# Finite Sample Nonparametric Tests for Linear Regressions*

Olivier Gossner[†] Karl H. Schlag[‡]

March 10, 2010

**Abstract**

We introduce several exact nonparametric tests for finite sample multivariate linear regressions, and compare their powers. This fills an important gap in the literature where the only known nonparametric tests are either asymptotic, or assume one covariate only.

## 1 Introduction

The question of testing parameters of a linear regression without assumptions beyond independence on the structure of the noise terms is a long standing one in Econometrics. Dating back to White (1980), several asymptotic solutions have been proposed. Although a large literature focuses on comparing the finite sample performances of asymptotical tests (see e.g. MacKinnon and White, 1985; Davidson and MacKinnon, 1993), it has already be pointed out that the use of asymptotic bounds for finite samples can be problematic (Greene, 2002, chapter 11). Exact finite sample non-parametric tests require the probability of type I errors to be below the specified

*Karl Schlag gratefully acknowledges financial support from the Spanish Ministerio de Educación y Ciencia, Grant MEC-SEJ2006-09993 and from the Barcelona Graduate School of Economics.

[†]Paris School of Economics, Paris and London School of Economics, London. email: gossner@pse.ens.fr

[‡]Department of Economics and Business, Universitat Pompeu Fabra, Spain, email: karl.schlag@upf.edu

significance level, for a given sample size. Such tests have been provided for one covariate by Dufour and Hallin (1993) when error terms are symmetric, and by Schlag (2008a) without this assumption, but their construction remains an open question for general linear regressions.

This paper introduces several exact finite sample nonparametric tests for general regressions and compares their power. In particular, our tests allow to derive exact confidence intervals for these coefficients. They rely on the knowledge of bounds on the range of values taken by the dependent variables. The impossibility results obtained by Bahadur and Savage (1956) and Dufour (2003) show that without such knowledge, only trivial tests are exact for a given finite sample. In practice, as data is usually based on outcomes measured on a bounded scale, cases in which the range of the endogenous variables is unbounded are the exception rather than the rule.

We present three different types of tests that we refer to as "Non-Standardized", "Bernoulli", and "Standardized". We derive bounds on the probabilities of type II errors that allow to select the most appropriate test given the sample size and the specific values of the covariates. We briefly summarize their construction. Each test relies on a linear combination of the dependent variables (such as in the OLS method) which is an unbiased estimator of the coefficient to be tested. Each element of the linear combination is a rescaling of the corresponding dependent variable. It is useful to think of the estimator as the sum of these independently distributed rescaled variables with unknown distributions.

The test we call "Non-Standardized" relies on Cantelli's inequality (Cantelli, 1910), on its strengthening for not too small deviations due to Bhattacharyya (1987), on a classical inequality of Hoeffding (1963), as well as on the Berry-Esseen inequality (Berry, 1941; Esseen, 1942; Shiganov, 1986) to bound the tail probabilities of the sum of the rescaled variables. This allows for the construction of an exact test, and for bounds on the power of such a test. We then discuss the choice of linear combination, and present several arguments in favor of using the linear combination that corresponds to the OLS estimator.

The "Bernoulli" test combines insights used in the mean tests of Schlag (2006, 2008b) with a bound for the sum of independent Bernoulli variables due to Hoeffding (1956). We first define a randomized test, using a mean preserving transformation that maps each rescaled variable into a binary random variable with identical ranges. In order to determine a critical region for this test, we bound the tails of the distribution of the sum using a result of Hoeffding (1956) showing that the worst case is attained when all Bernoulli variables are identically distributed. From this randomized test, we then define a nonrandomized test by rejecting the null hypothesis if the probability of rejection of the randomized test is above a specified threshold, thus following the same method as Schlag (2006, 2008b). A candidate for the linear combination of variables used for this test is the one that minimizes the largest absolute value of its coefficients, which is the solution of a linear programming problem.

The "Standardized" test relies both on the Berry-Esseen inequality and on a bound on the difference between the standard deviation of the estimator of the coefficient in the regression and an estimate thereof. A test statistic is constructed by dividing the estimator of the coefficient by the estimate of its standard deviation. It is enlightening to compare this test with that of White (1980). When the coefficient is estimated using OLS, and under some specification of the parameters defining our test, the test statistic is asymptotically equivalent to White's test statistic, and our bounds on the probability of type I and type II errors converge to those of White. In particular, the Standardized test performs asymptotically as well as White's test.

We investigate the performance of the Non-Standardized and the Bernoulli test in two canonical numerical examples involving one covariate in addition to the constant. We find that the tests perform well even for small sample sizes (e.g. $n = 40$). The Non-Standardized test does best when concerned with sufficiently small type II error probabilities. It also does best if the sample is suffcieintly large. Remarkably, the Bernoulli test does better in a variety of intermediate cases when type II error is not too small and the distribution of the covariates is not too asymmetric.

The Standardized test, which is not directly comparable with the first two, is expected to perform well in large samples, and when the noise terms of the regression are small compared to the bound on the exogenous variables.

The paper is organized as follows. Section 2 introduces the model. Sections 3, 4, and 5 successively introduce the Non-Standardized, Bernoulli and Standardized test. Section 6 presents numerical examples of applications of the first two. We conclude in Section 7.

# 2 Linear Regression

We consider a linear regression model with fixed regressors, given by

$$Y_i = X_i\beta + \varepsilon_i, \quad i = 1, .., n$$

where $X_i$ is the $i$-th row of a matrix $X \in \mathbb{R}^{n \times m}$, $\beta \in \mathbb{R}^m$ and $(\varepsilon_i)_i$ is a sequence of independent, not necessarily identically distributed, random variables with $E(\varepsilon_i) = 0$. The error terms $(\varepsilon_i)_i$ are unobservable while $Y = (Y_i)_i$ and $X$ are observable. The vector of parameters $\beta$ is unknown to the statistician. We assume uniform bounds on $Y_i$ and take, w.l.o.g., $Y_i \in [0, 1]$.

We derive exact tests at the level of significance $\alpha$ for the one-sided hypotheses $H_0 : \beta_j \leq \bar{\beta}_j$ against $H_1 : \beta_j > \bar{\beta}_j$ where $\bar{\beta}_j \in \mathbb{R}$. *Exact* means that the probability of a type I error of the test is proven to be below the specified significance level $\alpha$ for the regressors given by $X$. In particular, bounds on the probabilities of type I errors do not rely on asymptotic theory. For each test we provide upper bounds on the probability of type II error, independently of the realized value of $Y$.

