Higher Order Properties of the Wild Bootstrap Under Misspecification

Patrick Kline

Department of Economics

UC Berkeley / NBER

Andres Santos*

Department of Economics

UC San Diego

First Version: January, 2011

This Version: March, 2012

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. Nonetheless, we show that the wild bootstrap still has a lower maximal asymptotic risk as an estimator of the true distribution than a normal approximation, in shrinking neighborhoods of properly specified models. To assess the practical implications of this result, we conduct a Monte Carlo study contrasting the performance of the wild bootstrap with a normal approximation, and the traditional nonparametric bootstrap.

Keywords: Wild bootstrap, Misspecification, Edgeworth Expansion.

JEL Code: C12.

*Corresponding author. Address: 9500 Gilman Drive MS 0508, La Jolla, CA 92093-0508. Tel: (858) 534-2407. Fax: (858) 534-7040. Email: 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 heteroscedastic structure of unknown form. This robustness to arbitrary heteroscedasticity provides the wild bootstrap with a distinct advantage over the residual bootstrap of Freedman (1981) which requires homoscedastic errors. Moreover, theoretical results from Mammen (1993) indicate the wild bootstrap outperforms the nonparametric bootstrap when a large number of regressors are present and the errors obey the mean independence restriction $E[\epsilon|X] = 0$. These properties have led to increasing attention among economists concerned with heteroscedasticity 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), Kline and Santos (2011)). To date, however, the higher order properties of the wild bootstrap have only been studied under the assumption of proper model specification, where the errors are mean independent of the regressors. Liu (1988) first established that when this condition holds the wild bootstrap provides a refinement over a normal approximation.

Since the seminal work of White (1980a,b, 1982), economists have sought inference procedures robust to the possibilities of both unmodeled heteroscedasticity and misspecification (see Stock (2010) for a recent retrospective). In an important contribution, Mammen (1993) established that the wild bootstrap exhibits a form of robustness, showing that it remains consistent in the absence of proper model specification. In this paper, we contribute to the literature by examining whether, in addition to remaining consistent, the wild bootstrap continues to provide a refinement over the normal approximation under misspecification. Concretely, we study the higher order properties of the wild bootstrap when ϵ is uncorrelated with X but not necessarily mean independent of it – a setting commonly encountered in economics where parametric modeling is pervasive. It is precisely in such misspecified environments that heteroscedasticity is likely to arise making the higher order properties of the wild bootstrap of particular interest (White (1982)).

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. Despite this discordance, we establish that the wild bootstrap still possesses a lower asymptotic risk as an estimator of the true distribution of studentized test statistics than a normal approximation in shrinking neighborhoods of properly specified models. Heuristically, these results suggest the wild bootstrap should outperform a normal approximation provided misspecification is not "too severe." To assess the practical implications of this result, we conclude by conducting a Monte Carlo study contrasting the performance of the wild bootstrap with that of a normal approximation and 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 wild bootstrap distribution of T_n^* can provide size control asymptotically. In what follows, we explore whether such a procedure is additionally able to provide a refinement over the standard normal approximation.

2.1 Assumptions

In model (1), the regression can be made to include a constant by setting one of the components of the vector X to equal one almost surely. Because such a setting will require special care in our notation, we let $X = (1, \tilde{X}')'$ if X contains a constant and set $\tilde{X} = X$ otherwise. Throughout, for a matrix A, we also let $\|\cdot\|_F$ denote the Frobenius norm $\|A\|_F^2 \equiv \text{trace}\{A'A\}$. Given this notation, we introduce the following assumptions on the data generating process:

Assumption 2.1. *(i) $\{Y_i, X_i\}_{i=1}^n$ is i.i.d., satisfying* (1) *with $E[X\epsilon] = 0$; (ii) $E[\|XX'\|_F^\nu] < \infty$ and $E[\|XX'\epsilon^2\|_F^\nu] < \infty$ for some $\nu \geq 9$; (iii) $E[XX'] \equiv H_0$ and $\Sigma_0 \equiv E[XX'\epsilon^2]$ are full rank; (iv) For $Z \equiv (\tilde{X}', X'\epsilon, vech(\tilde{X}\tilde{X}')', vech(XX'\epsilon^2)')'$, 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 demand the existence of certain higher order moments of (Y, X) so that the appropriate approximate cumulants of T_n are finite. The requirements on the weights $\{W_i\}_{i=1}^n$ 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)). Unfortunately, this requirement rules out two frequently used wild bootstrap weights: Rademacher random variables and a weighting scheme originally proposed in Mammen (1993). Thus, while our results on approximate cumulants are applicable to these choices of weights, our results on Edgeworth expansions are not.

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:

$$\sigma^2 \equiv c'H_0^{-1}\Sigma_0 H_0^{-1}c \qquad \kappa \equiv E[(c'H_0^{-1}X)^3\epsilon^3]$$
$$\gamma_0 \equiv E[(c'H_0^{-1}X)^2 X\epsilon] \qquad \gamma_1 \equiv E[(c'H_0^{-1}X)(X'H_0^{-1}X)\epsilon] \tag{6}$$

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

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:*

$$L_n \equiv c'\{H_0^{-1} + H_0^{-1}\Delta_n H_0^{-1}\}\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} X_i\epsilon_i - \frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^{n}(c'H_0^{-1}X_i)\epsilon_i\{(\hat{\sigma}_R^2 - \sigma^2) - \frac{2}{n}\sum_{i=1}^{n}\gamma_0' H_0^{-1}X_i\epsilon_i\}$$

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

where $\Delta_n \equiv H_0 - H_n$, $\hat{\sigma}_R^2 \equiv c'H_0^{-1}\Sigma_n(\beta_0)H_0^{-1}c + 2c'H_0^{-1}\Delta_n H_0^{-1}\Sigma_0 H_0^{-1}c$ and $(\hat{\sigma}_s^*)^2 \equiv c'H_n^{-1}\Sigma_n^*(\hat{\beta})H_n^{-1}c$. It then follows that $T_n = L_n + o_p(n^{-\frac{1}{2}})$ and $T_n^* = L_n^* + o_{p^*}(n^{-\frac{1}{2}})$ almost surely.

Recall that in Assumption 2.1(ii) the matrix H_0 was defined to equal $E[XX']$. Therefore $\Delta_n \equiv H_0 - H_n$ is the estimation error in the Hessian and the first term in the definition of L_n 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) $\frac{2}{n}\sum_i \gamma_0' H_0^{-1}X_i\epsilon_i$ which reflects the 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 true cumulants of T_n and T_n^* for computing their second order Edgeworth expansions, whenever 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:*

$$\mathcal{X}_1(L_n) = -\frac{\kappa}{2\sigma^3\sqrt{n}} - \frac{\gamma_1}{\sigma\sqrt{n}} + \frac{2c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0}{\sigma^3\sqrt{n}} \qquad \mathcal{X}_1^*(L_n^*) = -\frac{E[W^3]\hat{\kappa}}{2\hat{\sigma}^3\sqrt{n}}$$

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

$$\mathcal{X}_3(L_n) = -\frac{2\kappa}{\sigma^3\sqrt{n}} + \frac{6c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0}{\sigma^3\sqrt{n}} + O(n^{-1}) \qquad \mathcal{X}_3^*(L_n^*) = -\frac{2E[W^3]\hat{\kappa}}{\hat{\sigma}^3\sqrt{n}} + O_{a.s.}(n^{-1}) \ .$$

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 when regressors and errors are merely uncorrelated. 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]

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:*

$$P(T_n \leq z) = \Phi(z) + \frac{\phi(z)\kappa}{6\sigma^3\sqrt{n}}(2z^2+1) - \frac{\phi(z)}{\sigma^3\sqrt{n}}(c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0(z^2+1) - \gamma_1\sigma^2) + o(n^{-\frac{1}{2}}) \tag{7}$$

$$P^*(T_n^* \leq z) = \Phi(z) + \frac{\phi(z)\hat{\kappa}E[W^3]}{6\hat{\sigma}^3\sqrt{n}}(2z^2+1) + o(n^{-\frac{1}{2}}) \qquad a.s. \tag{8}$$

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 some

[2] It is interesting to note that even under misspecification, if $c \neq 0$ solves $c'H_0^{-1}\Sigma_0 H_0^{-1}E[(c'H_0^{-1}X)^2 X\epsilon] = 0$, and $c'H_0^{-1}E[XX'H_0^{-1}X\epsilon] = 0$, then the approximate cumulants of the full sample and wild bootstrap statistics will still match. It seems unlikely, however, that a c of interest to the researcher would happen to satisfy these conditions.

of the second order terms in the Edgeworth 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, then the wild bootstrap will not provide the usual full refinement over a normal approximation.

2.4 Local Asymptotic Risk

Heuristically, the ability of the wild bootstrap to outperform the normal approximation hinges on the degree of misspecification of the regression model as measured by the magnitude of the parameters γ_0 and γ_1. However, various tests of model specification are available that reject with probability tending to one outside local neighborhoods of the true distribution. It is therefore natural to evaluate the performance of the wild bootstrap as an estimator of the finite sample distribution of T_n in neighborhoods local to proper model specification. Following Beran (1982) and Singh and Babu (1990) we assess such performance in terms of maximal local asymptotic risk.[3] Specifically, we compare limiting maximal local asymptotic risks for a variety of loss functions based on the Kolmogorov-Smirnov distance between the true distribution and its estimator.

Unfortunately, the local risk analysis requires us to significantly complicate notation, as it becomes imperative to explicitly study parameters as functions of the underlying distribution of the data. Towards this end, for P a measure on $\mathbf{R}^{1+d_x}$, let E_P denote expectation statements when $(Y, X) \sim P$. We will restrict our analysis to measures P such that $E_P[XX']$ is full rank, and hence:

$$\beta_0(P) \equiv (E_P[XX'])^{-1} E_P[XY] \ , \tag{9}$$

is well defined. Throughout what follows, let $\epsilon = Y - X'\beta_0(P)$. Note that we suppress the dependence of ϵ on P and, with some abuse of notation, we index the implied law of (X, ϵ) also by P. We further now need to make the dependence of previously defined moments on P explicit as well – e.g. $H_0(P) \equiv E_P[XX']$, $\Sigma_0(P) \equiv E_P[XX'\epsilon^2]$, and similarly for all expressions defined in (6).

In order to study local asymptotic risk, we must first define an appropriate local neighborhood. To this end, let $\mathbf{P}$ denote a set of probability distributions for (Y, X), which we assume is endowed with a metric $\|\cdot\|_{\mathbf{P}}$. For any $P_0 \in \mathbf{P}$ and any $0 < h \in \mathbf{R}$, we then define:

$$\mathbf{P}(P_0, h) \equiv \{P \in \mathbf{P} : \|P - P_0\|_{\mathbf{P}} \leq h\} \ . \tag{10}$$

[3] We thank an anonymous referee for suggesting we pursue this question.

Given the above definitions, we may now state our assumptions on $\mathbf{P}$ and associated metric $\|\cdot\|_{\mathbf{P}}$.

Assumption 2.3. *(i)* $\{Y_i, X_i\}_{i=1}^n$ *is i.i.d.; (ii) For some* $\tilde{\nu} \geq 18$, $\sup_{P\in\mathbf{P}} E_P[\|XX'\|_F^{\tilde{\nu}}] < \infty$ *and* $\sup_{P\in\mathbf{P}} E_P[\|XX'\epsilon^2\|_F^{\tilde{\nu}}] < \infty$*; (iii)* $\inf_{P\in\mathbf{P}} \lambda(H_0(P)) > 0$ *and* $\inf_{P\in\mathbf{P}} \lambda(\Sigma_0(P)) > 0$*, where* $\lambda(A)$ *is the smallest eigenvalue of* A*; (iv) For* $Z \equiv (\tilde{X}, X'\epsilon, vech(\tilde{X}\tilde{X}')', vech(XX'\epsilon^2))$, $\inf_{P\in\mathbf{P}} \lambda(E_P[ZZ']) > 0$*, and for* $\xi_{Z,P}$ *the characteristic function of* Z *under* P, $\sup_{P\in\mathbf{P}} |\xi_{Z,P}(t)| \leq F(t)$ *for some function* F *satisfying* $\sup_{\|t\|\geq\delta} F(t) < 1$ *for any* $\delta > 0$.

Assumption 2.4. *(i)* $\sigma(P)^2$, $\kappa(P)$, $\gamma_0(P)$ *and* $\gamma_1(P)$ *are continuous in* P *on* $\mathbf{P}$ *under* $\|\cdot\|_{\mathbf{P}}$.

Assumption 2.3(ii) strengthens the moment requirements relative to Assumption 2.1(ii) to ensure the wild bootstrap Edgeworth expansion holds except in a set whose probability vanishes sufficiently fast uniformly in $P \in \mathbf{P}$. In turn, Assumption 2.3(iii)-(iv) imposes that the conditions of Assumption 2.1(iii)-(iv) hold uniformly in $P \in \mathbf{P}$. In particular, we note that Assumption 2.1(iii)-(iv) and 2.3(iii)-(iv) are equivalent in the case that $\mathbf{P}$ is a singleton. The additional uniformity requirement of Assumptions 2.3(iii)-(iv) will enable us to show that the Edgeworth expansions for both the full sample and bootstrap statistics in fact hold uniformly in $P \in \mathbf{P}$.

In Assumption 2.4(i) we additionally demand that the metric $\|\cdot\|_{\mathbf{P}}$ be such that $\sigma(P)^2$, $\kappa(P)$, $\gamma_0(P)$ and $\gamma_1(P)$ are continuous in P under $\|\cdot\|_{\mathbf{P}}$ on $\mathbf{P}$. This requirement implies the Edgeworth expansions for T_n are continuous in the underlying distribution P with respect to the metric $\|\cdot\|_{\mathbf{P}}$ as well. Assumption 2.4(i) is satisfied, for example, if (Y, X) are bounded and $\|\cdot\|_{\mathbf{P}}$ metrizes the weak topology. Alternatively, in studying maximal local asymptotic risk for estimating the distribution of a univariate mean, Singh and Babu (1990) let $\|\cdot\|_{\mathbf{P}}$ denote the Kolmogorov-Smirnov metric and obtain continuity by uniformly bounding higher order moments of P in $\mathbf{P}$.

Under the additional Assumptions 2.3 and 2.4 we derive our asymptotic risk comparison.

Theorem 2.4. *Let Assumptions 2.2, 2.3 and 2.4 hold and* $L : [0,\infty) \to [0,\infty)$ *be a continuous increasing function satisfying* $\limsup_{a\to\infty} L(a)a^{-\vartheta} < \infty$ *for some* $0 < 9\vartheta < \tilde{\nu}$. *If* $E[W^3] = 1$, $P_0 \in \mathbf{P}$ *and in addition* $E_{P_0}[Y|X] = X'\beta_0(P_0)$*, then for any sequence* $h_n \downarrow 0$*:*

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} E_P[L(\sup_{z\in\mathbf{R}} \sqrt{n}|P(T_n \leq z) - P^*(T_n^* \leq z)|)] \\ \leq \liminf_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} L(\sup_{z\in\mathbf{R}} \sqrt{n}|P(T_n \leq z) - \Phi(z)|) \ . \quad (11)$$

Moreover, if in addition $L : [0,+\infty) \to [0,+\infty)$ is strictly increasing, and $P_0 \in \mathbf{P}$ is such that $\kappa(P_0) \neq 0$, then the inequality in (11) *is strict.*

Theorem 2.4 establishes that the asymptotic risk of the wild bootstrap is never larger than that of a normal approximation in shrinking neighborhoods of properly specified models. Moreover, the asymptotic risk is in fact strictly lower when the analysis is local to a distribution for which $\kappa(P_0) \neq 0$. Heuristically, this results from the effects of skewness ($\kappa(P)$) being more important than those of misspecification ($\gamma_0(P)$ and $\gamma_1(P)$) when P is local to a properly specified model. Thus, since the wild bootstrap distribution mimics the effects of skewness, while the normal approximation does not, it is able to deliver a lower asymptotic risk when $\kappa(P_0) \neq 0$.

The results of Theorem 2.4 also apply to any loss function satisfying a polynomial growth bound in the tails, provided sufficient moments exist. The more demanding moment requirements are employed in showing that the probability of an Edgeworth expansion not holding for the wild bootstrap distribution vanishes sufficiently fast uniformly in $P \in \mathbf{P}$.[4] Such a result, in turn enables us to show the wild bootstrap asymptotic risk remains finite despite the loss function diverging faster to infinity for higher degree polynomials. In the special case of quadratic loss, Theorem 2.4 requires to set $\tilde{\nu} = 18 + \delta$ for some $\delta > 0$ (compared to $\tilde{\nu} \geq 18$ in Assumption 2.3(ii)).

3 Monte Carlo

We turn now to a series of sampling experiments designed to assess the finite sample performance of the wild bootstrap in environments where misspecification is of concern. We begin with a short simulation study where the degree of heteroscedasticity and misspecification are varied parametrically. Specifically, we generate the variable Y according to the relationship:

$$Y_i = X_i^{(1)} + X_i^{(2)} + X_i^{(3)} + \psi X_i^{(1)} X_i^{(2)} + (1 + \lambda X_i^{(1)})\eta \ , \tag{12}$$

where $(V_i, X_i^{(2)}, X_3^{(3)})$ are distributed as trivariate standard normals with equal correlation of 0.2, and $X_i^{(1)} = \frac{\exp(V_i) - E[\exp(V_i)]}{\mathrm{Var}(\exp(V_i))^{1/2}}$. The lognormal specification of $X_i^{(1)}$ is meant to generate observations with high leverage, which Chesher and Jewitt (1987) have found can present serious obstacles to

[4] See also Bhattacharya and Qumsiyeh (1989) for a similar tradeoff between existence of moments and asymptotic loss under polynomial loss functions.

heteroscedasticity robust inference. Additionally, the log-normal specification provides us with an asymmetric covariate, which is helpful in avoiding overly optimistic results (Chesher (1995)). The error η is generated independently of the regressors as the mixture of a $N(-\frac{1}{9}, 1)$ variable with probability 0.9 and a $N(1, 4)$ variable with probability 0.1. Though all of the moments of η exist, it exhibits substantial skewness, a feature which influences the higher order approximate cumulants derived in Theorem 2.2.

