilon\sqrt{n}\ \forall\|t\|\ge\delta r_n\}} \quad (82)$$

for any $\epsilon > 0$. Moreover, since the smallest eigenvalue of $\Omega_n^{-\frac{1}{2}}$ equals $\|\Omega_n\|_o^{-\frac{1}{2}}$, we also have:

$$\#\{i : \|A_{in}\Omega_n^{-\frac{1}{2}}t\| \ge \epsilon\sqrt{n}\ \forall\|t\| \ge \delta r_n\} \ge \#\{i : \min\{|a_{in}|, a_{in}^2\} \ge \frac{\epsilon\sqrt{n}\|\Omega_n\|_o^{\frac{1}{2}}}{\delta r_n}\} . \quad (83)$$

Thus, as $\|\Omega_n\|_o^{\frac{1}{2}} \stackrel{a.s.}{\to} \|\Omega\|_o^{\frac{1}{2}} < \infty$ and $r_n \asymp \sqrt{n}$ we may almost surely pick ϵ^* such that $\epsilon^*\sqrt{n}\|\Omega_n\|_o^{\frac{1}{2}}/\delta r_n < \epsilon_0$ for n sufficiently large. In addition, by Assumption 2.2(ii), $\eta(\epsilon^*) < 1$; see page 207 in Bhattacharya and Rao (1976). Hence, by result (82) and condition (iii) we conclude that almost surely:

$$\limsup_{n\to\infty}\{r_n^6 \times \sup_{\delta r_n\le\|t\|}|\xi_n^*(t)|\} \le \limsup_{n\to\infty} r_n^6\eta(\epsilon^*)^{K_n(\epsilon_0)} = 0 , \quad (84)$$

verifying Condition (III") of Theorem 3.4 in Skovgaard (1986). The claim of the Lemma therefore follows by direct application of Theorem 3.4 in Skovgaard (1986). ■

Lemma B.3. *Suppose Assumptions 2.1(i)-(iv) and 2.2(i)-(ii) hold and let* $c \neq 0$, $T_{s,n}^* \equiv \sqrt{n}c'(\hat\beta^* - \hat\beta)/\hat\sigma_s^*$ *where* $(\hat\sigma_s^*)^2 \equiv c'H_n^{-1}\Sigma_n^*(\hat\beta)H_n^{-1}c$. *It then follows that almost surely, uniformly in* $z \in \mathbf{R}$*:*

$$P^*(T_{s,n}^* \le z) = \Phi(z) + \frac{\phi(z)\hat\kappa E[W^3]}{6\hat\sigma^3\sqrt{n}}(2z^2+1) + o(n^{-\frac{1}{2}}) . \quad (85)$$

Proof: We proceed by verifying the conditions of Theorem 3.2 in Skovgaard (1981). First, define:

$$a_{in} \equiv c'H_n^{-1}X_i(Y_i - X_i\hat{\beta}) \qquad\qquad a_i \equiv c'X_i(Y_i - X_i\beta_0) \ . \tag{86}$$

Since $\hat{\beta} \stackrel{a.s.}{\to} \beta_0$, $\|H_n^{-1} - I\|_o \stackrel{a.s.}{\to} 0$ and (X,ϵ) is bounded a.s. by Assumption 2.1(ii), we obtain:

$$\limsup_{n\to\infty}\{\max_{1\le i\le n}|a_{in} - a_i|\}$$
$$\le \limsup_{n\to\infty}\{\|c\|\|H_n^{-1} - I\|_o \max_{1\le i\le n}\|X_i\epsilon_i\|\} + \limsup_{n\to\infty}\{\|c\|\|H_n^{-1}\|_o\|\hat{\beta} - \beta_0\|\max_{1\le i\le n}\|X_i\|^2\} = 0 \ . \tag{87}$$

Let $V_{in} \equiv (a_{in}W_i, a_{in}^2(W_i^2 - 1))'$ and $V_i \equiv (a_iW_i, a_i^2(W_i^2-1))'$. By result (87), it then follows that:

$$\Omega_n \equiv \frac{1}{n}\sum_{i=1}^n E^*[V_{in}V_{in}'] \stackrel{a.s.}{\to} E[VV'] \ . \tag{88}$$

Assumption 2.2(ii) rules out Rademacher weights, which are the only ones satisfying $E[W] = 0$ and $P(W^2 = 1) = 1$. By Assumption 2.1(iii), $W \perp (Y,X)$, $c \neq 0$ and W not being Rademacher, it is then possible to show $E[VV']$ is full rank. Next, pick a δ_0 such that:

$$P(\min\{|(c'X)\epsilon|, (c'X)^2\epsilon^2\} \ge \delta_0) > 0 \ , \tag{89}$$

which is possible since $E[(c'X)^2\epsilon^2] > 0$ by Assumption 2.1(iii) and $c \neq 0$. By result (87), then:

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n 1\{\min\{|a_{in}|, a_{in}^2\} \ge \frac{\delta_0}{2}\} \ge \liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n 1\{\min\{|a_i|, a_i^2\} \ge \delta_0\} > 0 \qquad a.s. \ . \tag{90}$$

Defining $S_n \equiv \frac{1}{\sqrt{n}}\sum_i \Omega_n^{-\frac{1}{2}}V_{in}$, (88), (87) with Assumption 2.1(ii) and (90) verify conditions(i)-(iii) of Lemma B.2 respectively. Therefore, we can conclude that almost surely:

$$P^*(S_n \in B) = \sum_{j=0}^1 \int_B dP_j(-\Phi : \{\mathcal{X}_k^*(S_n)\}) + o(n^{-\frac{1}{2}}) \tag{91}$$

uniformly over all Borel sets B with $\Phi((\partial B)^\epsilon) \le C\epsilon$ for some constant C. This verifies condition (3.1) of Theorem 3.2 in Skovgaard (1981).