As shown by Pratt (1961), upper bounds on the maximal expected width of the confidence intervals can be derived from bounds on the probabilities of type II errors. Hence, our tests can be used to construct confidence intervals with guaranteed coverage.

Each test relies on a linear unbiased estimate $\hat{\beta}_j$ of the coefficient $\beta_j$ by considering $\tau_j \in \mathbb{R}^n$ such that $X'\tau_j = e_j$ where $e_{jj} = 1$ and $e_{jk} = 0$ for $k \neq j$ and setting $\hat{\beta}_j = \tau_j' Y$. The bounds on the probabilities of type II errors can be used to select the appropriate $\tau_j$ in each test and to compare the different tests. We let $\|\tau_j\|^2 = \sum_i \tau_{ij}^2$, and $\|\tau_j\|_\infty = \max_i |\tau_{ij}|$.

# 3 Non-Standardized Test

Our first test uses $\hat{\beta}_j = \tau_j' Y$ as test statistic. The test is called "Non-Standardized" as this test statistic is not divided by an estimate of its standard deviation.

In order to construct the test, we first use classical probability inequalities to bound the tail distribution of $\hat{\beta}_j$ in Subsection 3.1. Since some of these inequalities rely on the variance $\sigma_{\hat{\beta}_j}^2$ of $\hat{\beta}_j$, we present bounds on this variance in Subsection 3.2.

We then combine these bounds to construct an exact test in Subsection 3.3, and to bound the probability of type II error of this test in Subsection 3.4. In Subsection 3.5 we present some useful insights for assembling this test. Finally, we discuss the choice of $\tau_j$ in Subsection 3.6.

## 3.1 Tail Bounds

We present four methods for bounding the tail distribution of $\hat{\beta}$, based on Cantelli, Bhattacharyya, Hoeffding and Berry-Esseen's inequalities.

### 3.1.1 Cantelli

Cantelli's inequality (Cantelli, 1910) states that for a random variable $Z$ of variance $\sigma^2$ and for $k > 0$:

$$P\left(Z - EZ \geq k\sigma^2\right) \leq \frac{1}{1+k^2}. \tag{1}$$

**Proposition 1** *Let*

$$\varphi_C(\sigma, t) = \frac{\sigma^2}{\sigma^2 + t^2},$$

1. *For $\bar{t} > 0$ and $\beta_j \leq \bar{\beta}_j$,*

$$P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) \leq \varphi_C(\sigma_{\beta_j}, \bar{t}).$$

2. *For $\bar{t} > 0$ such that $\beta_j > \bar{\beta}_j + \bar{t}$,*

$$P\left(\hat{\beta}_j - \bar{\beta}_j < \bar{t}\right) \leq \varphi_C(\sigma_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}).$$

3. *For $\sigma, t > 0$, $\varphi_C$ is increasing in $\sigma$, decreasing in $t$.*

**Proof.** For $\bar{t} > 0$ and $\beta_j \leq \bar{\beta}_j$, applying Cantelli's inequality to $\hat{\beta}$ shows

$$\begin{aligned} P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) &\leq P\left(\hat{\beta}_j - \beta_j \geq \bar{t}\right) \\ &\leq \frac{\sigma^2_{\beta_j}}{\sigma^2_{\beta_j} + \bar{t}^2} \\ &= \varphi_C(\sigma_{\beta_j}, \bar{t}) \end{aligned}$$

which is point 1. For $\bar{t}$ such that $\beta_j > \bar{\beta}_j + \bar{t}$ we obtain

$$\begin{aligned} P\left(\hat{\beta}_j - \bar{\beta}_j < \bar{t}\right) &= P\left(-\hat{\beta}_j + \beta_j > \beta_j - (\bar{\beta}_j + \bar{t})\right) \\ &\leq \frac{\sigma^2_{\beta_j}}{\sigma^2_{\beta_j} + (\beta_j - \bar{\beta}_j - \bar{t})^2} \\ &= \varphi_C(\sigma_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}) \end{aligned}$$

which is point 2. Point 3 is immediate. ■

### 3.1.2 Bhattacharyya

The inequality due Bhattacharyya (1987) strengthens Cantelli's inequality using the third and fourth moments of the distribution as follows. Consider a random variable $Z$ with $EZ = 0$ and variance $\sigma^2$, and let $\gamma_1 = \frac{EZ^3}{\sigma^3}$ and $\gamma_2 = \frac{EZ^4}{\sigma^4}$. If $k^2 - k\gamma_1 - 1 > 0$ then

$$\Pr\left(Z \geq k\sigma\right) \leq \frac{\gamma_2 - \gamma_1^2 - 1}{\left(\gamma_2 - \gamma_1^2 - 1\right)\left(1 + k^2\right) + \left(k^2 - k\gamma_1 - 1\right)^2}. \tag{2}$$

The condition $k^2 - k\gamma_1 - 1 > 0$ imposes that $k$ has to be large enough, hence (2) only applies for deviations that are not too small.

**Proposition 2** *Let*

$$\varphi_Y(\sigma, t) = \begin{cases} \frac{2\sigma^4}{3\sigma^4 - 2\sigma^2 + t^4} & \text{if } \sigma^2 \leq \frac{t^2\|\tau_j\|_\infty}{\|\tau_j\|_\infty + 2t} \text{ and } \frac{t^2}{\sigma^2} - \frac{t\|\tau_j\|_\infty}{\sigma^2} - 1 > 0 \\ \frac{2\sigma^4}{2\sigma^2(\sigma^2 + t^2) + \left(t^2 - t\|\tau_j\|_\infty - \sigma^2\right)^2} & \text{if } \sigma^2 > \frac{t^2\|\tau_j\|_\infty}{\|\tau_j\|_\infty + 2t} \text{ and } \frac{t^2}{\sigma^2} - \frac{t\|\tau_j\|_\infty}{\sigma^2} - 1 > 0 \\ 1 & \text{if } \frac{t^2}{\sigma^2} - \frac{t\|\tau_j\|_\infty}{\sigma^2} - 1 \leq 0 \end{cases}$$

1. *For $\bar{t} > 0$ and $\beta_j \leq \bar{\beta}_j$,*

$$P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) \leq \varphi_Y(\sigma_{\beta_j}, \bar{t}).$$

2. *For $\bar{t} > 0$ such that $\beta_j > \bar{\beta}_j + \bar{t}$,*

$$P\left(\hat{\beta}_j - \bar{\beta}_j < \bar{t}\right) \leq \varphi_Y(\sigma_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}).$$

3. *For $\sigma, t > 0$ such that $\frac{t^2}{\sigma^2} - \frac{t\|\tau_j\|_\infty}{\sigma^2} - 1$, $\varphi_Y$ is increasing in $\sigma$, decreasing in $t$.*

Before applying inequality (2) to $Z = \hat{\beta}_j - \beta_j$ we bound the corresponding values of $\gamma_1$ and $\gamma_2$.