In (12), the parameter ψ captures the degree of interaction between the covariates $X_i^{(1)}$ and $X_i^{(2)}$. We will consider the case where the researcher mistakenly omits this interaction in the estimated specification, so that ψ parametrizes the severity of the resulting misspecification of the conditional mean. Concretely, we examine the ability of the wild bootstrap to control size when conducting inference on the coefficient β_1 in the following (potentially misspecified) linear regression model:

$$Y_i = \alpha + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + \beta_3 X_i^{(3)} + \epsilon \ . \tag{13}$$

When $\psi = 0$ in (12), the model is properly specified and the population regression coefficient β_1 equals one. Otherwise, the model is misspecified, and $\beta_1 = 1 + \psi b$ where $b \approx .41$ is the coefficient on $X_i^{(1)}$ from a projection of $X_i^{(1)} X_i^{(2)}$ onto the space spanned by $(1, X_i^{(1)}, X_i^{(2)}, X_i^{(3)})$.

In conducting our sampling experiments we examine tests regarding β_1 centered around the true population regression coefficient $1 + \psi b$. The following testing procedures are considered: (i) t-tests based upon the usual ("normal") heteroscedasticity robust variance estimates White (1980b), (ii) t-tests relying upon a Wild bootstrap critical value using three alternative weighting schemes ("Gamma", "Rademacher", and "Mammen"), and (iii) the nonparametric ("pairs") bootstrap.[5] As Hall and Horowitz (1996) have shown, the pairs bootstrap obtains a refinement without mean independence assumptions on the regression errors. Thus, theory suggests that the nonparametric bootstrap should exhibit an improvement over a standard normal approximation regardless of whether misspecification is present or not.

Tables 1 and 2 provide empirical rejection rates of one-sided and two-sided tests under different values of the parameters governing misspecification (ψ) and heteroscedasticity (λ). All rejection

[5] The "Gamma" specification draws $\{W_i\}_{i=1}^n$ from a recentered Gamma distribution with shape parameter 4 and scale 1/2 as suggested by Liu (1988). The "Rademacher" weights put probability $\frac{1}{2}$ on 1 and -1, while the skew correcting Mammen (1993) weights equal $\frac{1-\sqrt{5}}{2}$ with probability $\frac{\sqrt{5}+1}{2\sqrt{5}}$ and $\frac{\sqrt{5}+1}{2}$ with probability $1 - \frac{\sqrt{5}+1}{2\sqrt{5}}$. The nonparametric bootstrap computes the distribution of $\sqrt{n}c'(\hat{\beta} - \beta_0)/\hat{\sigma}$ under the empirical measure.

Table 1: Rejection rates for 0.05 nominal size - One sided tests

	Sample Size $n = 100$. Alternative Hypothesis $H_1 : \beta < 1 + \psi b$									
	Homoscedasticity $\lambda = 0$					Heteroscedasticity $\lambda = 1$				
ψ	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
-0.5	0.100	0.072	0.062	0.067	0.071	0.132	0.084	0.073	0.091	0.085
0.0	0.083	0.087	0.079	0.086	0.077	0.155	0.107	0.094	0.112	0.101
0.5	0.171	0.142	0.127	0.140	0.130	0.180	0.129	0.111	0.131	0.118
	Sample Size $n = 100$. Alternative Hypothesis $H_1 : \beta > 1 + \psi b$									
	Homoscedasticity $\lambda = 0$					Heteroscedasticity $\lambda = 1$				
ψ	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
-0.5	0.172	0.142	0.127	0.141	0.134	0.159	0.112	0.095	0.122	0.109
0.0	0.071	0.070	0.063	0.071	0.065	0.131	0.090	0.078	0.101	0.091
0.5	0.091	0.064	0.058	0.061	0.063	0.115	0.073	0.063	0.084	0.079
	Sample Size $n = 200$. Alternative Hypothesis $H_1 : \beta < 1 + \psi b$									
	Homoscedasticity $\lambda = 0$					Heteroscedasticity $\lambda = 1$				
ψ	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
-0.5	0.094	0.061	0.051	0.064	0.064	0.122	0.082	0.066	0.091	0.085
0.0	0.082	0.082	0.077	0.082	0.075	0.144	0.100	0.082	0.112	0.101
0.5	0.169	0.142	0.125	0.141	0.130	0.168	0.125	0.104	0.135	0.121
	Sample Size $n = 200$. Alternative Hypothesis $H_1 : \beta > 1 + \psi b$									
	Homoscedasticity $\lambda = 0$					Heteroscedasticity $\lambda = 1$				
ψ	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
-0.5	0.153	0.124	0.110	0.126	0.120	0.128	0.093	0.074	0.107	0.099
0.0	0.063	0.060	0.057	0.062	0.060	0.108	0.073	0.059	0.089	0.081
0.5	0.088	0.058	0.047	0.059	0.061	0.099	0.062	0.050	0.078	0.072

rates were computed using 200 bootstrap repetitions and 10,000 Monte Carlo replications.

The results in Table 1 indicate that all of the inference procedures provide comparable performance for properly specified homoscedastic models. The introduction of either heteroscedasticity ($\lambda = 1$) or misspecification ($\psi \neq 0$), however, drastically affects the accuracy of the conventional normal approximation, which all bootstrap procedures outperform in all specifications. As noted by White (1982), misspecification typically induces heteroscedasticity, so separating their effects is not straightforward without considering relatively contrived DGPs. We note that the bootstrap yields important improvements even at sample sizes of two hundred, for which the size distortion for a normal approximation ranges from 0.038 to 0.119. The relative performance of the wild and pairs bootstraps under misspecification is roughly comparable; the "Gamma" results, for example, are very similar to those of "pairs". The exact rankings, however, seem to depend on both the direction of misspecification and the presence of heteroscedasticity.

The ranking of the various techniques for two-sided tests in Table 2 is more clear cut, with the

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

	Sample Size $n = 100$. Alternative Hypothesis $H_1 : \beta \neq 1 + \psi b$									
	Homoscedasticity $\lambda = 0$					Heteroscedasticity $\lambda = 1$				
ψ	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
-0.5	0.200	0.155	0.139	0.150	0.112	0.218	0.135	0.125	0.136	0.081
0.0	0.097	0.105	0.098	0.102	0.078	0.215	0.137	0.124	0.138	0.082
0.5	0.194	0.152	0.136	0.144	0.111	0.220	0.140	0.130	0.141	0.085
	Sample Size $n = 200$. Alternative Hypothesis $H_1 : \beta \neq 1 + \psi b$									
	Homoscedasticity $\lambda = 0$					Heteroscedasticity $\lambda = 1$				
ψ	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
-0.5	0.172	0.124	0.109	0.122	0.092	0.179	0.113	0.094	0.111	0.069
0.0	0.087	0.095	0.087	0.092	0.075	0.180	0.114	0.098	0.112	0.071
0.5	0.184	0.136	0.120	0.131	0.100	0.191	0.122	0.103	0.122	0.076
	Sample Size $n = 400$. Alternative Hypothesis $H_1 : \beta \neq 1 + \psi b$									
	Homoscedasticity $\lambda = 0$					Heteroscedasticity $\lambda = 1$				
ψ	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
-0.5	0.149	0.107	0.087	0.105	0.073	0.138	0.089	0.072	0.085	0.052
0.0	0.073	0.078	0.071	0.077	0.062	0.134	0.088	0.072	0.084	0.054
0.5	0.153	0.112	0.094	0.111	0.081	0.148	0.095	0.075	0.095	0.058

nonparametric bootstrap performing best 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 findings regarding asymptotic risk. It is interesting to note that these conclusions seem to apply to the "Rademacher" and "Mammen" variants of the wild bootstrap, despite the fact that these weighting schemes fail to satisfy Cramer's condition (Assumption 2.2(ii)). Also notable is that the wild bootstraps are, in this design, bested by the conventional pairs bootstrap even under proper specification. By contrast, the monte carlo experiments of Horowitz (1997, 2001), found the wild bootstrap outperforming pairs under proper specification, though he considered designs with less severe leverage. Mammen (1993) also finds, in accordance with theory, the wild bootstrap outperforming pairs in simulations with a large number of regressors.

3.1 Mincer Regression

As a second exercise, we conduct a sampling experiment using actual earnings data from the Decennial Census. We work with an extract from the 2000 Integrated Public Use Microsample (Ruggles et al. (2010)) of 1.365 million native born black and white men ages 21-55 with unallocated wage

Table 3: Rejection rates for 0.05 nominal size

	Alternative Hypothesis $H_1 : \beta_1 < -.128$					Alternative Hypothesis $H_1 : \beta_1 \neq -.128$				
N	Normal	Gamma	Rademacher	Mammen	Pairs	Normal	Gamma	Rademacher	Mammen	Pairs
50[8]	0.195	0.112	0.135	0.159	0.110	0.136	0.112	0.097	0.112	0.087
100	0.110	0.075	0.067	0.092	0.048	0.092	0.075	0.063	0.078	0.058
200	0.075	0.065	0.054	0.068	0.048	0.071	0.065	0.058	0.066	0.059
400	0.061	0.060	0.054	0.058	0.050	0.061	0.060	0.057	0.059	0.056

data and at least six years of schooling who are not currently enrolled in school. Our exercise focuses on estimation of a conventional Mincer (1974) specification for average hourly wages (Y_i):

$$\log(Y_i) = \alpha + \beta_1 \mathrm{BLACK}_i + \beta_2 \mathrm{SCHOOL}_i + \beta_3 \mathrm{POTEXP}_i + \beta_4 \mathrm{POTEXP}_i^2 + \epsilon_i \ , \qquad (14)$$

where BLACK is an indicator that equals one if the respondent reports their race as black, SCHOOL is a categorical variable indicating the years of education obtained, and POTEXP is potential experience measured as the individual's age minus years of schooling minus six.[6] Such specifications have a long history in labor economics, and are heavily used today by researchers in a variety of fields. However, as demonstrated by Heckman, Lochner, and Todd (2006), the basic Mincer specification provides a rather crude approximation to the true conditional expectation of log wages, ignoring (among other factors) important nonseparabilities between race, schooling, and potential experience.

We examine the performance of the wild bootstrap in this environment by drawing random subsamples with replacement from the "population" of census microdata and conducting hypothesis tests regarding the coefficient β_1. Because blacks constitute only 7.8% of the wage observations in our census extract, the BLACK dummy exhibits skewness of the form likely to present an obstacle to conventional inference procedures. We note in passing that this problem is not purely hypothetical, such regressions are often used as evidence in employer lawsuits involving racial discrimination where sample sizes are typically small (Kaye and Freedman (2000)). In conducting our hypothesis tests we center our test statistic around the "population" value of β_1 which is approximately $-.128$.[7]

[6] Average hourly wages were constructed as total wage and salary income divided by usual hours worked per week times the number of weeks worked. We drop observations with average hourly wages less than four dollars an hour or greater than one hundred dollars an hour. Degrees were converted to years of schooling according to the following scheme: Associates Degrees (14 years of schooling), Bachelors (16 years of schooling), Masters (18 years of schooling), Professional degree (19 years of schooling), Doctorate (20 years of schooling).

[7] That is, when estimating (14) employing all 1.365 million obsevrations, we find $\beta_1 = -.128$.

[8] We ignore subsamples of the population containing no respondents reporting their race as black. When implementing the "pairs" bootstrap, we also ignore such bootstrap samples. These difficulties only arise when $N = 50$.

Table 3 reports actual size control for one sided and two sided tests, based on 10,000 Monte Carlo replications and 200 bootstrap repetitions. In accord with our results from Tables 1 and 2, we find that the conventional heteroscedasticity robust estimates perform poorly in small to moderate sample sizes. In both one-sided and two-sided tests, each of the wild bootstrap procedures yield improvements over the normal approximation, while the nonparametric pairs bootstrap yields the best performance overall. Notably, these simulations suggest that even in realistic moderately sized samples it can be important to go beyond the normal approximation when conducting inference.

4 Conclusion

We examined the higher order properties of the wild bootstrap under model misspecification, and found its Edgeworth expansion disagrees with that of the full sample statistic. However, while the wild bootstrap may not provide a traditional refinement, we additionally established that it has a lower maximal local asymptotic risk than a normal approximation in neighborhoods of properly specified models. Heuristically, these results suggest the wild bootstrap is robust in the sense that it will outperform a normal approximation provided misspecification is not "too severe." Our Monte Carlo studies confirm these results, showing the wild bootstrap provides better size control than a normal approximation in a variety of misspecified models.

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.
$\lvert\lambda\rvert$	For a vector λ of positive integers and $\lambda^{(i)}$ its i^{th} coordinate $\lvert\lambda\rvert = \sum_i \lambda^{(i)}$.
$D^\lambda f$	For $f : \mathbf{R}^d \to \mathbf{R}$ and $\lambda \in \mathbf{R}$, $D^\lambda f = \frac{\partial^{\lvert\lambda\rvert} f}{\partial^{\lambda^{(1)}} \ldots \partial^{\lambda^{(d)}}}$.
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 such that $E[\|Z\|_F^\delta] < \infty$ for some $\delta \geq 2$, and $\{c_n\}_{n=1}^\infty$ be a sequence of scalars such that $c_n^{-1} = o(n^\alpha)$ for some $\alpha \in [0, \frac{\delta-1}{2\delta})$. Then, it follows that:*

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

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

$$\begin{aligned} P(\|\frac{1}{n}\sum_{i=1}^n \{Z_i - E[Z_i]\}\|_F > c_n) &\leq P(\max_{1\leq l\leq k, 1\leq j\leq p} |\frac{kp}{n}\sum_{i=1}^n \{Z_i^{(l,j)} - E[Z_i^{(l,j)}]\}| > c_n) \\ &\leq \sum_{l=1}^k \sum_{j=1}^p P(|\frac{kp}{n}\sum_{i=1}^n \{Z_i^{(l,j)} - E[Z_i^{(l,j)}]\}| > c_n) \ . \end{aligned} \tag{15}$$

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

$$\begin{aligned} P(|\frac{kp}{n}\sum_{i=1}^n \{Z_i^{(l,j)} - E[Z_i^{(l,j)}]\}| > c_n) &\leq \frac{1}{c_n^\delta} E[|\frac{1}{n}\sum_{i=1}^n \{Z_i^{(l,j)} - E[Z_i^{(l,j)}]\}|^\delta] \\ &\leq \frac{C_1}{c_n^\delta} E[(\frac{1}{n^2}\sum_{i=1}^n (Z_i^{(l,j)} - E[Z_i^{(l,j)}])^2)^{\frac{\delta}{2}}] \leq \frac{C_2}{c_n^\delta n^{\frac{\delta}{2}}} (\mathrm{Var}(Z_i^{(l,j)}))^{\frac{\delta}{2}} \ , \end{aligned} \tag{16}$$

where we have used that $\delta \geq 2$. The result then follows from (15), (16), and $c_n^{-1} = o(n^\alpha)$ for $\alpha \in [0, \frac{\delta-1}{2\delta})$. ■

Lemma A.2. *If $\Delta_n \equiv H_0 - H_n$, $\hat{\sigma}_R^2 \equiv c' H_0^{-1}\Sigma_n(\beta_0)H_0^{-1}c + 2c'H_0^{-1}\Delta_n H_0^{-1}\Sigma_0 H_0^{-1}c$ and Assumptions 2.1 hold:*

(i) $P(\|\frac{1}{\sqrt{n}}\sum_i X_i\epsilon_i\| > M_n) = o(n^{-\frac{1}{2}})$ *for any* $M_n \uparrow \infty$ *with* $n^{\frac{1}{2}-\alpha} = o(M_n)$ *for some* $\alpha \in [0, \frac{\nu-1}{2\nu})$.

(ii) $P(\|H_n^{-1} - \sum_{j=0}^k (H_0^{-1}\Delta_n)^j H_0^{-1}\|_o > n^{-\alpha}) = o(n^{-\frac{1}{2}})$ *for any* $\alpha \in [0, \frac{(k+1)(\nu-1)}{2\nu})$.

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

(iv) $P(|\hat{\sigma}^2 - \hat{\sigma}_R^2 + \frac{2}{n}\sum_i \gamma_0' H_0^{-1}X_i\epsilon_i| > n^{-\alpha}) = o(n^{-\frac{1}{2}})$ *for any* $\alpha \in [0, \frac{\nu-1}{\nu})$.

Proof: Since $E[\|X\epsilon\|^\nu] < \infty$ 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 $n^{\frac{1}{2}-\alpha} = o(M_n)$ for some $\alpha \in [0, \frac{\nu-1}{2\nu})$ we must have:

$$P(\|H_0^{-1}\Delta_n\|_F \geq \frac{M_n}{\sqrt{n}}) = o(n^{-\frac{1}{2}}) \ . \tag{17}$$

Moreover, notice that if $\|H_0^{-1}\|_F\|\Delta_n\|_F < 1$, then $H_n^{-1} = \sum_{j=0}^{\infty}(H_0^{-1}\Delta_n)^j H_0^{-1}$. Hence, we obtain that:

$$\begin{aligned} P(\|H_n^{-1} - \sum_{j=0}^{k}(H_0^{-1}\Delta_n)^j H_0^{-1}\|_o > n^{-\alpha}) &\le P(\|\sum_{j\ge k+1}(H_0^{-1}\Delta_n)^j H_0^{-1}\|_o > n^{-\alpha} \text{ and } \|H_0^{-1}\Delta_n\|_F < 1) + o(n^{-\frac{1}{2}}) \\ &\le P(\frac{|\xi(H_0^{-1}\Delta_n)|^{k+1}}{1-|\xi(H_0^{-1}\Delta_n)|} > \frac{n^{-\alpha}}{\|H_0^{-1}\|_o}) + o(n^{-\frac{1}{2}}) \ , \end{aligned} \quad (18)$$

where $\xi((H_0^{-1}\Delta_n)^j)$ is the largest eigenvalue in absolute value of $(H_0^{-1}\Delta_n)^j$ and we have exploited $\|(H_0^{-1}\Delta_n)^j\|_o = |\xi((H_0^{-1}\Delta_n)^j)| = |\xi(H_0^{-1}\Delta_n)|^j$ for the second inequality. Moreover, since $|\xi(H_0^{-1}\Delta_n)| = \|H_0^{-1}\Delta_n\|_o \le \|H_0^{-1}\Delta_n\|_F$, result (17) implies that $P(|\xi(H_0^{-1}\Delta_n)| \ge 1/2) = o(n^{-\frac{1}{2}})$. Therefore, from (18) we are able to conclude that:

$$\begin{aligned} &P(\|H_n^{-1} - \sum_{j=0}^{k}(H_0^{-1}\Delta_n)^j H_0^{-1}\|_o > n^{-\alpha}) \\ &\le P(2|\xi(H_0^{-1}\Delta_n)|^{k+1} > \frac{n^{-\alpha}}{\|H_0^{-1}\|_o}) + o(n^{-\frac{1}{2}}) \le P(2\|H_0^{-1}\Delta_n\|_F^{k+1} > \frac{n^{-\alpha}}{\|H_0^{-1}\|_o}) + o(n^{-\frac{1}{2}}) \ . \end{aligned} \quad (19)$$

To conclude, exploit (19) and set $M_n = n^{\frac{1}{2}-\frac{\alpha}{k+1}}$ in (17) to obtain $P(2\|H_0^{-1}\Delta_n\|_F > n^{-\frac{\alpha}{k+1}}) = o(n^{-\frac{1}{2}})$.