Next, let $t^{(i)}$ denote the i^{th} coordinate of $t \in \mathbf{R}^2$ and define the functions $g_n, f_n : \mathbf{R}^2 \to \mathbf{R}$ by:

$$f_n(t) \equiv g_n(\Omega_n^{\frac{1}{2}}t) \qquad\qquad g_n(t) \equiv t^{(1)} \times (\frac{t^{(2)}}{\sqrt{n}} + \hat{\sigma}_n^2)^{-\frac{1}{2}} \ . \tag{92}$$

Note that by construction, $f_n(S_n) = T_{s,n}^*$, $f_n(0) = 0$ and $\|Df_n(0)\| = 1$. Further, define the set:

$$\Gamma_n \equiv \{t \in \mathbf{R}^2 : \|t\| \le \log(n)\} \ . \tag{93}$$

The functions g_n are differentiable everywhere except at $t \in \mathbf{R}^2$ with $t^{(2)} = -\hat{\sigma}_n^2\sqrt{n}$. However, since $\hat{\sigma}_n^2 \stackrel{a.s.}{\to} \sigma^2$ and $\|\Omega_n^{\frac{1}{2}}\|_o \stackrel{a.s.}{\to} \|\Omega^{\frac{1}{2}}\|_o$ we obtain that almost surely for n sufficiently large, f_n is differentiable on Γ_n. Moreover, since a.s. for n large enough $\|\Omega_n^{-\frac{1}{2}}\|_o\log(n)/\sqrt{n} \le \hat{\sigma}_n^2/2$ we obtain by direct calculation:

$$\limsup_{n\to\infty}\{\sqrt{n}\sup_{t\in\Gamma_n}\sup_{|\lambda|=3}|D^\lambda f_n(t)|\} \le \limsup_{n\to\infty}\{4\sqrt{n}\|\Omega_n^{\frac{1}{2}}\|_F^3 \times \max\{\frac{3}{4n}\times\frac{2^{\frac{5}{2}}}{\hat{\sigma}_n^5}, \frac{15\|\Omega_n^{\frac{1}{2}}\|_o\log(n)}{8n^{\frac{3}{2}}}\times\frac{2^{\frac{7}{2}}}{\hat{\sigma}_n^7}\} = 0 \tag{94}$$

almost surely; which verifies condition (3.11) of Theorem 3.2 in Skovgaard (1981). Similarly,

$$\limsup_{n\to\infty}\sqrt{n}\|\nabla^2 f_n(0)\|_F^2 = \limsup_{n\to\infty}\sqrt{n}\|\Omega_n^{\frac{1}{2}}\nabla^2 g_n(0)\Omega_n^{\frac{1}{2}}\|_F^2 \le \limsup_{n\to\infty}\{\sqrt{n}\|\Omega_n^{\frac{1}{2}}\|_F^2 \times \frac{1}{2n\hat{\sigma}_n^6}\} = 0 \tag{95}$$

almost surely, verifying condition (3.12) of Theorem 3.2 in Skovgaard (1981). Therefore, we conclude from (91), (94), (95), Theorem 3.2 and Remark 3.4 in Skovgaard (1981) that an Edgeworth expansion for $P^*(T^*_{s,n} \in B)$ holds almost surely for all sets B such that $\Phi((\partial B)^\epsilon) = O(\epsilon)$ (which includes all sets of the form $(-\infty, z]$)). In particular, (85) holds by Theorem 3.2 in Skovgaard (1981) and Theorem 2.2. ■

Proof of Theorem 2.3: The first claim of the Theorem follows from Lemma B.1, Lemma A.3 and Lemma 5(a) in Andrews (2002) while the second claim follows by Lemma B.3, Lemma A.6 and Lemma 5(a) in Andrews (2002). ■

References

Andrews, D. W. K. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica*, **70** 119–162.

Bhatia, R. (1997). *Matrix Analysis.* Springer, New York.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal edgeworth expansion. *The Annals of Statistics*, **6** 434–451.

Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions.* John Wiley & Sons, New York.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, **90** 414–427.

Cavaliere, G. and Taylor, A. M. R. (2008). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory*, **24** 43–71.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, **146** 162–169.

Davidson, R. and MacKinnon, J. G. (2010). Wild bootstrap tests for iv regression. *Journal of Business and Economic Statistics*, **28** 128–144.

de la Pena, V. H. and Gine, E. (1999). *Decoupling: From Dependence to Independence.* Springer-Verlag, New York.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, **9** 1218–1228.

Gonçalves, S. and Meddahi, N. (2009). Bootstrapping realized volatility. *Econometrica*, **77** 283–306.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

Horowitz, J. L. (1997). Bootstrap methods in econometrics: Theory and numerical performance. In *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress* (D. M. Kreps and K. F. Wallis, eds.), vol. 3. Cambridge University Press.

Horowitz, J. L. (2001). The bootstrap. In *Handbook of Econometrics* (J. J. Heckman and E. Leamer, eds.), vol. 5, chap. 52. Elsevier.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, **16** 1696–1708.

Liu, R. Y. and Singh, K. (1987). On a partial correction by the bootstrap. *The Annals of Statistics*, **15** 1713–1718.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, **21** 255–285.

Skovgaard, I. M. (1981). Transformation of an edgeworth expansion by a sequence of smooth functions. *Sandinavian Journal of Statistics*, **8** 207–217.

Skovgaard, I. M. (1986). On multivariate edgeworth expansions. *International Statistical Review*, **54** 169–186.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50** 1–25.

Wu, C. F. J. (1986). Jacknife, bootstrap, and other resampling methods in regression analysis. *Annals of Statistics*, **14** 1261–1295.