**Lemma 1**

$$\frac{E\left(\hat{\beta}_j-\beta_j\right)^3}{\sigma^3_{\beta_j}} \leq \frac{\|\tau_j\|_\infty}{\sigma_{\beta_j}} \tag{3}$$

*and*

$$\frac{E\left(\hat{\beta}_j-\beta_j\right)^4}{\sigma^4_{\beta_j}} \leq 3.$$

**Proof.** Using the polynomial expansion, and $E(X_i\beta - Y_i) = 0$ for every $i$, we obtain

$$E\left(\hat{\beta}_j-\beta_j\right)^3 = \sum_i \tau^3_{ij} E\left(X_i\beta - Y_i\right)^3.$$

Since $|X_i\beta - Y_i| \leq 1$, we obtain

$$\gamma_1 = \frac{E\left(\hat{\beta}_j-\beta_j\right)^3}{\sigma^3_{\beta_j}} = \frac{\sum_i \tau^3_{ij} E\left(X_i\beta - Y_i\right)^3}{\sigma^3_{\beta_j}} \leq \frac{\|\tau_j\|_\infty \sum_i \tau^2_{ij} E\left(X_i\beta - Y_i\right)^2}{\sigma^3_{\beta_j}} = \frac{\|\tau_j\|_\infty}{\sigma_{\beta_j}}.$$

Using the polynomial expansion again, we get

$$E\left(\hat{\beta}_j-\beta_j\right)^4 = \sum_i \tau^4_{ij} E\left(X_i\beta - Y_i\right)^4 + 3\sum_{i\neq k} \tau^2_{ij} E\left(X_i\beta - Y_i\right)^2 \tau^2_{kj} E\left(X_k\beta - Y_k\right)^2$$

and

$$\left(\sum_i \tau^2_{ij} E\left(X_i\beta - Y_i\right)^2\right)^2 = \sum_i \tau^4_{ij} E\left(X_i\beta - Y_i\right)^4 + \sum_{i\neq k} \tau^2_{ij} E\left(X_i\beta - Y_i\right)^2 \tau^2_{kj} E\left(X_k\beta - Y_k\right)^2.$$

From this we derive

$$E\left(\hat{\beta}_j-\beta_j\right)^4 = 3\left(\sum_i \tau^2_{ij} E\left(X_i\beta - Y_i\right)^2\right)^2 - 2\sum_i \tau^4_{ij} E\left(X_i\beta - Y_i\right)^4$$

and hence

$$\gamma_2 = \frac{E\left(\hat{\beta}_j-\beta_j\right)^4}{\sigma^4_{\beta_j}} = \frac{3\left(\sum_i \tau^2_{ij} E\left(X_i\beta - Y_i\right)^2\right)^2 - 2\sum_i \tau^4_{ij} E\left(X_i\beta - Y_i\right)^4}{\sigma^4_{\beta_j}} \leq 3\ .$$

■

**Proof of Proposition 2.** For the proof of point 1, we need only to consider the case where $\frac{\bar{t}^2}{\sigma^2} - \frac{\bar{t}\|\tau_j\|_\infty}{\sigma^2} - 1 > 0$, in which we can apply (2) to $\hat{\beta}_j - \beta_j$:

$$\begin{aligned} P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) &\leq P\left(\hat{\beta}_j - \beta_j \geq \bar{t}\right) \\ &\leq \frac{\gamma_2 - \gamma_1^2 - 1}{\left(\gamma_2 - \gamma_1^2 - 1\right)\left(1 + \left(\frac{\bar{t}}{\sigma_{\beta_j}}\right)^2\right) + \left(\left(\frac{\bar{t}}{\sigma_{\beta_j}}\right)^2 - \left(\frac{\bar{t}}{\sigma_{\beta_j}}\right)\gamma_1 - 1\right)^2} \\ &\leq \frac{2 - \gamma_1^2}{\left(2 - \gamma_1^2\right)\left(1 + \frac{\bar{t}^2}{\sigma^2_{\beta_j}}\right) + \left(\frac{\bar{t}^2}{\sigma^2_{\beta_j}} - \frac{\bar{t}}{\sigma_{\beta_j}}\gamma_1 - 1\right)^2}. \end{aligned} \tag{4}$$

We maximize (4), which is concave in $\gamma_1$, over all $\gamma_1 \leq \frac{\|\tau_j\|_\infty}{\sigma_{\beta_j}}$, holding $\sigma_{\beta_j}$ and $\|\tau_j\|_\infty$ fixed, and obtain

$$P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) \leq \begin{cases} \frac{2}{3-2\frac{\bar{t}^2}{\sigma^2_{\beta_j}}+\frac{\bar{t}^4}{\sigma^4_{\beta_j}}} & \text{if } \sigma^2_{\beta_j} \leq \frac{\bar{t}^2\|\tau_j\|_\infty}{\|\tau_j\|_\infty+2\bar{t}} \\ \frac{2-\frac{\|\tau_j\|^2_\infty}{\sigma^2_{\beta_j}}}{\left(2-\frac{\|\tau_j\|^2_\infty}{\sigma^2_{\beta_j}}\right)\left(1+\frac{\bar{t}^2}{\sigma^2_{\beta_j}}\right)+\left(\frac{\bar{t}^2}{\sigma^2_{\beta_j}}-\frac{\bar{t}\|\tau_j\|_\infty}{\sigma^2_{\beta_j}}-1\right)^2} & \text{if } \sigma^2_{\beta_j} \geq \frac{\bar{t}^2\|\tau_j\|_\infty}{\|\tau_j\|_\infty+2\bar{t}} \end{cases}$$
$$= \varphi_Y(\sigma_{\beta_j}, \bar{t})$$

which is point 1. The proof of point 2 is similar, and point 3 comes from the fact that both functionals defining $\varphi_Y$ when $\frac{t^2}{\sigma^2} - \frac{t\|\tau_j\|_\infty}{\sigma^2} - 1$ are increasing in $\sigma$, decreasing in $t$, and coincide when $\sigma^2 = \frac{t^2\|\tau_j\|_\infty}{\|\tau_j\|_\infty+2t}$. ■