Next, note that Corollary III.2.6 in Bhatia (1997) implies $|\xi(H_n^{-1}) - \xi(H_0^{-1})| \le \|H_n^{-1} - H_0^{-1}\|_F$. By part (ii) of the Lemma, it follows that $P(\|H_n^{-1}\|_o > 2\|H_0^{-1}\|_o) = o(n^{-\frac{1}{2}})$. Therefore, we obtain that:

$$P(\|\hat{\beta} - \beta_0\| > n^{-\alpha}) \le P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\epsilon_i\| > \frac{n^{\frac{1}{2}-\alpha}}{2\|H_0^{-1}\|_o}) + o(n^{-\frac{1}{2}}) \ . \quad (20)$$

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

In order to establish the final claim of the Lemma, note that Assumption 2.1(ii) and Lemma A.1 imply that for any $M_n \uparrow \infty$ such that $n^{\frac{1}{2}-\alpha} = o(M_n)$ for some $\alpha \in [0, \frac{\nu-1}{2\nu})$, we have:

$$P(\|\Sigma_n(\beta_0) - \Sigma_0\|_F > \frac{M_n}{\sqrt{n}}) = o(n^{-\frac{1}{2}}) \ . \quad (21)$$

Let $X_i^{(j)}$ denote the j^{th} coordinate of X_i, and note $\nu \ge 4$ and Assumption 2.1(ii) guarantee $E[(X_i^{(j)}X_i^{(k)}X_i^{(l)}X_i^{(m)})^2] < \infty$ and $E[(X_i^{(j)}X_i^{(k)}X_i^{(l)}\epsilon_i)^2] < \infty$. Hence, by Lemma A.1, there exists an $M > 0$ such that:

$$P(|\frac{1}{n}\sum_{i=1}^{n} X_i^{(j)}X_i^{(k)}X_i^{(l)}X_i^{(m)}| > M) = o(n^{-\frac{1}{2}}) \qquad P(|\frac{1}{n}\sum_{i=1}^{n} X_i^{(j)}X_i^{(k)}X_i^{(l)}\epsilon_i| > M) = o(n^{-\frac{1}{2}}) \ . \quad (22)$$

By direct calculation and part (iii) of the present Lemma, we then obtain that if $\alpha \in [0, \frac{\nu-1}{\nu})$, then we can conclude:

$$P(\|\Sigma_n(\hat{\beta}) - \Sigma_n(\beta_0)\|_F > n^{-\frac{\alpha}{2}}) = P(\|\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\{(X_i'(\hat{\beta}-\beta_0))^2 - 2\epsilon_iX_i'(\hat{\beta}-\beta_0)\}\|_F > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \quad (23)$$

Let $K > 0$ be such that $\|\Sigma_0\|_o < K$ and note (21) and (23) imply $P(\|\Sigma_n(\hat{\beta})\|_o > K) = o(n^{-\frac{1}{2}})$. Hence, we conclude from part (ii) of the present Lemma that for any $\alpha \in [0, \frac{\nu-1}{\nu})$ we have:

$$\begin{aligned} &P(|c'(H_n^{-1} - H_0^{-1})\Sigma_n(\hat{\beta})(H_n^{-1} - H_0^{-1})c| > n^{-\alpha}) \\ &\le P(K\|c\|^2\|H_n^{-1} - H_0^{-1}\|_o^2 > n^{-\alpha}) + P(\|\Sigma_n(\hat{\beta})\|_o > K) = o(n^{-\frac{1}{2}}) \ . \end{aligned} \quad (24)$$

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 that:

$$P(|c'(H_n^{-1} - H_0^{-1} - H_0^{-1}\Delta_nH_0^{-1})\Sigma_n(\hat{\beta})H_0^{-1}c| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ , \quad (25)$$

for any $\alpha \in [0, \frac{\nu-1}{\nu})$. Moreover, for any $\alpha \in [0, \frac{\nu-1}{\nu})$, exploiting (17), (21) and (23) we also conclude:

$$P(|c'H_0^{-1}\Delta_n H_0^{-1}(\Sigma_n(\hat{\beta}) - \Sigma_0)H_0^{-1}c| > n^{-\alpha}) \leq P(\|c'H_0^{-1}\|^2\|H_0^{-1}\Delta_n\|_o\|\Sigma_n(\hat{\beta}) - \Sigma_0\|_o > n^{-\alpha})$$
$$\leq P(\|c'H_0^{-1}\|\|H_0^{-1}\Delta_n\|_F > n^{-\frac{\alpha}{2}}) + P(\|c'H_0^{-1}\|\|\Sigma_n(\hat{\beta}) - \Sigma_0\|_F > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \ . \quad (26)$$

Since $\nu \geq 4$, Assumption 2.1(ii) implies $E[\|(c'H_0^{-1}X_i)^2\epsilon_i X_i\|^{\frac{\nu}{2}}] < \infty$, and hence Lemma A.1 implies that for any $\alpha \in [0, \frac{\nu-1}{\nu})$ we have $P(\|\frac{1}{n}\sum_i (c'H_0^{-1}X_i)^2\epsilon_i X_i - \gamma_0\| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}})$. Therefore, arguing as in (26),

$$P(\|\frac{1}{n}\sum_{i=1}^n \{\epsilon_i(c'H_0^{-1}X_i)^2X_i' - \gamma_0'\}(\hat{\beta} - \beta_0)\| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ , \quad (27)$$

for any $\alpha \in [0, \frac{\nu-1}{\nu})$. Next, exploit parts (i) and (ii) of the present Lemma and argue as in (26) to obtain:

$$P(|\gamma_0'(\hat{\beta} - \beta_0) - \frac{1}{n}\sum_{i=1}^n \gamma_0' H_0^{-1}X_i\epsilon_i| > n^{-\alpha}) \leq P(\|\gamma_0\|\|H_n^{-1} - H_0^{-1}\|_o\|\frac{1}{n}\sum_{i=1}^n X_i\epsilon_i\| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ . \quad (28)$$

Hence, by results (27), (28), and combining result (22) and part (iii) of the present Lemma we establish that:

$$P(|c'H_0^{-1}\Sigma_n(\hat{\beta})H_0^{-1}c - c'H_0^{-1}\Sigma_n(\beta_0)H_0^{-1}c + \frac{2}{n}\sum_{i=1}^n \gamma_0'H_0^{-1}X_i\epsilon_i| > n^{-\alpha})$$
$$= P(|\frac{2}{n}\sum_{i=1}^n \gamma_0'H_0^{-1}X_i\epsilon_i - \frac{2}{n}\sum_{i=1}^n (c'H_0^{-1}X_i)^2\epsilon_i X_i'(\hat{\beta} - \beta_0) + \frac{1}{n}\sum_{i=1}^n (c'H_0^{-1}X_i)^2(X_i'(\hat{\beta} - \beta_0))^2| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) \ . \quad (29)$$

for any $\alpha \in [0, \frac{\nu-1}{\nu})$. To conclude, note that by direct manipulations we obtain that:

$$\hat{\sigma}^2 = c'(H_n^{-1} - H_0^{-1})\Sigma_n(\hat{\beta})(H_n^{-1} - H_0^{-1})c + c'H_0^{-1}\Sigma_n(\hat{\beta})H_0^{-1}c + 2c'(H_n^{-1} - H_0^{-1})\Sigma_n(\hat{\beta})H_0^{-1}c \ , \quad (30)$$

and hence the final claim of the Lemma follows from (24), (25), (26) and (29). ■

Lemma A.3. *Let Assumptions 2.1(i)-(iii) hold and L_n be as in Theorem (2.1). Then for any $\alpha \in [0, \frac{2\nu-3}{2\nu})$:*

$$\limsup_{n\to\infty} \sqrt{n}P(|T_n - L_n| > n^{-\alpha}) = 0 \ .$$

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

$$T_n - L_n = c'\{H_n^{-1} - H_0^{-1} - H_0^{-1}\Delta_n H_0^{-1}\}\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i + \frac{(\sigma - \hat{\sigma})}{\hat{\sigma}\sigma}c'\{H_n^{-1} - H_0^{-1}\}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i$$
$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i\{-\frac{1}{2\sigma^3}(\hat{\sigma}^2 - \hat{\sigma}_R^2 + \frac{2}{n}\sum_{i=1}^n \gamma_0'H_0^{-1}X_i\epsilon_i) + \frac{3}{4\bar{\sigma}^5}(\hat{\sigma}^2 - \sigma^2)^2\} \ . \quad (31)$$

Fix $\alpha \in [0, \frac{2\nu-3}{2\nu})$. To study the right hand side of (31), first observe that Lemma A.2(i) and A.2(ii) imply that:

$$P(|c'\{H_n^{-1} - H_0^{-1} - H_0^{-1}\Delta_n H_0^{-1}\}\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i| > n^{-\alpha})$$
$$\leq P(\|c\|\|H_n^{-1} - H_0^{-1} - H_0^{-1}\Delta_n H_0^{-1}\|_o > \frac{1}{n^{\alpha+\delta}}) + P(\|\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i\| > n^{\delta}) = o(n^{-\frac{1}{2}}) \ , \quad (32)$$

for any δ such that $\alpha+\delta < \frac{\nu-1}{\nu}$ and $\delta > \frac{1}{2}-\frac{\nu-1}{2\nu}$ (which exists by $\alpha \in [0, \frac{2\nu-3}{2\nu})$). Moreover, by identical manipulations but exploiting Lemma A.2(i) and A.2(iv) instead, we can similarly conclude:

$$P(|\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^{n} c'H_0^{-1}X_i\epsilon_i\{\hat{\sigma}^2 - \hat{\sigma}_R^2 + \frac{2}{n}\sum_{i=1}^{n}\gamma_0' H_0^{-1}X_i\epsilon_i\}| > n^{-\alpha}) = o(n^{-\frac{1}{2}}) , \tag{33}$$

for any $\alpha \in [0, \frac{2\nu-3}{2\nu})$. Next, notice that (21), Lemma A.2(i) and the Cauchy-Schwarz inequality imply that:

$$P(|c'H_0^{-1}(\Sigma_n(\beta_0) - \Sigma_0)H_0^{-1}c| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) \qquad P(|\frac{1}{n}\sum_{i=1}^{n}\gamma_0' H_0^{-1}X_i\epsilon_i| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) , \tag{34}$$

for any $\alpha \in [0, \frac{\nu-1}{\nu})$. Therefore, we obtain from (30) together with (24) and (29) that for $\alpha \in [0, \frac{\nu-1}{\nu})$ we have:

$$P(|\hat{\sigma}^2 - \sigma^2| > n^{-\frac{\alpha}{2}}) = o(n^{-\frac{1}{2}}) . \tag{35}$$

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 (35) and manipulations as in (32) we can conclude for $\alpha \in [0, \frac{2\nu-3}{2\nu})$:

$$P(|\frac{(\hat{\sigma}^2-\sigma^2)^2}{\bar{\sigma}^5\sqrt{n}}\sum_{i=1}^{n} c'H_0^{-1}X_i\epsilon_i| > n^{-\alpha}) \leq P((\hat{\sigma}^2-\sigma^2)^2 > \frac{\epsilon^5}{n^{\alpha+\delta}}) + o(n^{-\frac{1}{2}}) = o(n^{-\frac{1}{2}}) , \tag{36}$$

by setting δ such that $\alpha+\delta < \frac{\nu-1}{\nu}$ and $\delta > \frac{1}{2} - \frac{\nu-1}{2\nu}$. Similarly, by $P(\hat{\sigma} > \epsilon) = o(n^{-\frac{1}{2}})$ and Lemma A.2(i) we obtain:

$$\begin{aligned} P(|\frac{(\sigma-\hat{\sigma})}{\hat{\sigma}\sigma}c'\{H_n^{-1} - H_0^{-1}\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\epsilon_i| > n^{-\alpha}) &\leq P(\frac{|\sigma-\hat{\sigma}|\|c\|}{\epsilon^2}\|H_n^{-1} - H_0^{-1}\|_o > \frac{1}{n^{\alpha+\delta}}) + o(n^{-\frac{1}{2}}) \\ &\leq P(|\sigma - \hat{\sigma}| > \frac{\epsilon^2}{\|c\|n^{\frac{\alpha+\delta}{2}}}) + P(\|H_n^{-1} - H_0^{-1}\|_o > \frac{1}{n^{\frac{\alpha+\delta}{2}}}) + o(n^{-\frac{1}{2}}) = o(n^{-\frac{1}{2}}) . \end{aligned} \tag{37}$$

where the final result follows from Lemma A.2(ii), equation (35) and $\alpha + \delta < \frac{\nu-1}{\nu}$. The Lemma is then established due to the decomposition in (31) and results (32), (33), (36) and (37). ■

Lemma A.4. *Let $\{A_{in}\}_{i=1}^n$ be a triangular array of $k \times p$ random matrices, and $\{c_n\}$ be a sequence of scalar valued random variables. Further assume that $\{A_{in}\}_{i=1}^n$ and c_n are both measurable functions of the data $\{Y_i, X_i\}_{i=1}^n$, that Assumptions 2.1(i) and 2.2(i) hold, and in addition:*

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\|A_{in}\|_F^{\delta} < \infty \qquad c_n^{-1} = o(n^{\alpha}) \qquad a.s. \tag{38}$$

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

$$P^*(\|\frac{1}{n}\sum_{i=1}^{n}A_{in}\{g(W_i) - E[g(W_i)]\}\|_F > c_n) \leq \sum_{l=1}^{k}\sum_{j=1}^{p}P^*(|\frac{kp}{n}\sum_{i=1}^{n}A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\}| > c_n) . \tag{39}$$

Next, apply Markov's, Marcinkiewicz and Rosenthal inequalities as in (16) to obtain for some C_1, C_2:

$$\begin{aligned} \sqrt{n}P^*(|\frac{1}{n}\sum_{i=1}^{n}A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\}| > c_n) &\leq \frac{\sqrt{n}}{c_n^{\delta}}E^*[|\frac{1}{n}\sum_{i=1}^{n}A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\}|^{\delta}] \\ &\leq \frac{\sqrt{n}C_1}{c_n^{\delta}}E^*[(\frac{1}{n^2}\sum_{i=1}^{n}(A_{in}^{(l,j)}\{g(W_i) - E[g(W_i)]\})^2)^{\frac{\delta}{2}}] \leq \frac{\sqrt{n}C_2}{c_n^{\delta}n^{\frac{\delta}{2}}}(\frac{1}{n}\sum_{i=1}^{n}(A_{in}^{(l,j)})^2\mathrm{Var}(g(W_i)))^{\frac{\delta}{2}} . \end{aligned} \tag{40}$$

The claim of the Lemma then follows by (38), (39), (40) and $\alpha \in [0, \frac{\delta-1}{2\delta})$ by hypothesis. ■

Lemma A.5. *Let $(\hat{\sigma}_s^*)^2 \equiv c' H_n^{-1} \Sigma_n^*(\hat{\beta}) H_n^{-1} c$ and $\{c_n\}$ be measurable scalar-valued functions of $\{Y_i, X_i\}_{i=1}^n$, and further suppose Assumptions 2.1(i)-(iii) and 2.2(i) hold. Then it follows that:*

(i) If $c_n^{-1} = o(n^\alpha)$ a.s. for some $\alpha \in [0, \frac{\omega\wedge\nu-1}{2(\omega\wedge\nu)})$, then $P^*(\|\hat{\beta}^* - \hat{\beta}\| > c_n) = o(n^{-\frac{1}{2}})$ a.s.

(ii) If $c_n^{-1} = o(n^\alpha)$ a.s. for some $\alpha \in [0, \frac{\omega\wedge(\nu/2)-1}{2(\omega\wedge(\nu/2))})$, then $P^*(|(\hat{\sigma}^*)^2 - (\hat{\sigma}_s^*)^2| > c_n^2) = o(n^{-\frac{1}{2}})$ a.s.