### 3.1.3 Hoeffding

We recall an inequality due to Hoeffding (1963, Theorem 2). Let $(Z_i)_{i=1}^n$ be independent random variables with $Z_i \in [a_i, b_i]$ for $i = 1, .., n$, and $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$. For $\bar{t} > 0$,

$$P\left(\bar{Z} - E\bar{Z} \geq \bar{t}\right) \leq \exp\left(-\frac{2n^2\bar{t}^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{5}$$

Relying on Hoeffding's inequality we show:

**Lemma 2** *Let*

$$\varphi_H(t) = \exp\left(-\frac{2t^2}{\|\tau_j\|^2}\right).$$

1. *For $\bar{t} > 0$ and $\beta_j \leq \bar{\beta}_j$,*

$$P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) \leq \varphi_H(\bar{t}).$$

2. *For $\bar{t} > 0$ such that $\beta_j > \bar{\beta}_j + \bar{t}$,*

$$P\left(\hat{\beta}_j - \bar{\beta}_j < \bar{t}\right) \leq \varphi_H(\beta_j - \bar{\beta}_j - \bar{t}).$$

3. *For $t > 0$, $\varphi_H$ is decreasing in $t$.*

**Proof.** We apply Hoeffding's inequality to $(Z_i)_i$ where $Z_i = n\tau_{ij}Y_i$. So $Z_i \in [0, n\tau_{ij}]$ for $\tau_{ij} \geq 0$ and $Z_i \in [n\tau_{ij}, 0]$ for $\tau_{ij} < 0$. For $\beta_j \leq \bar{\beta}_j$:

$$\begin{aligned} P(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}) &\leq P(\tau_j' Y - \beta_j \geq \bar{t}) \\ &\leq \exp\left(-\frac{2n^2\bar{t}^2}{\sum_i (n\tau_{ij})^2}\right) \\ &= \exp\left(-\frac{2\bar{t}^2}{\|\tau_j\|^2}\right) \end{aligned}$$

which is point 1. The proof of point 2 is similar, and point 3 is immediate. ■

### 3.1.4 Berry-Esseen

We recall the Berry-Esseen inequality (Berry, 1941; Esseen, 1942) with the constant as derived by Shiganov (1986). Let $(Z_i)_{1 \leq i \leq N}$ be a family of independent random variables with $Var(Z_i) = \sigma_i^2$. For $\bar{u} \in \mathbb{R}$,

$$\left| P\left(\frac{\sum_{i=1}^N (Z_i - EZ_i)}{\sqrt{\sum_{i=1}^N \sigma_i^2}} \leq \bar{u}\right) - \phi(\bar{u}) \right| \leq \frac{A}{\left(\sum_{i=1}^N \sigma_i^2\right)^{3/2}} \sum_{i=1}^N E|Z_i - EZ_i|^3 \tag{6}$$

where $A = 0.7915$ and $\phi$ is the cumulative density function of the standard normal distribution.

Using the Berry-Esseen inequality, we show

**Proposition 3** *Let*

$$\varphi_{BE}(\sigma, t) = \inf_{w > 0, b_1 \in \mathbb{R}} \frac{1 - \phi\left(\frac{t - b_1}{\sqrt{\sigma^2 + w^2}}\right) + A\frac{2\|\tau_j\|_\infty}{\sqrt{27}w}}{\phi(b_1/w)}.$$

1. *For* $\bar{t} > 0$ *and* $\beta_j \leq \bar{\beta}_j$,
$$P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) \leq \varphi_{BE}(\sigma_{\beta_j}, \bar{t}).$$

2. *For* $\bar{t}$ *such that* $\beta_j > \bar{\beta}_j + \bar{t}$,
$$P\left(\hat{\beta}_j - \bar{\beta}_j < \bar{t}\right) \leq \varphi_{BE}(\sigma_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}).$$

3. *For* $\sigma, t > 0$, $\varphi_{BE}$ *is increasing in* $\sigma$, *decreasing in* $t$.

The idea of the proof of Lemma 3 is to apply (6) to the random variables $Z_i = \tau_{ij} Y_i$. However, a difficulty arises from the fact that the right hand side in (6) is unbounded as a there is no lower bound on $\sum_{i=1}^{n} \sigma_i^2 = \sigma_{\beta_j}^2$. Our solution to this is to add additional random variables with known distribution to the family $(Z_i)_{1 \le i \le N}$ to guarantee such a lower bound. We eliminate this noise in a later step.

**Lemma 3** *Let $w > 0, \bar{u} \in \mathbb{R}$. With $Z \sim \mathcal{N}(0, w^2)$ independent of $(Y_i)_i$, and*

$$R(w) = \frac{\sum_i |\tau_{ij}|^3 E|Y_i - EY_i|^3}{\left(\sum_i \tau_{ij}^2 \sigma_i^2 + w^2\right)^{3/2}},$$

*we have*

$$P\left(\frac{\hat{\beta}_j - \beta_j + Z}{\sqrt{\sigma_{\beta_j}^2 + w^2}} \ge \bar{u}\right) \le 1 - \phi(\bar{u}) + AR(w).$$

**Proof.** We apply (6) to the family of random variables $Z_1, ..., Z_{n+N}$ where $Z_i = \tau_{ij} Y_i$ for $i \le n$ and $Z_i \sim \mathcal{N}\left(0, \frac{w^2}{N}\right)$ for $n+1 \le i \le n+N$. Let $K = E|\delta|^3$ for $\delta \sim \mathcal{N}(0, w^2)$. The right hand side in (6), up to the multiplicative constant $A$, becomes

$$\frac{\sum_{i=1}^{n} |\tau_{ij}|^3 E|Y_i - EY_i|^3 + KN\left(w/\sqrt{N}\right)^3}{\left(\sum_{i=1}^{n} \tau_{ij}^2 \sigma_i^2 + w^2\right)^{3/2}}.$$

As $N \to \infty$ this decreases and converges to $R(w)$, and the claim follows from (6). ■

Next we use Lemma 3 to obtain a bound on the upper tail of $\hat{\beta}_j - \beta_j$.

**Lemma 4**

$$P\left(\hat{\beta}_j - \beta_j \ge \bar{t}\right) \le \frac{1 - \phi\left(\frac{\bar{t} - b_1}{\sqrt{\sigma_{\beta_j}^2 + w^2}}\right) + AR(w)}{\phi(b_1/w)}.$$

**Proof.** We use the fact that $P(W_1 + W_2 \ge \bar{u}) \ge P(W_1 \ge -b_1) P(W_2 \ge \bar{u} + b_1)$ holds for all $b_1$, $\bar{u}$ and independent random variables $W_1$ and $W_2$. In our case, we write:

$$P\left(\hat{\beta}_j - \beta_j + Z \ge \bar{u}\sqrt{\sigma_{\beta_j}^2 + w^2}\right) = P\left(\hat{\beta}_j - \beta_j \ge \bar{u}\sqrt{\sigma_{\beta_j}^2 + w^2} + b_1\right)\phi(b_1/w).$$

Applying this to $\bar{u} = \frac{\bar{t} - b_1}{\sqrt{\sigma_{\beta_j}^2 + w^2}}$ and combining with Lemma 3 yields the result. ■

Our next task is to provide an upper bound on $R(w)$.

**Lemma 5**

$$R(w) \le \frac{2\|\tau_j\|_\infty}{\sqrt{27}w}.$$

**Proof.** Using $E\left|Y_i - EY_i\right|^3 \leq \sigma_i^2$, $\left|\tau_{ij}\right|^3 \leq \left\|\tau_j\right\|_\infty \tau_{ij}^2$, and that for $x \geq 0$,

$$\frac{x}{\left(x+w^2\right)^{3/2}} \leq \frac{2}{\sqrt{27}w},$$

we derive

$$\begin{aligned} R\left(w\right) &= \frac{\sum_i \left|\tau_{ij}\right|^3 E\left|Y_i - \mu_i\right|^3}{\left(\sum_i \tau_{ij}^2 E\left(Y_i - \mu_i\right)^2 + w^2\right)^{3/2}} \\ &\leq \frac{\left\|\tau_j\right\|_\infty \sum_i \left|\tau_{ij}\right|^2 E\left(Y_i - \mu_i\right)^2}{\left(\sum_i \tau_{ij}^2 E\left(Y_i - \mu_i\right)^2 + w^2\right)^{3/2}} \\ &\leq \frac{2\left\|\tau_j\right\|_\infty}{\sqrt{27}w}. \end{aligned} \tag{7}$$

■

**Proof of Proposition 3.** Using Lemmata 4 and 5, we obtain that for $\beta_j \leq \bar{\beta}_j$:

$$\begin{aligned} P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) &\leq P\left(\hat{\beta}_j - \beta_j \geq \bar{t}\right) \\ &\leq \inf_{w>0, b_1 \in \mathbb{R}} \frac{1 - \phi\left(\frac{\bar{t} - b_1}{\sqrt{\sigma_{\beta_j}^2 + w^2}}\right) + A\frac{2\|\tau_j\|_\infty}{\sqrt{27}w}}{\phi\left(b_1/w\right)} \end{aligned}$$

which is point 1. For point 2, we apply point 1 to $Y' = 1_n - Y$. For $\beta_j$ such that $\beta_j > \bar{\beta}_j + \bar{t}$,

$$\begin{aligned} P\left(\hat{\beta}_j - \bar{\beta}_j < \bar{t}\right) &\leq P\left(\tau_j' Y - \bar{\beta}_j \leq \bar{t}\right) \\ &= P\left(\tau_j^T\left(1_n - Y\right) - \left(\tau_j^T 1_n - \beta_j\right) \geq \beta_j - \bar{\beta}_j - \bar{t}\right) \\ &\leq \varphi_{BE}(\sigma_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}). \end{aligned}$$

Point 3 is immediate. ■

## 3.2 Bounds on $\sigma_{\beta_j}$

In order to construct a test and bound its power based on the inequalities presented in Subsections 3.1.1 to 3.1.4, we need both a bound on $\sigma_{\beta_j}$ under the null hypothesis, and a bound on $\sigma_{\beta_j}$ as a function of the unknown parameter $\beta_j$. Therefore we let

$$\bar{\sigma}_{\beta_j}^2 = \bar{\sigma}_{\beta_j}^2(\beta_j) = \max \sum_{i=1}^{n} \tau_{ij}^2 Var(Y_i),$$

where the maximum is taken over $z \in \mathbb{R}^m$ with $z_j = \beta_j$ and all random variables $Y_i$ with values in $[0,1]$ such that $EY_i = X_i z$. It is easy to see that one can restrict attention to Bernoulli random variables, so that

$$\bar{\sigma}^2_{\beta_j} = \max_{z \in \mathbb{R}^m} \left\{ \sum_i \tau_{ij}^2 X_i z \left(1 - X_i z\right) : z_j = \beta_j, \ Xz \in [0,1]^n \right\} .$$

The above expression shows that $\bar{\sigma}^2_{\beta_j}$ can easily be computed numerically. Also, let

$$\bar{\sigma}_{0,\beta_j} = \max_{\beta_j \leq \bar{\beta}_j} \bar{\sigma}_{\beta_j}\left(\beta_j\right) .$$

The following lemma, proven in Appendix A, provides upper bounds on $\bar{\sigma}_{\beta_j}$ and $\bar{\sigma}_{0,\beta_j}$.

**Lemma 6**

$$\bar{\sigma}^2_{\beta_j} \leq \frac{1}{4}\left\|\tau_j\right\|^2 - \frac{1}{n}\left(\beta_j - \frac{1}{2}\sum_i \tau_{ij}\right)^2$$

*and*

$$\bar{\sigma}^2_{0,\beta_j} \leq \frac{\|\tau_j\|^2}{4}.$$

Note that when the first regressor is constant, i.e., when $X_{i1} = 1$ for all $i$, we have $\sum_i \tau_{i1} = 1$ and $\sum_i \tau_{ij} = 0$ for $j > 1$, so that the above bound on $\bar{\sigma}^2_{\beta_j}$ only depends on $\tau_j$ through $\|\tau_j\|$, and is decreasing in $\|\tau_j\|$. Appendix A also presents tighter bounds on $\bar{\sigma}_{\beta_j}$ and $\bar{\sigma}_{0,\beta_j}$.