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

Proof: Since $\hat{\beta} \xrightarrow{a.s.} \beta$, and $E[\|XX'\|_F^\nu] < \infty$, $E[\|X\epsilon\|^\nu] < \infty$ by Assumption 2.1(ii), we obtain $\limsup \frac{1}{n}\sum_i \|X_i(Y_i - X_i'\hat{\beta}\|^\nu < \infty$ *a.s.*. Therefore, $\|H_n^{-1}\|_o \xrightarrow{a.s.} \|H_0^{-1}\|_o < \infty$, $E[|W|^\omega] < \infty$, $\alpha \in [0, \frac{\omega\wedge\nu-1}{2(\omega\wedge\nu)})$ and Lemma A.4 imply that:

$$P^*(\|\hat{\beta}^* - \hat{\beta}\| > c_n) \leq P^*(\|H_n^{-1}\|_o \|\frac{1}{n}\sum_{i=1}^n X_i(Y_i - X_i\hat{\beta})W_i\| > c_n) = o(n^{-\frac{1}{2}}) \qquad a.s. \tag{41}$$

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

$$\begin{aligned} P^*(|(\hat{\sigma}^*)^2 - (\hat{\sigma}_s^*)^2| > c_n^2) &= P^*(|c' H_n^{-1}\{\frac{1}{n}\sum_{i=1}^n X_iX_i'(X_i'(\hat{\beta}^* - \hat{\beta}))^2 - \frac{2}{n}\sum_{i=1}^n X_iX_i'\epsilon_i^* X_i'(\hat{\beta}^* - \hat{\beta})\} H_n^{-1} c| > c_n^2) \\ &\leq P^*(\|c\|^2 \|H_n^{-1}\|_o^2 \{\|\frac{1}{n}\sum_{i=1}^n X_iX_i'(X_i'(\hat{\beta}^* - \hat{\beta}))^2\|_o + \|\frac{2}{n}\sum_{i=1}^n X_iX_i'\epsilon_i^* X_i'(\hat{\beta}^* - \hat{\beta})\|_o\} > c_n^2) \ . \end{aligned} \tag{42}$$

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

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

Note Assumption 2.1(ii) implies that for any (j,k), $E[\|X^{(j)}X^{(k)}XX'\|^{\frac{\nu}{2}}] < \infty$ and $E[\|X^{(j)}X^{(k)}X\epsilon\|^{\frac{\nu}{2}}] < \infty$. Hence, $\hat{\beta} \xrightarrow{a.s.} \beta$ yields $\limsup \frac{1}{n}\sum_i \|X_i^{(j)}X_i^{(k)}X_i(Y_i - X_i'\hat{\beta})\|^{\frac{\nu}{2}} < \infty$ *a.s.*. Lemma A.4 and part (i) of this Lemma, then imlpy:

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

almost surely. Moreover, since $E[|X^{(j)}X^{(k)}X^{(l)}X^{(m)}|] < \infty$ by Assumption 2.1(ii), part (i) of the Lemma yields:

$$P^*(\|c\|^2 \|H_n^{-1}\|_o^2 \|\frac{1}{n}\sum_{i=1}^n X_iX_i'(X_i'(\hat{\beta}^* - \hat{\beta}))^2\|_o > c_n^2) = o(n^{-\frac{1}{2}}) \ , \tag{45}$$

almost surely. The second claim of the Lemma then follows from (42)-(45).

Next, note that $\|H_n^{-1}\|_o \xrightarrow{a.s.} \|H_0^{-1}\|_o < \infty$ by Assumption 2.1(iii), and $\hat{\sigma}^2 \xrightarrow{a.s.} \sigma$ together with Lemma A.4 imply:

$$P^*(|(\hat{\sigma}_s^*)^2 - \sigma^2| > \epsilon) \leq P^*(|(\hat{\sigma}_s^*)^2 - \hat{\sigma}^2| > \epsilon - |\hat{\sigma}^2 - \sigma^2|)$$

$$\leq P^*(\|\frac{1}{n}\sum_{i=1}^n X_i X_i'(Y_i - X_i\hat{\beta})^2(W_i^2 - 1)\|_F > \frac{\epsilon - |\hat{\sigma}^2 - \sigma^2|}{\|c\|^2\|H_n^{-1}\|_o^2}) = o(n^{-\frac{1}{2}}) \qquad a.s. \;, \quad (46)$$

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

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

$$T_{s,n}^* \equiv \frac{\sqrt{n}c'}{\hat{\sigma}_s^*}(\hat{\beta}^* - \hat{\beta}) \qquad\qquad (\hat{\sigma}_s^*)^2 \equiv c'H_n^{-1}\Sigma_n^*(\hat{\beta})H_n^{-1}c \;. \quad (47)$$

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

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 $\gamma \in [0, \frac{\omega\wedge\nu - 1}{2(\omega\wedge\nu)})$, part (i) of Lemma A.5 then establishes that:

$$P^*(|T_n^* - T_{s,n}^*| > n^{-\alpha}) \leq P^*(\frac{\sqrt{n}|\hat{\sigma}^* - \hat{\sigma}_s^*|}{\hat{\sigma}^*\hat{\sigma}_s^*} \times \|c\|\|\hat{\beta}^* - \hat{\beta}\| > n^{-\alpha})$$

$$\leq P^*(\sqrt{n}|\hat{\sigma}^* - \hat{\sigma}_s^*| > \frac{\epsilon}{n^{\alpha-\gamma}}) + P^*(\|\hat{\beta}^* - \hat{\beta}\| > \frac{1}{n^{\gamma}\|c\|}) + P^*(\hat{\sigma}^*\hat{\sigma}_s^* < \epsilon)$$

$$= P^*(\sqrt{n}|\hat{\sigma}^* - \hat{\sigma}_s^*| > \frac{\epsilon}{n^{\alpha-\gamma}}) + o(n^{-\frac{1}{2}}) \qquad a.s. \;. \quad (48)$$

Since for any $\alpha \in [0, \frac{(2\omega)\wedge\nu - 2}{(2\omega)\wedge\nu} - \frac{1}{2(\omega\wedge\nu)})$ we may pick $\gamma \in [0, \frac{\omega\wedge\nu - 1}{2(\omega\wedge\nu)})$ so that $\alpha - \gamma + \frac{1}{2} \in [0, \frac{\omega\wedge(\nu/2) - 1}{\omega\wedge(\nu/2)})$, the claim of the Lemma then follows from result (48) 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:*

$$E[L_n] = -\frac{\kappa}{2\sigma^3\sqrt{n}} - \frac{\gamma_1}{\sigma\sqrt{n}} + \frac{2c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0}{\sigma^3\sqrt{n}} \qquad\qquad E^*[L_n^*] = -\frac{E[W^3]\hat{\kappa}}{2\hat{\sigma}^3\sqrt{n}} \;.$$

Proof: We first derive an expression for $E[L_n]$. Note that $E[XX'] = H_0$ and $E[X\epsilon] = 0$ imply:

$$E[c'H_0^{-1}\Delta_n H_0^{-1}\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i]$$

$$= c'E[\frac{1}{n}\sum_{i=1}^n H_0^{-1}(I - X_iX_i')H_0^{-1}\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i] = -\frac{1}{\sigma\sqrt{n}}E[(c'H_0^{-1}X)X'H_0^{-1}X\epsilon] \quad (49)$$

due to the i.i.d. assumption. Similarly, exploiting $E[(c'H_0^{-1}X)\epsilon] = H_0^{-1}E[\Delta_n]H_0^{-1} = 0$ yields:

$$E[\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n (c'H_0^{-1}X_i)\epsilon_i(\hat{\sigma}_R^2 - \sigma^2)]$$

$$= E[\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n (c'H_0^{-1}X_i)\epsilon_i\{c'H_0^{-1}(\Sigma_n(\beta_0) - \Sigma_0)H_0^{-1}c + 2c'H_0^{-1}\Delta_n H_0^{-1}\Sigma_0 H_0^{-1}c\}]$$

$$= \frac{1}{2\sigma^3\sqrt{n}}\{E[(c'H_0^{-1}X)^3\epsilon^3] - 2E[\epsilon(c'H_0^{-1}X)^2X'H_0^{-1}]\Sigma_0 H_0^{-1}c\} \;. \quad (50)$$

The expression for $E[L_n]$ can then be obtained from (49), (50) and by analogous arguments concluding:

$$E[\frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n (c'H_0^{-1}X_i)\epsilon_i \times \frac{2}{n}\sum_{i=1}^n \gamma_0'H_0^{-1}X_i\epsilon_i] = \frac{c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0}{\sigma^3\sqrt{n}} \;. \quad (51)$$

In order to compute $E^*[L_n^*]$, observe that $W \perp (Y, X)$ and $E[W^2] = 1$ by Assumption 2.2(i) imply that:

$$E^*[L_n^*] = -\frac{1}{2\hat{\sigma}^3} E^*[\frac{c' H_n^{-1}}{\sqrt{n}} \sum_{i=1}^n X_i \epsilon_i^* \frac{1}{n} \sum_{i=1}^n c' H_n^{-1} X_i X_i' H_n^{-1} c e_i^2 (W_i^2 - 1)] = -\frac{E[W^3]\hat{\kappa}}{2\hat{\sigma}^3 \sqrt{n}} , \tag{52}$$

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 $E[XX'] = H_0$, $E[X\epsilon] = 0$ and direct calculations yield:

$$\begin{aligned} & E[(c' H_0^{-1} \Delta_n H_0^{-1} \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n X_i \epsilon_i)^2] \\ & \quad = E[(\frac{c'}{n} \sum_{i=1}^n H_0^{-1}(H_0 - X_i X_i') H_0^{-1} \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n X_i \epsilon_i)^2] = \frac{1}{\sigma^2 n^2} E[(c' H_0^{-1}(H_0 - X_i X_i') H_0^{-1} (\sum_{k=1}^n X_k \epsilon_k))^2] \\ & \qquad + \frac{(n-1)}{\sigma^2 n^2} E[\{c' H_0^{-1}(H_0 - X_i X_i) H_0^{-1} \sum_{k=1}^n X_k \epsilon_k\}\{c' H_0^{-1}(H_0 - X_j X_j') H_0^{-1} \sum_{k=1}^n X_k \epsilon_k\}] = O(n^{-1}) . \end{aligned} \tag{53}$$

Similarly, exploiting the i.i.d. assumption together with $E[X\epsilon] = 0$ and $E[H_0 - XX'] = 0$ we obtain:

$$\begin{aligned} & E[(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n c' H_0^{-1} X_i \epsilon_i)(c' H_0^{-1} \Delta_n H_0^{-1} \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n X_i \epsilon_i)] \\ & \qquad = \frac{1}{n^2 \sigma^2} E[(\sum_{i=1}^n c' H_0^{-1} X_i \epsilon_i)(c' \sum_{i=1}^n H_0^{-1}(H_0 - X_i X_i') H_0^{-1})(\sum_{i=1}^n X_i \epsilon_i)] \\ & \qquad = \frac{1}{n\sigma^2} E[(c' H_0^{-1} X \epsilon)(c' H_0^{-1} X \epsilon - c' H_0^{-1} X X' H_0^{-1} X \epsilon)] = O(n^{-1}) . \end{aligned} \tag{54}$$

Exploiting identical arguments to (53) on the squares of L_n and the Cauchy-Schwarz inequality and arguments identical to those in (54) to address cross terms, it is then straightforward to establish that:

$$E[L_n^2] = E[(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n c' H_0^{-1} X_i \epsilon_i)^2] + O(n^{-1}) = \frac{c' E[H_0^{-1} X X' \epsilon^2 H_0^{-1}] c}{\sigma^2} + O(n^{-1}) = 1 + O(n^{-1}) . \tag{55}$$

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:

$$\frac{1}{\hat{\sigma}^4 n^2} E^*[(\sum_{i=1}^n a_{in} \epsilon_i^*)^2 (\sum_{i=1}^n a_{in}^2 \{(\epsilon_i^*)^2 - e_i^2\})] = \frac{1}{\hat{\sigma}^4 n^2} \sum_{i=1}^n a_{in}^4 e_i^4 (E[W^4] - 1) = O_{a.s}(n^{-1}) . \tag{56}$$

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:

$$\begin{aligned} & \frac{1}{4\hat{\sigma}^6 n^3} E^*[(\sum_{i=1}^n a_{in} \epsilon_i^*)^2 (\sum_{i=1}^n a_{in}^2 \{(\epsilon_i^*)^2 - e_i^2\})^2] \\ & \quad = \frac{1}{4\hat{\sigma}^6 n^3} \{\sum_{i=1}^n E^*[a_{in}^2 (\epsilon_i^*)^2 (\sum_{k=1}^n a_{kn}^2 \{(\epsilon_k^*)^2 - e_k^2\})^2] + \sum_{i=1}^n \sum_{j \neq i} E^*[(a_{in} \epsilon_i^*)(a_{jn} \epsilon_j^*)(\sum_{k=1}^n a_{kn}^2 \{(\epsilon_k^*)^2 - e_k^2\})^2]\} \\ & \quad = \frac{1}{4\hat{\sigma}^6 n^3} \{\sum_{i=1}^n \sum_{k=1}^n a_{in}^2 a_{kn}^4 E^*[(\epsilon_i^*)^2 \{(\epsilon_k^*)^2 - e_k^2\}^2] + 2 \sum_{i=1}^n \sum_{j \neq i} a_{in}^3 e_i^3 a_{jn}^3 e_j^3 (E[W^3])^2\} . \end{aligned} \tag{57}$$

Therefore, expanding the square, noting that $\frac{1}{n}\sum_i a_{in}^2 e_i^2 = \hat{\sigma}^2$ and exploiting (56) and (57):

$$E^*[(L_n^*)^2] = \frac{1}{n\hat{\sigma}^2}E^*[(\sum_{i=1}^n a_{in}\epsilon_i^*)^2] + O_{a.s.}(n^{-1}) = 1 + O_{a.s.}(n^{-1}) \ , \tag{58}$$

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:*

$$E[L_n^3] = -\frac{7\kappa}{2\sigma^3\sqrt{n}} - \frac{3\gamma_1}{\sigma\sqrt{n}} + \frac{12c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0}{\sigma^3\sqrt{n}} + O(n^{-1}) \qquad E^*[(L_n^*)^3] = -\frac{7E[W^3]\hat{\kappa}}{2\hat{\sigma}^3\sqrt{n}} + O_{a.s.}(n^{-1}) \ . \tag{59}$$

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

$$\Gamma_n \equiv c'H_0^{-1}\Delta_n H_0^{-1}\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i\epsilon_i - \frac{1}{2\sigma^3\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i\{(\hat{\sigma}_R^2 - \sigma^2) - \frac{2}{n}\sum_{i=1}^n \gamma_0' H_0^{-1}X_i\epsilon_i\} \ . \tag{60}$$

Notice that $L_n = \frac{1}{\sigma\sqrt{n}}c'\sum_i H_0^{-1}X_i\epsilon_i + \Gamma_n$. Under Assumption 2.1(ii), it can be shown that $E[\Gamma_n^3] = O(n^{-\frac{3}{2}})$ and similarly that $E[(\frac{1}{\sqrt{n}}\sum_i c'H_0^{-1}X_i\epsilon_i)^3] = O(n^{-\frac{1}{2}})$. Therefore, by direct calculation and Holder's inequality:

$$\begin{aligned}
E[L_n^3] &= E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^3] + 3E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^2\Gamma_n] + 3E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)\Gamma_n^2] + E[\Gamma_n^3] \\
&= E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^3] + 3E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^2\Gamma_n] + O(n^{-1}) \ .
\end{aligned} \tag{61}$$

Hence, we can establish the first claim of the Lemma by analyzing the remaining terms in (61). Note that

$$E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^3] = \frac{1}{\sigma^3\sqrt{n}}E[(c'H_0^{-1}X)^3\epsilon^3] \ , \tag{62}$$

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

$$\begin{aligned}
&E[(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^2\frac{c'H_0^{-1}\Delta_n H_0^{-1}}{\sqrt{n}\sigma}\sum_{i=1}^n X_i\epsilon_i] \\
&= \frac{1}{\sigma^3 n^{\frac{5}{2}}}E[\{\sum_{i=1}^n (c'H_0^{-1}X_i)^2\epsilon_i^2 + \sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i\sum_{j\neq i}c'H_0^{-1}X_j\epsilon_j\}\sum_{k=1}^n c'H_0^{-1}(H_0 - X_kX_k')H_0^{-1}\sum_{l=1}^n X_l\epsilon_l] \\
&= -\frac{c'H_0^{-1}\Sigma_0 H_0^{-1}c}{\sigma^3\sqrt{n}}E[(c'H_0^{-1}X)(X'H_0^{-1}X)\epsilon] - \frac{2}{\sigma^3\sqrt{n}}E[(c'H_0^{-1}X)(\gamma_0'H_0^{-1}X)\epsilon^2] + O(n^{-\frac{3}{2}}) \ .
\end{aligned} \tag{63}$$

By analogous arguments we can compute the remaining terms in $E[(\frac{1}{\sigma\sqrt{n}}\sum_i c'H_0^{-1}X_i\epsilon_i)^2\Gamma_n]$ and obtain:

$$\frac{1}{2\sigma^5}E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^3 c'H_0^{-1}\{\Sigma_n(\beta_0) - \Sigma_0\}H_0^{-1}c] = \frac{3c'H_0^{-1}\Sigma_0 H_0^{-1}c}{2\sigma^5\sqrt{n}}E[(c'H_0^{-1}X)^3\epsilon^3] + O(n^{-\frac{3}{2}}) \tag{64}$$

$$\frac{1}{\sigma^5}E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^3\{c'H_0^{-1}\Delta_n H_0^{-1}\Sigma_0 H_0^{-1}c\}] = -\frac{3c'H_0^{-1}\Sigma_0 H_0^{-1}c}{\sigma^5\sqrt{n}}\gamma_0'H_0^{-1}\Sigma_0 H_0^{-1}c + O(n^{-\frac{3}{2}}) \tag{65}$$

$$\frac{1}{\sigma^5}E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n c'H_0^{-1}X_i\epsilon_i)^3\{\frac{1}{n}\sum_{i=1}^n \gamma_0'H_0^{-1}X_i\epsilon_i\}] = \frac{3c'H_0^{-1}\Sigma_0 H_0^{-1}c}{\sigma^5\sqrt{n}}c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0 + O(n^{-\frac{3}{2}}) \ . \tag{66}$$

The first claim of the Lemma then follows by combining the results from (61)-(66).