## 3.3 Test Cutoff

Let

$$\varphi(\sigma,t) = \min\left\{\varphi_C(\sigma,t), \varphi_Y(\sigma,t), \varphi_H(t), \varphi_{BE}(\sigma,t)\right\}.$$

It follows from Propositions 1-3 and from the definition of $\bar{\sigma}_{0,\beta_j}$ that, under $H_0$ and for $t > 0$:

$$P(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}) \leq \varphi(\bar{\sigma}_{0,\beta_j}, \bar{t}).$$

$\varphi$ is continuously decreasing in $\bar{t}$, $\lim_{\bar{t}\to 0} \varphi(\bar{\sigma}_{0,\beta_j}, \bar{t}) = 1$, and $\lim_{\bar{t}\to\infty} \varphi(\bar{\sigma}_{0,\beta_j}, \bar{t}) = 0$. Hence, for $0 < \alpha < 1$, there is a unique solution $\bar{t}_N$ to $\varphi(\bar{\sigma}_{0,\beta_j}, \bar{t}) = \alpha$. We define the Non-Standardized test as the one that rejects the null hypothesis when $\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}_N$. This is an exact test with the probability of a type I error bounded above by $\alpha$.

## 3.4 Type II Error

Given $\beta_j > \bar{\beta}_j + \bar{t}_N$, the following bound on the type II error probability follows from Propositions 1-3, from the definition of $\bar{\sigma}_{\beta_j}$ and of $\bar{t}_N$:

$$P\left(\hat{\beta}_j - \bar{\beta}_j < \bar{t}_N\right) \leq \varphi\left(\bar{\sigma}_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}_N\right). \tag{8}$$

## 3.5 Computation

The value of $\bar{\sigma}_{\beta_j}$ used in the construction of the test can be computed numerically. Alternatively, one can use upper bounds on these in order to define the test. For instance, relying on Lemma 6, one can replace $\bar{\sigma}^2_{0,\beta_j}$ by $\frac{1}{4}\left\|\tau_j\right\|^2$ in the definition of $\varphi$, thus obtaining a larger value for $\bar{t}_N$. By doing so, one obtains an exact test which is less powerful, but more easily computable. With this replacement, the bound on the probability of a type I error derived using Cantelli's inequality is not binding in the equation determining $\bar{t}_N$ if $\alpha < 0.284$. To see this, assume that $\bar{t}_N$ is binding under the bound derived from Cantelli's inequality. Then

$$\varphi_C\left(\frac{\|\tau_j\|}{2}, \bar{t}_N\right) = \frac{\frac{1}{4}\left\|\tau_j\right\|^2}{\frac{1}{4}\left\|\tau_j\right\|^2 + \bar{t}_N^2} = \alpha,$$

hence $\bar{t}_N = \frac{1}{2}\sqrt{\frac{1-\alpha}{\alpha}}\left\|\tau_j\right\|$ and

$$\varphi_H(\bar{t}_N) = \exp\left(-\frac{2\bar{t}_N^2}{\left\|\tau_j\right\|^2}\right) = \exp\left(-\frac{1-\alpha}{2\alpha}\right) \geq \alpha,$$

which implies that $\alpha > 0.284$.

Similarly, using the fact that $\bar{\sigma}^2_{\beta_j} \leq \left\|\tau_j\right\|^2/4$ holds for all $\bar{\sigma}_{\beta_j}$, it follows that the bound on the type II error probability of the Non-Standardized test is not determined by Hoeffding's inequality if the type II error of the Non-Standardized test is above 0.285. Indeed,

$$\varphi_C(\bar{\sigma}_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}_N) \leq \varphi_C\left(\frac{\|\tau_j\|}{2}, \beta_j - \bar{\beta}_j - \bar{t}_N\right) < \varphi_H(\beta_j - \bar{\beta}_j - \bar{t}_N)$$

holds if $\varphi_H(\beta_j - \bar{\beta}_j - \bar{t}_N) > 0.285$.

Finally, note that a necessary condition for Bhattacharyya's inequality to be applied, for type I or for type II error probability, is that the bound derived using

Cantelli's inequality is below 0.5. This is because

$$\frac{z^2}{\sigma_{\beta_j}^2} - \frac{z\,\|\tau_j\|_\infty}{\sigma_{\beta_j}^2} - 1 > 0$$

implies that $z > \sigma_{\beta_j}$, and hence that $\varphi_C(\sigma_{\beta_j}, z) < 0.5$.

## 3.6 Choice of $\tau_j$

In what precedes, $\tau_j$ is an unspecified vector with the property that $X'\tau_j = e_j$. An appropriate choice of $\tau_j$ is one that minimizes the bound on the probability of type II error provided by (8). Examination of (8) shows that one would ideally simultaneously want $\tau_j$ to minimize the rejection threshold $\bar{t}_N$, $\bar{\sigma}_{\beta_j}$, $\|\tau_j\|$, and $\|\tau_j\|_\infty$, in order to minimize $\bar{t}_N$, $\tau_j$ should minimize $\bar{\sigma}_{0,\beta_j}$, $\|\tau_j\|$,and $\|\tau_j\|_\infty$.

These conditions are intuitive. A good unbiased estimator is one with minimal variance, hence minimization of $\bar{\sigma}_{0,\beta_j}$ and $\bar{\sigma}_{\beta_j}$. In the homoskedastic case, the unbiased estimator with minimal variance, i.e., the OLS estimator, is also the one that minimizes $\|\tau_j\|$. Finally, minimizing $\|\tau_j\|_\infty$ can be interpreted as a condition that no single observation should be too influential.

Except in some particular cases of interest, including the examples studied in Section 6, we do not provide explicit formulas for $\bar{\sigma}_{0,\beta_j}$ and $\bar{\sigma}_{\beta_j}$, but these can be computed as the solutions of simple maximization problems.

The best choice of $\tau_j$ can also be computed numerically. We provide some heuristic arguments that are confirmed in the numerical examples presented in Section 6. A natural choice is to choose $\tau_j$ to minimize the rejection threshold $\bar{t}_N$ and to only consider the bounds on the probabilities of type II errors thereafter. As shown in the previous section, when $\alpha < 0.284$ then the bound on probability of type I error derived from Cantelli's inequality is never binding. Berry-Esseen's inequality targets small and moderate deviations while Hoeffding's inequality concerns large deviations. Hence, we expect that the bound based on Hoeffding's inequality is lower then that under Berry-Esseen's inequality even when $\tau_j$ is chosen so as to minimize the latter. Bhattacharyya's inequality, as a variant of Cantelli's, relies heavily on $\|\tau_j\|$ being small when $\bar{\sigma}_{0,\beta_j}$ is bounded by $\|\tau_j\|/2$ as in Lemma 6. Anticipating that either Hoeffding's inequality or Bhattacharyya's inequality with minimal $\|\tau_j\|$ is best at minimizing $\bar{t}_N$, one needs to choose $\|\tau_j\|$ minimal, hence as in the OLS estimator.

The discussion above indicates that the choice of $\tau_j$ corresponding to the OLS estimator, $e_j'(X'X)^{-1}X'$, is a good choice when using the Non-Standardized test. It has the additional advantage that results are easily comparable to those based on tests that assume normally distributed errors. Under this choice of $\tau_j$ the cutoff $\bar{t}$ derived from Hoeffding's inequality (see point 1 of Lemma 2) for determining whether or not to reject the null hypothesis at significance level $\alpha$ is given by

$$\sqrt{\frac{-\ln\alpha}{2}(X'X)^{-1}_{jj}}. \tag{9}$$

# 4 Bernoulli Test

In this section we build on an exact test of Schlag (2006) for testing the mean of a random variable with bounded support based on an independent sample. We extend this test to nonidentically distributed random variables with bounded support and apply it to our linear unbiased estimate by interpreting the estimate as an average.

Consider $\tau_j \in \mathbb{R}^n$ such that $X'\tau_j = e_j$. Let $d \in \mathbb{R}^n$ and $Z_i = n\left(\tau_{ij}Y_i + d_i\right)$. Then $(Z_i)_i$ are independently distributed with $E\bar{Z} = \beta_j + d^s$ where $\bar{Z} = \frac{1}{n}\sum_i Z_i$ and $d^s = \sum_i d_i$. Let $a = n\min\{d_i, \tau_{ij} + d_i : i = 1,..,n\}$ and $b = n\max\{d_i, \tau_{ij} + d_i : i = 1,..,n\}$. Then $Z_i \in [a,b]$ for all $i$.

We first construct a test given $\tau_j$ and $d$, and later discuss the choice of these parameters.

Let $f$ be a random transformation on the domain $[a,b]$ defined by

$$P\left(f\left(z\right)=1\right) = \frac{z-a}{b-a} \text{ and } P\left(f\left(z\right)=0\right) = \frac{b-z}{b-a},$$

and let $W_i = f(Z_i)$. Then $(W_i)_i$ is an independent, not necessarily identically distributed, sequence of random variables with $W_i \in \{0,1\}$ and $E\bar{W} = \left(E\bar{Z} - a\right)/\left(b-a\right)$ where $\bar{W} = \frac{1}{n}\sum_i W_i$.

We successively construct a randomized test, that depends on the realization of $(W_i)_i$ given $(Y_i)_i$, and a non-randomized test, that only depends on $(Y_i)_i$.

## 4.1 A Randomized Test

In this subsection, we construct a randomized test based on one realization of the family $(W_i)_i$. Given $\beta_j$, we let $p_{\beta_j}$ denote the expected proportion of 1's in $(W_i)_i$, it is given by

$$p_{\beta_j} = E\bar{W} = \frac{\beta_j + d^s - a}{b - a}.$$

Let $\bar{p} = p_{\bar{\beta}_j}$. The null hypothesis $H_0 : \beta_j \leq \bar{\beta}_j$ can be restated as

$$H_0 : E\bar{W} \leq \bar{p}.$$

The family $(W_i)_i$ is a family of independent, non identically distributed Bernoulli random variables. Relying on a result of Hoeffding (1956), we show that testing for $H_0$ reduces to testing for the probability of success in a binomial distribution, hence to the case in which $(W_i)_i$ is i.i.d.. For $0 < p < 1$ and $k \in \{0, ..., n\}$, we let

$$B(k, p) = \sum_{i=k}^{n} \binom{n}{i} p^i (1 - p)^{n-i}.$$

**Proposition 4** *For $\alpha' > 0$, let $\bar{k} = \bar{k}(\bar{p}, \alpha')$ be the smallest integer such that $B(\bar{k}, \bar{p}) \leq \alpha'$. Let*

$$r_{\alpha'}(\bar{W}) = \begin{cases} 1 & \text{if} \quad n\bar{W} \geq \bar{k} \\ \frac{\alpha' - B(\bar{k}, \bar{p})}{B(\bar{k}-1, \bar{p}) - B(\bar{k}, \bar{p})} & \text{if} \quad n\bar{W} = \bar{k} - 1 \\ 0 & \text{if} \quad n\bar{W} \leq \bar{k} - 2 \end{cases}$$

*and let*

$$\psi_0(k, p, \alpha') = 1 - r_{\alpha'}\left(\frac{k-1}{n}\right) B(k-1, p) - \left(1 - r_{\alpha'}\left(\frac{k-1}{n}\right)\right) B(k, p).$$

*Assume $\bar{k} > n\bar{p} + 1$.*

1. *If $\beta_j \leq \bar{\beta}_j$ then $Er_{\alpha'}(\bar{W}) \leq \alpha'$.*

2. *If $p_{\beta_j} > \bar{k}/n$ then*

$$Er_{\alpha'}(\bar{W}) \geq 1 - \psi_0(\bar{k}, p_{\beta_j}, \alpha').$$

Consider a randomized test that rejects $H_0$ with probability $r_{\alpha'}(\bar{W})$. Point 1 shows that the type I error probability of this test is bounded by $\alpha'$. A bound on the type II error probability is given by point 2.