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

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

$$E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3(\frac{1}{2\hat\sigma^3}\{(\hat\sigma_s^*)^2-\hat\sigma^2\})^3] = O_{a.s.}(n^{-\frac{3}{2}})\ . \quad (68)$$

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:

$$E^*[(L_n^*)^3] = E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3\{\frac{1}{\hat\sigma^3} - \frac{3((\hat\sigma_s^*)^2-\hat\sigma^2)}{2\hat\sigma^5} + \frac{3((\hat\sigma_s^*)^2-\hat\sigma^2)^2}{4\hat\sigma^7} - \frac{((\hat\sigma_s^*)^2-\hat\sigma^2)^3}{8\hat\sigma^9}\}]$$

$$= \frac{E[W_i^3]}{\hat\sigma^3\sqrt{n}} \times \frac{1}{n}\sum_{i=1}^n a_{in}^3 e_i^3 - \frac{3}{2\hat\sigma^5}E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3\{(\hat\sigma_s^*)^2-\hat\sigma^2\}] + O_{a.s}(n^{-\frac{3}{2}})\ . \quad (69)$$

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

$$E^*[(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_{in}\epsilon_i^*)^3\{\frac{3}{2\hat\sigma^5 n}\sum_{i=1}^n a_{in}^2\{(\epsilon_i^*)^2 - e_i^2\}\}]$$

$$= \frac{3}{2\hat\sigma^5 n^{\frac{3}{2}}} \times \frac{1}{n}\sum_{i=1}^n a_{in}^5 E^*[(\epsilon_i^*)^3\{(\epsilon_i^*)^2-e_i^2\}] + \frac{9}{2\hat\sigma^5 n^{\frac{5}{2}}}E^*[\{\sum_{i=1}^n a_{in}(\epsilon_i^*)\sum_{j\neq i} a_{jn}^2(\epsilon_j^*)^2\}\sum_{k=1}^n a_{kn}^2\{(\epsilon_k^*)^2 - e_i^2\}]$$

$$= \frac{9}{2\hat\sigma^5\sqrt{n}} \times \frac{1}{n}\sum_{i=1}^n a_{in}^2 e_i^2 \times \frac{E[W^3]}{n}\sum_{i=1}^n a_{in}^3 e_i^3 + O_{a.s.}(n^{-\frac{3}{2}})\ . \quad (70)$$

The second claim of the Lemma is then established by (69) and (70). ■

Proof of Theorem 2.1: The first claim of the Theorem is an immediate consequence of Lemma A.3 and $\nu \geq 9$. For the second claim, note that in lieu of Lemma A.6 and $\omega \wedge \nu \geq 9$, it suffices to show that $T_{n,s}^* = L_n^* + o_{p^*}(n^{-\frac{1}{2}})$ *a.s.*. For notational simplicity, let $a_{in} \equiv c'H_n^{-1}X_i(Y_i - X_i'\hat\beta)$ and apply Markov's inequality to conclude that:

$$P^*(|(\hat\sigma_s^*)^2 - \hat\sigma^2| > \frac{C}{\sqrt{n}}) = P^*(|\frac{1}{n}\sum_{i=1}^n a_{in}^2(W_i^2-1)| > \frac{C}{\sqrt{n}})$$

$$\leq \frac{n}{C^2}E^*[(\frac{1}{n}\sum_{i=1}^n a_{in}^2(W_i^2-1))^2] = \frac{1}{C^2 n}\sum_{i=1}^n a_{in}^4 E[(W_i^2-1)^2]\ . \quad (71)$$

However, under our moment assumptions, $\frac{1}{n}\sum_i a_{in}^4 E[(W_i^2-1)^2] \stackrel{a.s.}{\to} E[(c'H_0^{-1}X)^4\epsilon_i^4]E[(W^2-1)^2] < \infty$, and therefore from (71) 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 - Proof of Theorems 2.3 and 2.4

In what follows, we let Φ_V and ϕ_V denote the cdf and pdf of a multivariate normal random variable in $\mathbf{R}^d$ with zero mean and covariance matrix V. We also let I_d denote the identity matrix in $\mathbf{R}^d$, and with some abuse of notation, when $d = 1$ we simply denote $\Phi_{I_1} = \Phi$ and $\phi_{I_1} = \phi$. For a random variable $U \in \mathbf{R}^d$ and $k = (k_1, \ldots, k_d)$ a vector of nonnegative integers, we let $\mathcal{X}_k(U)$ denote the k^{th}-cummulant of U. That is, for $|k| = \sum_{i=1}^d k_i$, $\mathcal{X}_k(U)$ satisfies $i^{|k|}\mathcal{X}_k(U) = \frac{\partial^{|k|}}{\partial^{k_1}t_1\ldots\partial^{k_d}t_d}\log\xi_U(t)\Big|_{t=0}$, where $i = \sqrt{-1}$ and ξ_U denotes the characteristic function of U. For

$j \in \{0,1\}$, the Cramer-Edgeworth densities $P_j(-\phi_V : \{\mathcal{X}_k(U)\})$ are $P_0(-\phi_V : \{\mathcal{X}_k(U)\})(u) = \phi_V(u)$, and:

$$P_1(-\phi_V : \{\mathcal{X}_k(U)\})(u) = -\sum_{|k|=3} \frac{\mathcal{X}_k(U)}{k!} \frac{D^{|k|}}{\partial^{k_1} u_1 \ldots \partial^{k_d} u_d} \phi_V(u) \ . \tag{72}$$

For $j \in \{0,1\}$, the Cramer-Edgeworth measure $P_j(-\Phi_V : \{\mathcal{X}_k(U)\})$ is the measure with corresponding density $P_j(-\phi_V : \{\mathcal{X}_k(U)\})$. See also Chapter 2.7 in Bhattacharya and Rao (1976) for a more general definition when $j > 1$.

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

$$P(L_n \leq z) = \Phi(z) + \frac{\phi(z)\kappa}{6\sigma^3\sqrt{n}}(2z^2+1) - \frac{\phi(z)}{\sigma^3\sqrt{n}}(c'H_0^{-1}\Sigma_0 H_0^{-1}\gamma_0(z^2+1) - \gamma_1\sigma^2) + o(n^{-\frac{1}{2}}) \ .$$

Proof: For Z as in Assumption 2.1(iv), 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) $\Omega_n \stackrel{a.s.}{\to} \Omega$ with Ω full rank, (ii) $\limsup_{n\to\infty} \frac{1}{n}\sum_i \ |a_{in}|^9 < \infty$ 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, it follows that:*

$$P^*(S_n \in B) = \sum_{j=0}^{1} \int_B dP_j(-\Phi_{I_2} : \{\mathcal{X}_k^*(S_n)\}) + o(n^{-\frac{1}{2}}) \qquad a.s.$$

uniformly over all Borel sets $B \subset \mathbf{R}^2$ with $\int_{(\partial B)^\epsilon} d\Phi_{I_2}(u) \leq C\epsilon$ 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:

$$\rho_n(t) \equiv \frac{1}{3!\|t\|^3}|\mathcal{X}_3^*(t'S_n)| = \frac{1}{3!\|t\|^3}|E^*[(t'S_n)^3]| \ , \tag{73}$$

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

$$\begin{aligned}\rho_n(t) \leq \frac{1}{n^{\frac{3}{2}}\|t\|^3}\sum_{i=1}^n E^*[|t'\Omega_n^{-\frac{1}{2}}V_{in}|^3] \leq \frac{\|\Omega_n^{-\frac{1}{2}}\|_o^3}{n^{\frac{3}{2}}}\sum_{i=1}^n E^*[\|V_{in}\|^3] \\ \leq \frac{4\|\Omega_n^{-\frac{1}{2}}\|_o^3}{n^{\frac{3}{2}}}\sum_{i=1}^n \{E^*[|a_{in}|^3|W_i|^3] + E^*[a_{in}^6|W_i^2-1|^3]\} \ .\end{aligned} \tag{74}$$

Note that $\Omega_n \stackrel{a.s.}{\to} \Omega$ with Ω full rank by hypothesis, implies $\|\Omega_n^{-\frac{1}{2}}\|_o \stackrel{a.s.}{\to} \|\Omega^{-\frac{1}{2}}\|_o < \infty$. Moreover, since $\{a_{in}\}_{i=1}^n$ is not random with respect to P^*, we obtain from condition (ii) and result (74) that almost surely:

$$\limsup_{n\to\infty}\{\sup_{t\in\mathbf{R}^2}\sqrt{n}\rho_n(t)\} \leq \limsup_{n\to\infty}\{4\|\Omega_n^{-\frac{1}{2}}\|_o^3(E[|W|^3] + E[|W^2-1|^3])\frac{1}{n}\sum_{i=1}^n\{|a_{in}|^3 + a_{in}^6\}\} < \infty \ . \tag{75}$$

Therefore, we conclude that conditions (I) and (II) of Theorem 3.4 in Skovgaard (1986) are satisfied for any sequence $\{r_n\}$ that is measurable with respect to $\{Y_i, X_i\}_{i=1}^n$ and satisfies for some $\varrho > 0$:

$$\frac{r_n}{n^{\frac{1}{2}}} \stackrel{a.s.}{\to} 0 \qquad\qquad \frac{n^{\frac{1}{2}+\varrho}}{r_n^2} \stackrel{a.s.}{\to} 0 \ . \tag{76}$$

In particular, we note that if $r_n \asymp n^{\frac{5}{18}}$ almost surely, then it satisfies (76).

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:

$$\limsup_{n\to\infty}\{\sup_{0<h<\delta n^{\frac{5}{18}}, t\in\mathbf{R}^2} |\frac{d^4}{dh^4}\log(\xi_n^*(\frac{th}{\|t\|}))| \times n^{\frac{10}{18}}\} < \infty \ . \tag{77}$$

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), Jensen's inequality, $\{a_{in}\}_{i=1}^n$ being nonrandom with respect to P^* and direct calculation it then follows that:

$$\begin{aligned}|\xi_{in}^*(t) - 1| &\leq \frac{\|t\|^2}{2n}E^*[\|\Omega_n^{-\frac{1}{2}}V_{in}\|^2] \leq \frac{\|t\|^2\|\Omega_n^{-\frac{1}{2}}\|_o}{2n}\{\sum_{i=1}^n E^*[\|V_{in}\|^{4.5}]\}^{\frac{2}{4.5}} \\ &\leq \frac{\|t\|^2\|\Omega_n^{-\frac{1}{2}}\|_o}{2n^{\frac{5}{9}}}(2^{3.5}(E[|W_i|^{4.5} + |W^2-1|^{4.5}])\frac{1}{n}\sum_{i=1}^n\{|a_{in}|^{4.5} + |a_{in}|^9\})^{\frac{2}{4.5}} \ .\end{aligned} \tag{78}$$

Condition (ii), $E[|W|^9] < \infty$ and $\|\Omega_n^{-\frac{1}{2}}\|_o \stackrel{a.s.}{\to} \|\Omega^{-\frac{1}{2}}\|_o < \infty$, then imply that almost surely there is a $\delta > 0$ with:

$$\limsup_{n\to\infty}\{\sup_{\|t\|\leq\delta n^{\frac{5}{18}}} |\xi_{in}^*(t) - 1|\} < \frac{1}{2} \ . \tag{79}$$

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:

$$\begin{aligned}\limsup_{n\to\infty}\{\sup_{0<h<\delta n^{\frac{5}{18}}, t\in\mathbf{R}^2} |\frac{d^4}{dh^4}\log(\xi_n^*(\frac{th}{\|t\|}))|\} &\leq \limsup_{n\to\infty}\{\sup_{0<h<\delta n^{\frac{5}{18}}, t\in\mathbf{R}^2}\sum_{i=1}^n|\frac{d^4}{dh^4}\log(\xi_{in}^*(\frac{th}{\|t\|}))|\} \\ \leq \limsup_{n\to\infty}\{\sup_{\|t\|\leq\delta n^{\frac{5}{18}}}\sum_{i=1}^n\sum_{|\lambda|=4}|D^\lambda\log(\xi_{in}^*(t))|\} &\leq \limsup_{n\to\infty}\{16\sum_{i=1}^n E^*[\|\frac{\Omega_n^{-\frac{1}{2}}V_{in}}{\sqrt{n}}\|^4]\} \ ,\end{aligned} \tag{80}$$

where the final inequality holds by Lemma 9.4 in Bhattacharya and Rao (1976) and result (79) implying $|\xi_{in}^*(t)-1| < \frac{1}{2}$ for all $\|t\| \leq \delta n^{\frac{5}{18}}$ and all $1 \leq i \leq n$ for n large enough. Moreover, we also have

$$\limsup_{n\to\infty}\{n^{\frac{10}{18}}\sum_{i=1}^n E^*[\|\frac{\Omega_n^{-\frac{1}{2}}V_{in}}{\sqrt{n}}\|^4]\} \leq \limsup_{n\to\infty}\{\frac{n^{\frac{10}{18}}}{n} \times \frac{2\|\Omega_n^{-\frac{1}{2}}\|_o^4}{n}\sum_{i=1}^n\{a_{in}^4E[W^4] + a_{in}^8E[(W^2-1)^4]\}\} = 0 \tag{81}$$

almost surely, by condition (i), (ii) and $E[W^8] < \infty$. It follows from (80) and (81) that (77) holds almost surely, which verifies condition (IV) of Theorem 3.4 in Skovgaard (1986) with $r_n \asymp n^{\frac{5}{18}}$.

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

$$\limsup_{n\to\infty}\{n^\alpha \times \sup_{\delta n^{\frac{5}{18}}\leq\|t\|} |\xi_n^*(t)|\} < \infty \ . \tag{82}$$

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

$$A_{in} \equiv \begin{pmatrix} a_{in} & 0 \\ 0 & a_{in}^2 \end{pmatrix} \qquad\qquad l_{3,n} \equiv \sup_{\|t\|=1}\frac{\frac{1}{n}\sum_{i=1}^n E^*[|t'A_{in}U|^3]}{[\frac{1}{n}\sum_{i=1}^n E^*[(t'A_{in}U)^2]]^{\frac{3}{2}}}\frac{1}{\sqrt{n}} \ , \tag{83}$$

where $l_{3,n}$ is the Lipaunov coefficient. Letting λ_n denote the smallest eigenvalue of Ω_n, we obtain:

$$l_{3,n} \leq \frac{1}{\sqrt{n}\lambda_n^{\frac{3}{2}}} \times \frac{1}{n}\sum_{i=1}^n E^*[\|A_{in}U\|^3] \leq \frac{1}{\sqrt{n}\lambda_n^{\frac{3}{2}}} \times \frac{4E[|W|^3 + |W^2-1|^3]}{n}\sum_{i=1}^n\{|a_{in}|^3 + a_{in}^6\} \ , \tag{84}$$

where the first inequality is (8.12) in Bhattacharya and Rao (1976), and the second was derived in (74). For λ the smallest eigenvalue of Ω, condition (i) implies $\lambda_n \stackrel{a.s.}{\to} \lambda > 0$, and hence condition (ii) implies there almost surely exists a $\tau > 0$ such that $l_{3,n} \le (\tau\sqrt{n})^{-1}$ for n large. Since $\Omega_n^{-\frac{1}{2}}$, A_{in} are not random with respect to P^*, then:

$$\sup_{\delta n^{\frac{5}{18}} \le \|t\|} |\xi_n^*(t)| \le \sup_{\delta n^{\frac{5}{18}} \le \|t\| \le \tau\sqrt{n}} |\prod_{i=1}^n \xi_U(\frac{A_{in}\Omega_n^{-\frac{1}{2}}}{\sqrt{n}} t)| + \sup_{\tau\sqrt{n} \le \|t\|} |\prod_{i=1}^n \xi_U(\frac{A_{in}\Omega_n^{-\frac{1}{2}}}{\sqrt{n}} t)| \ . \tag{85}$$

By Theorem 8.9 in Bhattacharya and Rao (1976), $|\prod_i \xi_U(\frac{A_{in}\Omega_n^{-\frac{1}{2}}}{\sqrt{n}} t)| \le \exp\{-\frac{1}{6}\|t\|^2\}$ for all $\|t\| \le l_{3,n}^{-1}$, and hence:

$$\sup_{\delta n^{\frac{5}{18}} \le \|t\| \le \tau\sqrt{n}} |\prod_{i=1}^n \xi_U(\frac{A_{in}\Omega_n^{-\frac{1}{2}}}{\sqrt{n}} t)| \le \exp\{-\frac{\delta^2}{6} n^{\frac{10}{18}}\} \ , \tag{86}$$

due to $l_{3,n} \le (\tau\sqrt{n})^{-1}$ for n large. Moreover, observe that for any $\epsilon > 0$ we also have:

$$\sup_{\tau\sqrt{n} \le \|t\|} \prod_{i=1}^n |\xi_U(\frac{A_{in}\Omega_n^{-\frac{1}{2}}}{\sqrt{n}} t)| \le \{\eta(\epsilon)\}^{\#\{i : \|A_{in}\Omega_n^{-\frac{1}{2}} t\| \ge \epsilon\sqrt{n} \ \forall \|t\| \ge \tau\sqrt{n}\}} \ . \tag{87}$$

However, since the smallest eigenvalue of $\Omega_n^{-\frac{1}{2}}$ equals $\|\Omega_n\|_o^{-\frac{1}{2}}$, it additionally follows that:

$$\#\{i : \|A_{in}\Omega_n^{-\frac{1}{2}} t\| \ge \epsilon\sqrt{n} \ \forall \|t\| \ge \tau\sqrt{n}\} \ge \#\{i : \min\{|a_{in}|, a_{in}^2\} \ge \frac{\epsilon\sqrt{n}\|\Omega_n\|_o^{\frac{1}{2}}}{\tau\sqrt{n}}\} \ . \tag{88}$$

Thus, as $\|\Omega_n\|_o^{\frac{1}{2}} \stackrel{a.s.}{\to} \|\Omega\|_o^{\frac{1}{2}} < \infty$, we may almost surely pick ϵ^* such that $\epsilon^*\|\Omega_n\|_o^{\frac{1}{2}}/\tau < \epsilon_0$ for n sufficiently large. In addition, by Assumption 2.2(ii), $\eta(\epsilon^*) < 1$ (see page 207 in Bhattacharya and Rao (1976)). Hence, by (87):

$$\sup_{\tau\sqrt{n} \le \|t\|} \prod_{i=1}^n |\xi_U(\frac{A_{in}\Omega_n^{-\frac{1}{2}}}{\sqrt{n}} t)| \le \eta(\epsilon^*)^{K_n(\epsilon_0)} \ , \tag{89}$$

for n sufficiently large. Therefore, combining (85), (86) and (89) together with condition (iii) establishes (82), thus 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 $z \in \mathbf{R}$:*

$$P^*(T_{s,n}^* \le z) = \Phi(z) + \frac{\phi(z)\hat{\kappa}E[W^3]}{6\hat{\sigma}^3\sqrt{n}}(2z^2 + 1) + o(n^{-\frac{1}{2}}) \ . \tag{90}$$

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

$$a_{in} \equiv c' H_n^{-1} X_i(Y_i - X_i\hat{\beta}) \qquad\qquad a_i \equiv c' H_0^{-1} X_i(Y_i - X_i\beta_0) \ . \tag{91}$$

Since $E[\|XX'\|_F^9] < \infty$, $E[\|XX'\epsilon^2\|_F^9] < \infty$, the law of large numbers and $\hat{\beta} \stackrel{a.s.}{\to} \beta_0$ and $\|H_n^{-1} - H_0^{-1}\|_o \stackrel{a.s.}{\to} 0$ yield:

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n |a_{in}|^9 \stackrel{a.s.}{\to} E[|c' H_0^{-1} X\epsilon|^9] < \infty \ . \tag{92}$$

Let $V_{in} \equiv (a_{in}W_i, a_{in}^2(W_i^2 - 1))'$ and $V_i \equiv (a_i W_i, a_i^2(W_i^2 - 1))'$. The same arguments as in (92) then imply:

$$\Omega_n \equiv \frac{1}{n}\sum_{i=1}^n E^*[V_{in}V_{in}'] \stackrel{a.s.}{\to} E[VV'] \ . \tag{93}$$

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, observe that for any $0 < M < \infty$, $\hat{\beta} \stackrel{a.s.}{\to} \beta$ and $\|H_n^{-1} - H_0^{-1}\| \stackrel{a.s.}{\to} 0$ imply that:

$$\sup_{\|X\epsilon\| \leq M, \|XX'\|_F \leq M} |c'H_n^{-1}X(Y - X'\hat{\beta}) - c'H_0^{-1}X(Y - X'\beta_0)| \stackrel{a.s.}{\to} 0 \ . \tag{94}$$

Moreover, since $E[(c'H_0^{-1})^2\epsilon^2] > 0$ by Assumption 2.1(iii) and $c \neq 0$, there exist a $\delta_0 > 0$ and an $M < \infty$ such that:

$$P(\min\{|(c'H_0^{-1}X)\epsilon|, (c'H_0^{-1}X)^2\epsilon^2\} \geq \delta_0 \text{ and } \max\{\|X\epsilon\|, \|XX'\|_F\} \leq M) > 0 \ . \tag{95}$$

As a consequence of result (94), it then follows that almost surely we must have:

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n 1\{\min\{|a_{in}|, a_{in}^2, \} \geq \frac{\delta_0}{2} \text{ and } \max\{\|X_i\epsilon_i\|, \|X_iX_i'\|_F\} \leq M\} \\ \geq \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^n 1\{\min\{|a_i|, a_i^2\} \geq \delta_0 \text{ and } \max\{\|X_i\epsilon_i\|, \|X_iX_i'\|_F\} \leq M\} > 0 \ . \end{aligned}\tag{96}$$

Defining $S_n \equiv \frac{1}{\sqrt{n}}\sum_i \Omega_n^{-\frac{1}{2}} V_{in}$, (92), (93) and (96) verify conditions(i)-(iii) of Lemma B.2 respectively. Therefore, we can conclude that uniformly over all Borel sets $B \subseteq \mathbf{R}^2$ with $\int_{(\partial B)^\epsilon} d\Phi_{I_2}(u) \leq C\epsilon$ for some constant C, we have:

$$P^*(S_n \in B) = \sum_{j=0}^1 \int_B dP_j(-\Phi_{I_2} : \{\mathcal{X}_k^*(S_n)\}) + o(n^{-\frac{1}{2}}) \ . \tag{97}$$

Result (97) 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:

$$f_n(t) \equiv g_n(\Omega_n^{\frac{1}{2}} t) \qquad\qquad g_n(t) \equiv t^{(1)} \times (\frac{t^{(2)}}{\sqrt{n}} + \hat{\sigma}_n^2)^{-\frac{1}{2}} \ . \tag{98}$$

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

$$\Gamma_n \equiv \{t \in \mathbf{R}^2 : \|t\| \leq \log(n)\} \ . \tag{99}$$

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 $\|\Omega_n^{-\frac{1}{2}}\|_o \log(n)/\sqrt{n} \leq \hat{\sigma}_n^2/2$ we obtain by direct calculation:

$$\limsup_{n\to\infty}\{\sqrt{n}\sup_{t\in\Gamma_n}\sup_{|\lambda|=3}|D^\lambda f_n(t)|\} \leq \limsup_{n\to\infty}\{4\sqrt{n}\|\Omega_n^{\frac{1}{2}}\|_F^3 \times \max\{\frac{3}{4n} \times \frac{2^{\frac{5}{2}}}{\hat{\sigma}_n^5}, \frac{15\|\Omega_n^{\frac{1}{2}}\|_o \log(n)}{8n^{\frac{3}{2}}} \times \frac{2^{\frac{7}{2}}}{\hat{\sigma}_n^7}\} = 0 \tag{100}$$

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

$$\limsup_{n\to\infty}\sqrt{n}\|\nabla^2 f_n(0)\|_F^2 = \limsup_{n\to\infty}\sqrt{n}\|\Omega_n^{\frac{1}{2}}\nabla^2 g_n(0)\Omega_n^{\frac{1}{2}}\|_F^2 \leq \limsup_{n\to\infty}\{\sqrt{n}\|\Omega_n^{\frac{1}{2}}\|_F^2 \times \frac{1}{2n\hat{\sigma}_n^6}\} = 0 \tag{101}$$

almost surely, verifying condition (3.12) of Theorem 3.2 in Skovgaard (1981). Therefore, we conclude from (97), (100), (101), 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 $\int_{(\partial B)^\epsilon} d\Phi(u) = O(\epsilon)$ (which includes all sets of the form $(-\infty, z]$)). In particular, (90) 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). ■

Proof of Theorem 2.4: The proof relies on Lemmas C.1 and C.3 in the Supplemental Appendix, which establish uniform versions (in $P \in \mathbf{P}$) of the Edgeworth expansions in Theorem 2.3. First define:

$$\Delta_n(z,P) \equiv \frac{\phi(z)}{\sigma(P)^3}(c'H_0(P)^{-1}\Sigma_0(P)H_0(P)^{-1}\gamma_0(P)(z^2+1) - \gamma_1(P)\sigma^2(P))$$
$$\mathcal{E}_n(z,P) \equiv \Phi(z) + \frac{\phi(z)\kappa(P)}{6\sigma(P)^3\sqrt{n}}(2z^2+1) - \frac{\phi(z)}{\sigma(P)^3\sqrt{n}}(c'H_0(P)^{-1}\Sigma_0(P)H_0(P)^{-1}\gamma_0(P)(z^2+1) - \gamma_1(P)\sigma^2(P))$$
$$\mathcal{E}_n^*(z) \equiv \Phi(z) + \frac{\phi(z)E[W^3]}{6\sqrt{n}}(2z^2+1) \times (\frac{|\hat{\kappa}|}{\hat{\sigma}^3} \wedge C_0) \times \text{sign}\{\hat{\kappa}\}\ ,$$

where $(\sup_{P\in\mathbf{P}} |\kappa(P)|)/(\inf_{P\in\mathbf{P}} \sigma(P)^3) < C_0$. Note that $\sup_{z\in\mathbf{R}} |\phi(z)z^2| < \infty$, while by Assumptions 2.3(ii)-(iii) and $c \neq 0$, $\sigma(P)^{-1}$, $\|H_0(P)^{-1}\|_o$, $\sigma^2(P)$ and $\|\Sigma_0(P)\|_o$ are bounded in $P \in \mathbf{P}$. Therefore, there exist M_0, M_1 such that:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} \sup_{z\in\mathbf{R}} |\Delta_n(z,P)| \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} \{M_0\|\gamma_0(P)\| + M_1|\gamma_1(P)|\} = 0 \qquad (102)$$

where the final equality follows from $P \mapsto \gamma_0(P)$ and $P \mapsto \gamma_1(P)$ being continuous under $\|\cdot\|_{\mathbf{P}}$ by Assumption 2.4(ii), and $\gamma_0(P_0) = 0$, $\gamma_1(P_0) = 0$ due to $E_{P_0}[Y|X] = X'\beta_0(P_0)$. Similarly, by continuity of $P \mapsto \sigma(P)$ and $P \mapsto \kappa(P)$:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} \sup_{z\in\mathbf{R}} |\frac{\phi(z)\kappa(P)}{6\sigma(P)^3}(2z^2+1) - \frac{\phi(z)\kappa(P_0)}{6\sigma(P_0)^3}(2z^2+1)| \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} M_2|\frac{\kappa(P)}{\sigma(P)^3} - \frac{\kappa(P_0)}{\sigma(P_0)^3}| = 0 \quad (103)$$

for some $M_2 < \infty$. Therefore, combining (102), (103), Lemma C.1 and the continuity and monotonicity of $a \mapsto L(a)$:

$$\begin{aligned}
&\liminf_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} L(\sup_{z\in\mathbf{R}} \sqrt{n}|P(T_n \leq z) - \Phi(z)|) \\
&\quad \geq \liminf_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} L(\max\{\sup_{z\in\mathbf{R}}\{|\frac{\phi(z)\kappa(P)}{6\sigma(P)^3}(2z^2+1)| - |\Delta_n(z,P)| - \sqrt{n}|P(T_n \leq z) - \mathcal{E}_n(z,P)|\}, 0\}) \\
&\quad = L(\sup_{z\in\mathbf{R}} |\frac{\phi(z)\kappa(P_0)}{6\sigma(P_0)^3}(2z^2+1)|)\ . \qquad (104)
\end{aligned}$$

Moreover, by Lemma C.3, there exist sets A_n such that $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_n^c) = O(n^{-\frac{\tilde{\nu}}{18}})$, and:

$$\sup_{z\in\mathbf{R}} |P^*(T_n^* \leq z) - \mathcal{E}_n^*(z)| \leq \delta_n \qquad (105)$$

for some deterministic $\delta_n = o(n^{-\frac{1}{2}})$, whenever $\{Y_i, X_i\}_{i=1}^n \in A_n$. Furthermore, since $\limsup_{a\to\infty} L(a)a^{-\vartheta} < \infty$ by hypothesis, it follows that there exists a $C > 0$ such that $L(a) \leq Ca^{\vartheta}$ for a sufficiently large. Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} E_P[L(\sup_{z\in\mathbf{R}} \sqrt{n}|P(T_n \leq z) - P^*(T_n^* \leq z)|)1\{\{Y_i, X_i\}_{i=1}^n \in A_n^c\}] \\
&\qquad \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} L(\sqrt{n})P(\{Y_i, X_i\}_{i=1}^n A_n^c) \leq \limsup_{n\to\infty} Cn^{\frac{\vartheta}{2}} \times O(n^{-\frac{\tilde{\nu}}{18}}) = 0\ , \quad (106)
\end{aligned}$$

where the final equality holds by $9\vartheta < \tilde{\nu}$. Hence, by Lemma C.1, (105) and (106), for some deterministic $\gamma_n = o(1)$:

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} E_P[L(\sup_{z\in\mathbf{R}} \sqrt{n}|P(T_n \leq z) - P^*(T_n \leq z)|)] \\
&\quad \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} E_P[L(\sup_{z\in\mathbf{R}}\{\sqrt{n}|\Phi(z) + \frac{\phi(z)\kappa(P)}{6\sigma(P)^3\sqrt{n}}(2z^2+1) - \mathcal{E}_n^*(z)| + |\Delta_n(z,P)| + \gamma_n\})] + o(1) \\
&\quad \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} E_P[L(M_2|\frac{\kappa(P)}{\sigma(P)^3} - (\frac{|\hat{\kappa}|}{\hat{\sigma}^3} \wedge C_0) \times \text{sign}\{\hat{\kappa}\}| + \tilde{\gamma}_n)] + o(1)\ , \qquad (107)
\end{aligned}$$

where the final inequality holds for some deterministic $\tilde{\gamma}_n = o(1)$ due to (103) and $E[W^3] = 1$. By Lemma C.3:

$$\sup_{P\in\mathbf{P}} P(|\frac{\kappa(P)}{\sigma(P)^3} - \frac{\hat{\kappa}}{\hat{\sigma}^3}| > \epsilon) = O(n^{-\frac{\tilde{\nu}}{8}}) \ , \tag{108}$$

for any $\epsilon > 0$. Therefore, since $\sup_{P\in\mathbf{P}} |\kappa(P)|/\sigma(P)^3 < C_0$, results (107), (108), $\tilde{\gamma}_n = o(1)$ and continuity of L imply:

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{P\in\mathbf{P}(P_0,h_n)} E_P[L(M_2|\frac{\kappa(P)}{\sigma(P)^3} - (\frac{|\hat{\kappa}|}{\hat{\sigma}^3} \wedge C_0) \times \text{sign}\{\hat{\kappa}\}| + \tilde{\gamma}_n)] \\ &\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \{L(M_2\epsilon + \tilde{\gamma}_n) + L(2M_2C_0 + \tilde{\gamma}_n) \times P(|\frac{\kappa(P)}{\sigma(P)^3} - \frac{\hat{\kappa}}{\hat{\sigma}^3}| > \epsilon)\} = L(M_2\epsilon) \ . \end{aligned} \tag{109}$$

Hence, since ϵ in (109) was arbitrary, the first claim of the Theorem follows from (104), (107) and (109). Finally, note that if $\kappa(P_0) \neq 0$, then $\sup_{z\in\mathbf{R}} |\frac{\phi(z)\kappa(P_0)}{6\sigma(P_0)^3}(2z^2+1)| > 0$, and hence (11) holds strictly by $L : [0, +\infty) \to [0, +\infty)$ being strictly increasing and setting ϵ sufficiently small in (109). ■

References

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

Beran, R. 1982, Estimated Sampling Distributions: The Bootstrap and Competitors. The Annals of Statistics 10, 212–225.

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.

Bhattacharya R. and Qumsiyeh M. 1989, Second Order and L^p-Comparisons between the Bootstrap and Empirical Edgeworth Expansion Methodologies. The Annals of Statistics 17, 160–169.

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.

Chesher, A. 1987, A Mirror Image Invariance for M-Estimators, Econometrica 63, 207-211.

Chesher, A. and Jewitt, I., 1987, The Bias of a Heteroscedasticity Consistent Covariance Matrix Estimator, Econometrica 55, 1217–1222.

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.

Durrett, R. 1996, Probability: Theory and Examples Second Edition. Duxbury Press, Belmont.

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.

Hall, P., and Horowitz, J. L., 1996, Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. Econometrica, 64, 891–916.

Heckman, J. J.,, Lochner, L. J.,, and Todd, P. E, 2006, Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond in: E. A. Hanushek and F. Welch, (Eds.), Handbook of Education Economics, Vol. 1, Chap. 7. Elsevier.

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

Horowitz, J. L., 2001, The bootstrap. in: J. J. Heckman and E. Leamer, (Eds.), Handbook of Econometrics, Vol. 5, Chap. 52. Elsevier.

Kaye, D. H. and Freedman, D. A., 1981, Reference Guide on Statistics. in Reference Manual on Scientific Evidence, Federal Judicial Center Accessed online at: http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf

Kline, P., and Santos, A., 2011, A Score Based Approach to Wild Bootstrap Inference. The Journal of Econometric Methods 1, in press.

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

Mackinnon, J. G. and White, H., 1985, Some Heteroscedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. Journal of Econometrics 29, 305–325.

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

Mincer, J., 1974, Schooling, Experience and Earnings. Columbia University Press for National Bureau of Economic Research, New York

Ruggles, S. J., Alexander, T. J., Genadek, K., Goeken, R., Schroeder, M. B., and Sobek, M., 2010, Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Singh K. and Babu G. J., 1990, On Asymptotic Optimality of the Bootstrap. Scandinavian Journal of Statistics 17, 1–9.

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

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

Stock, J. H., 2010, The Other Transformation in Econometric Practice: Robust Tools for Inference. Journal of Economic Perspectives 24, 83-94.

White, H., 1980a, Using Least Squares to Approximate Unknown Regression Functions. International Economic Review 21, 149–170.

White, H., 1980b, A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. Econometrica 48, 817–838.

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.

Supplemental Appendix - Auxiliary Lemmas for the proof of Theorem 2.4

Throughout Appendix C, we employ the notation of Section 2.4, which emphasizes the dependence on $P \in \mathbf{P}$.

Lemma C.1. *Let Assumptions 2.3 hold, and denote the Edgeworth expansion for $P(T_n \leq z)$ by:*

$$\mathcal{E}_n(z,P) \equiv \Phi(z) + \frac{\phi(z)\kappa(P)}{6\sigma(P)^3\sqrt{n}}(2z^2+1) - \frac{\phi(z)}{\sigma(P)^3\sqrt{n}}(c'H_0(P)^{-1}\Sigma_0(P)H_0(P)^{-1}\gamma_0(P)(z^2+1) - \gamma_1(P)\sigma^2(P)) \ . \quad (110)$$

If $c \neq 0$, then it follows that $\limsup_{n\to\infty}\sup_{P\in\mathbf{P}}\sup_{z\in\mathbf{R}}\sqrt{n}|P(T_n \leq z) - \mathcal{E}_n(z,P)| = 0$.

Proof: For fixed $P \in \mathbf{P}$, the validity of the Edgeworth expansion has already been established in Theorem 2.3. We establish the Lemma by showing Assumption 2.3 controls all approximation errors uniformly. Specifically, in lieu of (15) and (16) note that with $\tilde{\nu}$ in place of ν: Lemma A.2(i) holds uniformly in $P \in \mathbf{P}$ due to $\sup_{P\in\mathbf{P}} E_P[\|X\epsilon\|^{\tilde{\nu}}] < \infty$ by Assumption 2.3(ii); Lemma A.2(ii) holds uniformly in $P \in \mathbf{P}$ due to Assumptions 2.3(ii)-(iii) implying:

$$0 < \inf_{P\in\mathbf{P}} \|H_0(P)^{-1}\|_o < \sup_{P\in\mathbf{P}} \|H_0(P)^{-1}\|_F < \infty \ , \quad (111)$$

and $\sup_{P\in\mathbf{P}} E_P[\|XX'\|_F^{\tilde{\nu}}] < \infty$ by Assumption 2.3(ii); Lemma A.2(iii) holds uniformly in $P \in \mathbf{P}$ by (111); and Lemma A.2(iv) holds uniformly in $P \in \mathbf{P}$ by $E_P[\|XX'\|_F^{\tilde{\nu}}]$, $E_P[\|XX'\epsilon^2\|_F^{\tilde{\nu}}]$, $E_P[\|(c'H_0(P)^{-1}X)^2\epsilon X\|^{\frac{\tilde{\nu}}{2}}]$, $\|\gamma_0(P)\|$ and $\|\Sigma_0(P)\|_F$ being uniformly bounded in $P \in \mathbf{P}$ by Assumptions 2.3(ii)-(iii) and result (111). Similarly, since $\inf_{P\in\mathbf{P}} \sigma(P) > 0$ by Assumption 2.3(iii), we get by result (111) and $\sup_{P\in\mathbf{P}} \|\gamma_0(P)\| < \infty$ by Assumptions 2.3(ii)-(iii), that the arguments in Lemma A.3 hold uniformly in $P \in \mathbf{P}$. Therefore we obtain for any $\alpha \in [0, \frac{2\tilde{\nu}-3}{2\tilde{\nu}})$:

$$\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} \sqrt{n}P(|T_n - L_n(P)| > n^{-\alpha}) = 0 \ . \quad (112)$$

Let $Z \in \mathbf{R}^{d_z}$ be as in Assumption 2.3(iv), set $S_n(P) = \frac{1}{\sqrt{n}}\sum_i (Z_i - E_P[Z_i])$, $V(P) = E_P[ZZ']$ and $\Phi_{V(P)}$ to be a mean zero Gaussian measure on $\mathbf{R}^{d_z}$ with covariance $V(P)$. For $\mathcal{X}_k(S_n(P))$ the k^{th} cumulant of $S_n(P)$ under P, and P_j the Cramer-Edgeworth measures we next aim to show that for any Borel set B and all $P \in \mathbf{P}$:

$$|P(S_n(P) \in B) - \sum_{j=0}^{1}\int_B dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\})| \leq \delta_n + \Phi_{V(P)}((\partial B)^{2e^{-dn}}) \quad (113)$$

where $\delta_n = o(n^{-\frac{1}{2}})$ and $d > 0$ are independent of B and P. The validity of the Edgeworth expansion in (113) pointwise in $P \in \mathbf{P}$ is immediate from Assumption 2.3 and Theorem 20.1 in Bhattacharya and Rao (1976). Most of their error bounds can be controlled uniformly by $\sup_{P\in\mathbf{P}} E_P[\|Z\|^4] < \infty$. The only necessary modifications to their arguments is in their equation (20.22) which can be controlled uniformly due to $\inf_{P\in\mathbf{P}} \lambda(E_P[ZZ']) > 0$ by Assumption 2.3(iv), and in their equations (20.29)-(20.34), which can be controlled uniformly in $P \in \mathbf{P}$ since:

$$\sup_{\|t\|\geq \frac{\sqrt{n}}{16E_P[\|Z\|^3]}} |\xi_{Z,P}(t/\sqrt{n})| \leq \sup_{\|t\|\geq (16\sup_{P\in\mathbf{P}} E_P[\|Z\|^3])^{-1}} |\xi_{Z,P}(t)| \leq \sup_{\|t\|\geq (16\sup_{P\in\mathbf{P}} E_P[\|Z\|^3])^{-1}} F(t) < 1 \ , \quad (114)$$

due to Assumption 2.3(iv). The remaining arguments in establishing (113) are identical to their proof and therefore omitted; see also Lemma 2 in Singh and Babu (1990) for the univariate case.

Next, let $G_P : \mathbf{R}^{d_z} \to \mathbf{R}$ be such that $L_n(P) = \sqrt{n} G_P(\frac{1}{n}\sum_i Z_i)$, and note $G_P(E_P[Z]) = 0$. Further define $g_{n,P}(z) = \sqrt{n} G_P(E_P[Z] + z/\sqrt{n})$ and note $L_n(P) = g_{n,P}(S_n(P))$. Exploiting result (113), we aim to establish that:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{z\in\mathbf{R}} \sqrt{n}|P(L_n(P) \le z) - \mathcal{E}_n(z,P)| = \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{z\in\mathbf{R}} \sqrt{n}|P(g_{n,P}(S_n(P)) \le z) - \mathcal{E}_n(z,P)| = 0 \quad (115)$$

The validity of (115) pointwise in P follows from Assumption 2.3 and Theorem 2 in Bhattacharya and Ghosh (1978). The arguments leading to a uniform result are similar, and we describe only the necessary modifications. To this end, let $K > 0$ satisfy $\sup_{P\in\mathbf{P}} \|E_P[ZZ']\|_F < K < \infty$, which is feasible by Assumption 2.3(ii), and define $M_n \equiv \{z \in \mathbf{R}^{d_z} : \|z\| < K\log(n)\}$. By Assumption 2.3(ii), $\{\mathcal{X}_k(S_n(P))\}_{k=1}^3$ are bounded in $P \in \mathbf{P}$, and hence:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sum_{j=0}^{1} \sqrt{n}|\int_{(M_n)^c} dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\})| = 0 \ . \quad (116)$$

Since in addition $\nabla g_{n,P}(\tilde{z})$ is uniformly bounded on $(\tilde{z}, P) \in M_n \times \mathbf{P}$ and n by Assumption 2.3(ii)-(iii), Lemma 2.1 in Bhattacharya and Ghosh (1978) holds uniformly in $P \in \mathbf{P}$. For each $z \in \mathbf{R}$, then define the set $A_{n,P}(z) \equiv \{\tilde{z} \in \mathbf{R}^{d_z} : g_{n,P}(\tilde{z}) \le z\}$ and note that by continuity $\partial A_{n,P}(z) \subseteq \{\tilde{z} \in \mathbf{R}^{d_z} : g_{n,P}(\tilde{z}) = z\}$. Moreover, $\nabla g_{n,P}(\tilde{z})$ being uniformly bounded on $M_n \times \mathbf{P}$ further implies that if $\tilde{z} \in \partial A_{n,P}(z) \cap M_n$, $\tilde{z}' \in M_n$, and $\|\tilde{z} - \tilde{z}'\| \le \epsilon$, then by the mean value theorem $g_{n,P}(\tilde{z}') \in z \pm M\epsilon$ for some M not depending on P, $\tilde{z}$ or n. Hence, $(\partial A_{n,P}(z))^\epsilon \cap M_n \subseteq \{\tilde{z} \in \mathbf{R}^{d_z} : g_{n,P}(\tilde{z}') \in z \pm M\epsilon\}$, and since $\sup_{P\in\mathbf{P}} \int_{M_n^c} d\Phi_{V(P)}(\tilde{z}) = o(n^{-\frac{1}{2}})$ by (116), we conclude:

$$\begin{aligned}\int_{(\partial A_{n,P}(z))^{2e^{-dn}}} d\Phi_{V(P)}(\tilde{z}) &= \int_{(\partial A_{n,P}(z))^{2e^{-dn}} \cap M_n} d\Phi_{V(P)}(\tilde{z}) + o(n^{-\frac{1}{2}}) \\ &\le 2\sum_{j=0}^{1} \int_{\{\tilde{z}: g_{n,P}(\tilde{z}) \in z \pm M\epsilon\}} dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\}) + o(n^{-\frac{1}{2}}) \le O(e^{-dn}) + o(n^{-\frac{1}{2}}) \ , \end{aligned} \quad (117)$$

where the first inequality holds for n large enough uniformly in P by arguing as in (20.37) in Bhattacharya and Rao (1976), while the second inequality holds by Lemma 2.1 in Bhattacharya and Ghosh (1978), Corollary 3.2 in Bhattacharya and Rao (1976) and Assumptions 2.3(ii)-(iv). Therefore, by (113) and (117):

$$\sup_{P\in\mathbf{P}} \sup_{z\in\mathbf{R}} |P(L_n(P) \le z) - \sum_{j=0}^{1} \int_{A_{n,P}(z)} dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\})| = o(n^{-\frac{1}{2}}) \ , \quad (118)$$

where we have used that $L_n(P) \le z$ if and only if $S_n(P) \in A_{n,P}(z)$. Replacing equation (2.20) in Bhattacharya and Ghosh (1978) with result (118), claim (115) then follows using the same arguments in the proof of Theorem 2 in Bhattacharya and Ghosh (1978) and noting that due to Assumption 2.3(ii)-(iii) the arguments in Lemmas A.8 and A.9 hold uniformly in $P \in \mathbf{P}$. The claim of the Lemma then follows from (112), (118), Assumptions 2.3(ii)-(iii) implying the coefficients in $\mathcal{E}_n(\cdot, P)$ are bounded in $P \in \mathbf{P}$ and Lemma 5 in Andrews (2002). ■

Lemma C.2. *Let Assumptions 2.2(i)-(ii), 2.3(i)-(iii) hold and $T^*_{s,n}$ be as in* (47). *It then follows that for any $9 \le \zeta \le 2\tilde{\nu}$, and $\alpha \in [0, \frac{(2\omega)\wedge\zeta - 2}{(2\omega)\wedge\zeta} - \frac{1}{2(\omega\wedge\zeta)})$ there exists a deterministic sequence $\delta_n = o(n^{-\frac{1}{2}})$ and sets $A_n \subseteq \mathbf{R}^{n(d_x+1)}$ such that $P^*(|T^*_{s,n} - T^*_n| > n^{-\alpha}) \le \delta_n$ whenever $\{Y_i, X_i\}_{i=1}^n \in A_n$ and $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin A_n) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$.*

Proof: Let K_0 satisfy $\sup_{P\in\mathbf{P}}\{\|H_0(P)^{-1}\|_o^\zeta E_P[\|X\epsilon\|^\zeta]\} < K_0 < \infty$ which is possible by Assumption 2.3(ii)-(iii), and:

$$A_{0n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \frac{1}{n}\sum_{i=1}^n \|H_n^{-1}\|_o^\zeta \|X_i(Y_i - X_i'\hat{\beta})\|^\zeta < K_0\} \ . \quad (119)$$

For any $\alpha_0 \in [0, \frac{\omega\wedge\zeta-1}{2(\omega\wedge\zeta)})$, we then obtain from (41) together with (39) and (40) that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{0n}$,

$$P^*(\|\hat{\beta}^* - \hat{\beta}\| > n^{-\alpha_0}) \leq \frac{C_0 K_0}{n^{(\frac{1}{2}-\alpha_0)(\omega\wedge\zeta)}} \tag{120}$$

for some constant $C_0 > 0$. Similarly, let $\sup_{P\in\mathbf{P}}\{(2d_x^2)^{\frac{\zeta}{2}}\|c\|^{\zeta}\|H_0(P)^{-1}\|_o^{\zeta} E_P[\|X\|^{\frac{3\zeta}{2}}|\epsilon|^{\frac{\zeta}{2}}]\} < K_1 < \infty$, and:

$$A_{1n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \frac{(2d_x^2)^{\frac{\zeta}{2}}\|c\|^{\zeta}}{n}\sum_{i=1}^n \|H_n^{-1}\|_o^{\zeta}\|X_i\|^{\frac{3\zeta}{2}}|(Y_i - X_i'\hat{\beta})|^{\frac{\zeta}{2}} < K_1\} \ . \tag{121}$$

For $X_i^{(l)}$ the l^{th} coordinate of X_i, we obtain by (39) and (40) that for any $1 \leq j \leq k \leq d_x$ and $\alpha_1 \in [0, \frac{\omega\wedge(\zeta/2)-1}{2(\omega\wedge(\zeta/2))})$:

$$P^*(\|c\|^2\|H_n^{-1}\|_o^2\|\frac{2d_x^2}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(k)}X_i\epsilon_i^*\| > n^{-\alpha_1}) \leq \frac{C_1 K_1}{n^{(\frac{1}{2}-\alpha_1)(\omega\wedge(\zeta/2))}} \tag{122}$$

for some $C_1 > 0$ whenever $\{Y_i, X_i\}_{i=1}^n \in A_{1n}$. Set $\sup_{P\in\mathbf{P}}\{\|c\|^2\|H_0(P)^{-1}\|_o^2 E_P[\|XX'\|_F\|X\|^2]\} < K_2 < \infty$, and:

$$A_{2n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \|c\|^2\|H_n^{-1}\|_o^2\frac{1}{n}\sum_{i=1}^n \|X_iX_i'\|_F\|X_i\|^2 < K_2\} \ . \tag{123}$$

We then obtain from (42), (43), (44), together with (120) and (121) that for any $\alpha_1 \in [0, \frac{\omega\wedge(\zeta/2)-1}{2(\omega\wedge(\zeta/2))})$ there exists a constant $C_2 > 0$ (depending on K_0, K_1, K_2, ω and ζ) such that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{0n} \cap A_{1n} \cap A_{2n}$:

$$P^*(|(\hat{\sigma}^*)^2 - (\hat{\sigma}_s^*)^2| > n^{-\alpha_1}) \leq \frac{C_2}{n^{(\frac{1}{2}-\alpha_1)(\omega\wedge(\zeta/2))}} \ . \tag{124}$$

Let $\sup_{P\in\mathbf{P}}\{\|c\|^4\|H_0(P)^{-1}\|_o^4 E_P[\|XX'\epsilon^2\|_F^2]\} < K_3 < \infty$ which is possible by Assumption 2.3(ii), and define:

$$A_{3n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \|c\|^4\|H_n^{-1}\|_o^4\frac{1}{n}\sum_{i=1}^n \|X_iX_i'(Y_i - X_i\hat{\beta})^2\|_F^2 < K_3\} \ . \tag{125}$$

The inequalities (39) and (40) then imply that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{3n}$, for any $\epsilon > 0$ we obtain that:

$$P^*(|(\hat{\sigma}_s^*)^2 - \hat{\sigma}^2| > \epsilon) \leq \frac{C_3}{\epsilon^2 n} \ . \tag{126}$$

Therefore, setting $\inf_{P\in\mathbf{P}}\sigma^2(P) > \epsilon_0 > 0$, which is feasible by Assumption 2.3(iii) and letting $A_{4n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \hat{\sigma}^2 > \epsilon_0\}$, we obtain from (126) that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{3n} \cap A_{4n}$ we must have:

$$P^*((\hat{\sigma}_s^*)^2 < \epsilon_0/2) \leq \frac{2C_3}{\epsilon_0 n} \ . \tag{127}$$

Letting $A_n = \bigcap_{j=0}^4 A_{jn}$, we then obtain from (48) together with (120), (126) and (127) and Assumptions 2.2(ii), 2.3(ii) that the desired deterministic sequence $\delta_n = o(n^{-\frac{1}{2}})$ exists.

To conclude the proof, we next show that $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_n^c) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$. To this end, note that:

$$\sup_{P\in\mathbf{P}} P(\|H_n - H_0(P)\|_F > \eta) = O(n^{-\frac{\tilde{\nu}}{2}}) \tag{128}$$

for any $\eta > 0$ due to (15), (16) and Assumption 2.3(ii). Moreover, since $\sup_{P\in\mathbf{P}}\|H_0(P)^{-1}\|_o > 0$ by Assumption 2.3(iii), (128) implies $\sup_{P\in\mathbf{P}} P(\|H_0(P)^{-1}(H_n - H_0(P))\|_F > \eta) = O(n^{-\frac{\tilde{\nu}}{2}})$, and therefore (18) and (19) yield:

$$\sup_{P\in\mathbf{P}} P(\|H_n^{-1} - H_0(P)^{-1}\|_F > \eta) = O(n^{-\frac{\tilde{\nu}}{2}}) \ . \tag{129}$$

Therefore, by (129) and Assumption 2.3(iii) there exists an $M_0 > 0$ such that $\sup_{P\in\mathbf{P}} P(\|H_n^{-1}\|_F > M_0) = O(n^{-\frac{\tilde{\nu}}{2}})$. It then follows by Assumption 2.3(ii) and (15) and (16), that for any $\eta > 0$ we have:

$$\sup_{P\in\mathbf{P}} P(\|\hat{\beta} - \beta_0\| > \eta) \leq \sup_{P\in\mathbf{P}} P(\|\frac{1}{n}\sum_{i=1}^n X_i\epsilon_i\| > \frac{\epsilon}{M_0}) + O(n^{-\frac{\tilde{\nu}}{2}}) = O(n^{-\frac{\tilde{\nu}}{2}}) \ . \tag{130}$$

Since (129), the mean value theorem and Assumption 2.3(iii) yield $\sup_{P\in\mathbf{P}} P(|\|H_n^{-1}\|_o^\zeta - \|H_0(P)^{-1}\|_o^\zeta| > \eta) = O(n^{-\frac{\tilde{\nu}}{2}})$, and $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_iX_i'\|_F^\zeta - E_P[\|XX'\|_F^\zeta]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$ by Assumption 2.3(ii) and (15)-(16):

$$\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^n \|H_n^{-1}\|_o^\zeta \|X_i(Y_i - X_i'\hat{\beta})\|^\zeta - \|H_0(P)^{-1}\|_o^\zeta E_P[\|X\epsilon\|^\zeta]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}}) \tag{131}$$

due to (130). Since (131) holds for any $\eta > 0$, the definition of A_{0n} and the constant K_0 in turn imply that:

$$\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_{0n}^c) \leq \sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^n \|H_n^{-1}\|_o^\zeta \|X_i(Y_i - X_i'\hat{\beta})\|^\zeta - \|H_0(P)^{-1}\|_o^\zeta E_P[\|X\epsilon\|^\zeta]|$$
$$> K_0 - \sup_{P\in\mathbf{P}} \|H_0(P)^{-1}\|_o^\zeta E_P[\|X\epsilon\|^\zeta]) = O(n^{-\frac{\tilde{\nu}}{2\zeta}}) \ . \tag{132}$$

Analogously, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_i\|^{2\zeta} - E_P[\|X\|^{2\zeta}]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$ due to (15), (16) and Assumption 2.3(ii) implying $\sup_{P\in\mathbf{P}} E_P[(\|X\|^{2\zeta})^\delta] < \infty$ for any $\delta \leq \tilde{\nu}/\zeta$. Similarly, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_i\|^\zeta \|X_i\epsilon_i\|^{\frac{\zeta}{2}} - E_P[\|X\|^\zeta \|X\epsilon\|^{\frac{\zeta}{2}}]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$, and therefore from (129), (130) and arguing as in (131) and (132):

$$\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_{1n}^c) = O(n^{-\frac{\tilde{\nu}}{2\zeta}}) \ . \tag{133}$$

The same arguments, but bounding $\sup_{P\in\mathbf{P}} E_P[(\|XX'\|_F\|X\|^2)^\delta]^2 \leq \sup_{P\in\mathbf{P}}\{E_P[\|XX'\|_F^{2\delta}]E_P[\|X\|^{4\delta}]\} < \infty$ for $\delta \leq \tilde{\nu}/2$, and $\sup_{P\in\mathbf{P}} E_P[(\|XX'X\|^2)^\delta]^2 \leq \sup_{P\in\mathbf{P}}\{E_P[\|XX'\|^{4\delta}]E_P[\|X\|^{4\delta}]\} < \infty$ for $\delta \leq \tilde{\nu}/4$, yields:

$$\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_{2n}^c) = O(n^{-\frac{\tilde{\nu}}{4}}) \qquad \sup_{P\in\mathbf{P}} \max\{P(\{Y_i, X_i\}_{i=1}^n \in A_{3n}^c), P(\{Y_i, X_i\}_{i=1}^n \in A_{4n}^c)\} = O(n^{-\frac{\tilde{\nu}}{8}}) \tag{134}$$

The lemma then follows from $P(\{Y_i, X_i\}_{i=1}^n \in A_n^c) \leq \sum_{j=1}^4 P(\{Y_i, X_i\}_{i=1}^n \in A_{jn}^c)$, (132), (133) and (134). ■

Lemma C.3. *Let Assumptions 2.2, 2.3(i)-(iii) hold, and* $(\sup_{P\in\mathbf{P}} |\kappa(P)|)/(\inf_{P\in\mathbf{P}} \sigma(P)^3) < C_0$. *In addition, denote*

$$\mathcal{E}_n^*(z) \equiv \Phi(z) + \frac{\phi(z)E[W^3]}{6\sqrt{n}}(2z^2 + 1) \times (\frac{|\hat{\kappa}|}{\hat{\sigma}^3} \wedge C_0) \times \mathit{sign}\{\hat{\kappa}\} \ , \tag{135}$$

then there exist a deterministic $\delta_n = o(n^{-\frac{1}{2}})$ *and sets* $A_n \subset \mathbf{R}^{n(1+d_x)}$ *such that* $\sup_{z\in\mathbf{R}} |P^*(T_n \leq z) - \mathcal{E}_n^*(z)| \leq \delta_n$ *whenever* $\{Y_i, X_i\}_{i=1}^n \in A_n$ *and in addition* $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin A_n) = O(n^{-\frac{\tilde{\nu}}{18}})$. *Additionally, for any* $\epsilon > 0$:

$$\sup_{P\in\mathbf{P}} P(|\frac{\kappa(P)}{\sigma(P)^3} - \frac{\hat{\kappa}}{\hat{\sigma}^3}| > \epsilon) = O(n^{-\frac{\tilde{\nu}}{8}}) \tag{136}$$

Proof: We first proceed as in Lemmas B.2 and B.3 by verifying the conditions of Theorems 3.4 in Skovgaard (1986) and 3.2 in Skovgaard (1981) respectively. Throughout, let $a_{in} \equiv c'H_n^{-1}X_i(Y_i - X_i\hat{\beta})$, $V_{in} \equiv (a_{in}W_i, a_{in}^2(W_i^2 - 1))$, $\Omega_n \equiv \frac{1}{n}\sum_i E^*[V_{in}V_{in}']$ and $S_n \equiv \frac{1}{\sqrt{n}}\sum_i \Omega_n^{-\frac{1}{2}} V_{in}$. We first aim to show there exist sets B_n such that $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_n) = O(n^{-\frac{\tilde{\nu}}{18}})$, and that there exists a deterministic sequence $b_n = o(n^{-\frac{1}{2}})$ satisfying:

$$P^*(S_n \in B) = \sum_{j=0}^1 \int_B dP_j(-\Phi_{I_2} : \{\mathcal{X}_k^*(S_n)\}) + b_n \ , \tag{137}$$

uniformly over all Borel sets B with $\int_{(\partial B)^\epsilon} d\Phi_{I_2}(u) \leq C\epsilon$ whenever $\{Y_i, X_i\}_{i=1}^n \in B_n$. To this end, let $a_i \equiv c'H_0(P)^{-1}X_i\epsilon_i$, $V_i \equiv (a_iW_i, a_i^2(W_i^2-1))$ and $\Omega(P) \equiv E_P[VV']$. By Assumption 2.3(ii)-(iii) and Exercise 3.8 in Durrett (1996), there exists a $1 > \delta_0 > 0$ such that $\inf_{P\in\mathbf{P}} P(|a_i|^2 > \delta_0) > 0$, and hence by Assumption 2.3(iii):

$$\inf_{P\in\mathbf{P}} P(|a_i|^2 > \delta_0 \text{ and } \max\{\|X\epsilon\|, \|XX'\|_F\} \leq M_0) > \epsilon_0 \tag{138}$$

for some $M_0 < \infty$ and some $\epsilon_0 > 0$. We can now define the sequence of sets B_n, by $B_n = \bigcap_{j=0}^4 B_{jn}$, where:

$$\begin{aligned}
B_{0n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : \sup_{t\in\mathbf{R}^2} \tfrac{1}{3!\|t\|^3}|E^*[(t'S_n)^3]| \leq n^{-\frac{5}{18}}\} \\
B_{1n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : \|\Omega_n^{-\frac{1}{2}}\|_o(\tfrac{1}{n}\textstyle\sum_i\{|a_{in}|^{4.5} + |a_{in}|^9\})^{\frac{2}{4.5}} < 2\sup_{P\in\mathbf{P}}\{\|\Omega(P)^{-\frac{1}{2}}\|_o(E_P[|a_i|^{4.5}] + E_P[|a_i|^9])^{\frac{2}{4.5}}\}\} \\
B_{2n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : 2n^{-1}\|\Omega_n^{-\frac{1}{2}}\|_o^4 \textstyle\sum_i\{a_{in}^4 E[W^4] + a_{in}^8 E[(W^2-1)^4]\} \leq n^{\frac{4}{9}}\} \\
B_{3n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : \|\Omega_n^{-1}\|_o^{\frac{3}{2}} n^{-1}\textstyle\sum_i\{|a_{in}|^3 + |a_{in}|^6\} < 2\sup_{P\in\mathbf{P}}\{\|\Omega(P)^{-1}\|_o^{\frac{3}{2}} E_P[|a_i|^3 + |a_i|^6]\}\} \\
B_{4n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : n^{-1}\textstyle\sum_i 1\{\min\{|a_{in}|, a_{in}^2\} \geq \delta_0/2\} > \epsilon_0/2 \text{ and } \|\Omega_n\|_o^{\frac{1}{2}} < 2\sup_{P\in\mathbf{P}}\|\Omega(P)\|_o^{\frac{1}{2}}\}
\end{aligned}$$

Then note that whenever $\{Y_i, X_i\}_{i=1}^n \in B_n$: (i) $\{Y_i, X_i\}_{i=1}^n \in B_{0n}$ implies Conditions (I) and (II) in Theorem 3.4 in Skovgaard (1986) are satisfied with $r_n \asymp n^{\frac{5}{18}}$; (ii) $\{Y_i, X_i\}_{i=1}^n \in B_{1n} \cap B_{2n}$ implies together with results (78)-(81) that Condition (IV) in Theorem 3.4 in Skovgaard (1986) is satisfied; (iii) $\{Y_i, X_i\}_{i=1}^n \in B_{3n} \cap B_{4n}$ implies by (84)-(86), together with setting $\epsilon < (\delta_0 \sup_{P\in\mathbf{P}}\{\|\Omega(P)^{-1}\|_o^{\frac{3}{2}} E_P[|a_i|^3 + |a_i|^6]\})/(2\sup_{P\in\mathbf{P}}\|\Omega(P)\|_o^{\frac{1}{2}})$ in equation (88), Assumption 2.2(ii) and (89) that Condition III" of Theorem 3.4 in Skovgaard (1986) also holds. Therefore, the existence of the desired deterministic sequence $b_n = o(n^{-\frac{1}{2}})$ follows from Theorem 3.4 in Skovgaard (1986).

We now verify $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_n) = O(n^{-\frac{\tilde\nu}{18}})$. To this end, let δ satisfy $1 \leq \delta \leq 9$. By result (129) and Assumption 2.3(iii), there exists a $0 < M_1 < \infty$ such that $\sup_{P\in\mathbf{P}} P(\|H_n^{-1}\|_F > M_1) = O(n^{-\frac{\tilde\nu}{2}})$. Moreover, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_iX_i\|_F^\delta - E_P[\|X_iX_i\|_F^\delta]| > \eta) = O(n^{-\frac{\tilde\nu}{2\delta}})$ for any $\eta > 0$ due to $\tilde\nu \geq 18$, and results (15) and (16). Hence, by Assumption 2.3(ii) there exists a $0 < M_2 < \infty$ such that $\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_i \|X_iX_i'\|_F^\delta > M_2) = O(n^{-\frac{\tilde\nu}{2\delta}})$. Combining these results, we then obtain that:

$$\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_{i=1}^n |c'H_n^{-1}X_iX_i'(\hat\beta - \beta_0)|^\delta > \eta) \leq \sup_{P\in\mathbf{P}} P(M_1M_2\|c\|^\delta\|\hat\beta - \beta_0\|^\delta > \eta) + O(n^{-\frac{\tilde\nu}{2\delta}}) = O(n^{-\frac{\tilde\nu}{2\delta}}) , \tag{139}$$

where the final equality follows from (130) and $\delta \geq 1$. Next, note that by (15), (16) and Assumption 2.3(ii) we have $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_i\epsilon_i\|^\delta - E_P[\|X\epsilon\|^\delta]| > \eta) = O(n^{-\frac{\tilde\nu}{\delta}})$ for any $\eta > 0$. Therefore, by Assumption 2.3(ii), there exists a $0 < M_3 < \infty$ such that $\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_i \|X_i\epsilon_i\|^\delta > M_3) = O(n^{-\frac{\tilde\nu}{\delta}})$, and thus we have:

$$\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_{i=1}^n |c'(H_n^{-1} - H_0^{-1})X_i\epsilon_i|^\delta > \eta) = O(n^{-\frac{\tilde\nu}{2}}) + O(n^{-\frac{\tilde\nu}{\delta}}) \tag{140}$$

due to result (129). Moreover, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^n |a_i|^\delta - E_P[|a_i|^\delta]| > \eta) = O(n^{-\frac{\tilde\nu}{\delta}})$ for any $\eta > 0$ by the same arguments and Assumption 2.3(ii). Therefore, combining (139) and (140) we can conclude that:

$$\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^n |a_{in}|^\delta - E_P[|a_i|^\delta]| > \eta) = O(n^{-\frac{\tilde\nu}{2\delta}}) . \tag{141}$$

Hence, result (141), the definition of Ω_n and $\Omega(P)$ and a_{in}, a_i being nonstochastic with respect to L^*, imply:

$$\sup_{P\in\mathbf{P}} P(\|\Omega_n - \Omega(P)\|_F > \eta) = O(n^{-\frac{\tilde\nu}{8}}) \tag{142}$$

for any $\eta > 0$. In addition, Assumptions 2.3(iii)-(iv) and $E[(W^2-1)^2] > 0$ by Assumption 2.2(i)-(ii) imply that $\inf_{P\in\mathbf{P}} \lambda(\Omega(P)) > 0$, where $\lambda(\Omega(P))$ denotes the smallest eigenvalue of $\Omega(P)$. Hence, arguing as in (18)-(19):

$$\sup_{P\in\mathbf{P}} P(\|\Omega_n^{-1} - \Omega(P)^{-1}\|_o > \eta) = O(n^{-\frac{\tilde{\nu}}{8}}) \ , \tag{143}$$

for any $\eta > 0$. Therefore, employing (73)-(74) for B_{0n} and results (141) and (143) allow us to obtain the bounds:

$$\begin{array}{llll} \sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin B_{0n}) = & O(n^{-\frac{\tilde{\nu}}{12}}) & \sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin B_{1n}) & = O(n^{-\frac{\tilde{\nu}}{18}}) \\ \sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin B_{2n}) = & O(n^{-\frac{\tilde{\nu}}{16}}) & \sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin B_{3n}) & = O(n^{-\frac{\tilde{\nu}}{12}}) \end{array} . \tag{144}$$

Moreover, we also note by direct calculation that results (129) and (130) imply that (for M_0 as in (138)):

$$\sup_{P\in\mathbf{P}} P(\sup_{\max\{\|XX'\|_F,\|X\epsilon\|\}\le M_0} |c'H_0(P)^{-1}X\epsilon - c'H_n^{-1}X(Y-X\hat{\beta})| > \eta) = O(n^{-\frac{\tilde{\nu}}{2}}) \ . \tag{145}$$

Hence, since $0 < \delta_0 < 1$, we obtain that on a set with probability $1 - O(n^{-\frac{\tilde{\nu}}{2}})$ (uniformly in $P\in\mathbf{P}$) we have:

$$\frac{1}{n}\sum_{i=1}^n 1\{\min\{|a_{in}|, a_{in}^2\} \ge \frac{\delta_0}{2}\} \ge \frac{1}{n}\sum_{i=1}^n 1\{a_i^2 \ge \delta_0 \text{ and } \max\{\|X_i\epsilon_i\|, \|X_iX_i'\|_F\} \le M_0\} \ . \tag{146}$$

Thus, by (146), Bernstein's inequality and (138), together with (142) we conclude that $\sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin B_{4n}) = O(n^{-\frac{\tilde{\nu}}{8}})$. Result (137) then follows by (144) and $P(\{Y_i,X_i\}_{i=1}^n \notin B_n) \le \sum_{j=0}^4 P(\{Y_i,X_i\}_{i=1}^n \notin B_{jn})$.

Next, we aim to exploit result (137) to establish the existence of sets C_n such that $P(\{Y_i,X_i\}_{i=1}^n \notin C_n) = O(n^{-\frac{\tilde{\nu}}{18}})$ and a deterministic sequence $c_n = o(n^{-\frac{1}{2}})$ such that whenever $\{Y_i,X_i\}_{i=1}^n \in C_n$, then uniformly in $z\in\mathbf{R}$:

$$P^*(T_{s,n}^* \le z) = \Phi(z) + \frac{\phi(z)E[W^3]}{6\sqrt{n}}(2z^2+1) \times \frac{\hat{\kappa}}{\hat{\sigma}^3} + c_n \ . \tag{147}$$

To this end, define $C_n = B_n \cap (\bigcap_{j=0}^2 C_{jn})$ where the sets C_{jn} are given by:

$$\begin{array}{ll} C_{0n} & \equiv \{\{Y_i,X_i\}_{i=1}^n : \hat{\sigma}^2 > \frac{1}{2}\inf_{P\in\mathbf{P}} \sigma^2(P) \text{ and } \|\Omega_n\|_F < \sup_{P\in\mathbf{P}} 2\|\Omega(P)\|_F\} \\ C_{1n} & \equiv \{\{Y_i,X_i\}_{i=1}^n : |E^*[(L_n^*)^2] - 1| \le n^{-\frac{3}{4}}\} \\ C_{2n} & \equiv \{\{Y_i,X_i\}_{i=1}^n : |E^*[(L_n^*)^3] + (7E[W^3]\hat{\kappa})/(2\hat{\sigma}^3\sqrt{n})| \le n^{-\frac{3}{4}}\} \end{array} .$$

Then note that whenever $\{Y_i,X_i\}_{i=1}^n \in C_n$: (i) $\{Y_i,X_i\}_{i=1}^n \in B_n$ and (137) implies condition (3.1) of Theorem 3.2 in Skovgaard (1981) is satisfied; (ii) $\{Y_i,X_i\}_{i=1}^n \in C_{0n}$ and result (100) verifies condition (3.11) of Theorem 3.2 in Skovgaard (1981), while $\{Y_i,X_i\}_{i=1}^n \in C_{0n}$ and result (101) verifies condition (3.12). The Edgeworth expansion in (147) then holds due to Theorem 3.2 and Remark 3.4 in Skovgaard (1981), Lemma A.7 and $\{Y_i,X_i\}_{i=1}^n \in C_{1n}\cap C_{2n}$. Moreover, by (142) and (134), $\sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin C_{0n}) = O(n^{-\frac{\tilde{\nu}}{8}})$, while from (56), (57) and (141), together with (134) we obtain $\sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin C_{1n}) = O(n^{-\frac{\tilde{\nu}}{8}})$ (note in Lemma A.8, $a_{in} = c'H_n^{-1}X_i$, and not $a_{in} = c'H_n^{-1}(Y_i - X_i\hat{\beta})$ as used in (141)). Finally, by direct calculation, we also obtain from (67)-(70) and (141), together with (134) that $\sup_{P\in\mathbf{P}} P(\{Y_i,X_i\}_{i=1}^n \notin C_{2n}) = O(n^{-\frac{\tilde{\nu}}{18}})$, and hence (147) follows.

Finally, note $\hat{\kappa} = n^{-1}\sum_i a_{in}^3$, (134), (139) and (140) verify (136), which implies $\sup_{P\in\mathbf{P}} P(\frac{|\hat{\kappa}|}{\hat{\sigma}^3} > C_0) = O(n^{-\frac{\tilde{\nu}}{8}})$. The Lemma then follows from (147), Lemma C.2 and Lemma 5 in Andrews (2002). ■