Observe that $r_{\alpha'}(\bar{W})$ is the rejection probability under the randomized binomial test for testing $p \leq \bar{p}$ against $p > \bar{p}$ at level $\alpha'$ given $n$ i.i.d. observations, using the most powerful test derived from the Neyman-Pearson lemma (see, e.g., Lehmann and Romano, 2005, Example 3.4.2).

**Proof.** Theorem 5 in Hoeffding (1956) shows that, if $k \geq nE\bar{W}$, then $P(n\bar{W} \geq k) \leq B(k, E\bar{W})$. Similarly, if $k < nE\bar{W}$, then $P(n\bar{W} \geq k) \geq B(k, E\bar{W})$.

Now we prove point 1. With $\lambda = r_{\alpha'}((\bar{k}-1)/n)$, $0 \leq \lambda < 1$ and

$$r_{\alpha'}(\bar{W}) = \lambda \mathbb{1}_{n\bar{W} \geq k-1} + (1-\lambda)\mathbb{1}_{n\bar{W} \geq k}.$$

Assume that $E\bar{W} \leq \bar{p}$. Then $\bar{k} - 1 > nE\bar{W}$. Taking expectations in the previous equation, using Hoeffding's inequality and the fact that $B(k, p)$ is increasing in $p$, we obtain

$$\begin{aligned} Er_{\alpha'}(\bar{W}) &= \lambda P(n\bar{W} \geq k-1) + (1-\lambda)P(n\bar{W} \geq k) \\ &\leq \lambda B(\bar{k}-1, E\bar{W}) + (1-\lambda)B(\bar{k}, E\bar{W}) \\ &\leq \lambda B(\bar{k}-1, \bar{p}) + (1-\lambda)B(\bar{k}, \bar{p}) = \alpha'. \end{aligned}$$

Point 2 follows as $E\bar{W} > \bar{k}/n$ implies:

$$\begin{aligned} Er_{\alpha'}(\bar{W}) &= \lambda P(n\bar{W} \geq k-1) + (1-\lambda)P(n\bar{W} \geq k) \\ &\geq \lambda B(\bar{k}-1, E\bar{W}) + (1-\lambda)B(\bar{k}, E\bar{W}). \end{aligned}$$

■

## 4.2 Non-Randomized Bernoulli Test

The randomized test of Subsection 4.1 relies on one realization of the family $(W_i)_i$, drawn from $(Y_i)_i$ to decide whether or not to reject $H_0$. Given $Y = (Y_i)_i$, $E(r_{\alpha'}(\bar{W})|Y)$ is the probability that this randomized test rejects the null hypothesis at significance level $\alpha'$, conditional on the observation of $(Y_i)_i$. Note that $E(r_{\alpha'}(\bar{W})|Y)$ is a function of $Y$, hence is known to the observer.

As in Schlag (2006, 2008b), we use Markov's inequality (first appearing in Bienaymé, 1853) to construct a nonrandomized test from the randomized test.

**Proposition 5** *For $0 < \theta < 1$, let $\alpha' = \theta\alpha$ and let $\psi(\tau_j, d, \theta) = \psi_0(\bar{k}, p_{\beta_j}, \theta\alpha)$. Assume $\bar{k} > n\bar{p} + 1$.*

1. *If $\beta_j \leq \bar{\beta}_j$ then $P(E(r_{\alpha'}(\bar{W})|Y)) \geq \theta) \leq \alpha$,*

*2. for* $\beta_j > \bar{\beta}_j$*,*

$$P(E(r_{\alpha'}(\bar{W})|Y)) < \theta) \leq \frac{\psi(\tau_j, d, \theta)}{1-\theta}.$$

Define the Bernoulli test as the test that rejects $H_0$ if $E(r_{\alpha'}(\bar{W})|Y)) \geq \theta$. Point 1 shows that this is an exact test with significance level $\alpha$, and point 2 provides a bound on the type II error probability.

**Proof.** For point 1, let $\beta_j \leq \bar{\beta}_j$. From point 1 of Proposition 5, $E(r_{\alpha'}(\bar{W})|Y) \leq \theta\alpha$. Applying Markov's inequality to the non-negative random variable $E(r_{\alpha'}(\bar{W})|Y)$ of expectation $Er_{\alpha'}(\bar{W})$ shows

$$P(E(r_{\alpha'}(\bar{W})|Y)) \geq \theta) \leq \frac{Er_{\alpha'}(\bar{W})}{\theta} \leq \alpha.$$

For point 2, we apply Markov's inequality to $1 - E(r_{\alpha'}(\bar{W})|Y)$:

$$P(E(r_{\alpha'}(\bar{W})|Y)) < \theta) = P(1 - E(r_{\alpha'}(\bar{W})|Y) > 1-\theta) \leq \frac{1 - Er_{\alpha'}(\bar{W})}{1-\theta},$$

which together with point 2 of Proposition 5 implies the result. ■

## 4.3 Choice of the Parameters

The last step is to choose the parameters $\theta$, $\tau_j$ and $d$ used in the construction of the Bernoulli test to minimize the bound on type II error probability presented in Proposition 5 for given $\beta_j$ with $\beta_j > \bar{\beta}_j$.

Recall that the bound on the type II error probability provided by Proposition 5 point 2 is the multiple $1/(1-\theta)$ of the type II error probability of the randomized binomial test with significance level $\theta\alpha$ for testing $p \leq p_{\bar{\beta}_j}$ against $p > p_{\bar{\beta}_j}$, where the type II error probability is evaluated at $p = p_{\beta_j}$. As such, the bound on the type II error probability of the Bernoulli test only depends on $\theta$, $p_{\bar{\beta}_j}$, $p_{\beta_j}$, $\alpha$ and $n$, where $p_{\bar{\beta}_j} = (\bar{\beta}_j + d^s - a)/(b-a)$ and $p_{\beta_j} - p_{\bar{\beta}_j} = (\beta_j - \bar{\beta}_j)/(b-a)$. While $p_{\bar{\beta}_j}$ and $p_{\beta_j}$ are invariant to adding a constant $\varepsilon$ to each $d_i$, this translation increases $a$ by $n\varepsilon$, so we can assume w.l.o.g. that $a = 0$. It follows that $b \geq n\left\|\tau_j\right\|_\infty$. In fact, for given $\tau_j$ and $b_0$ with $b_0 \geq n\left\|\tau_j\right\|_\infty$ one can find $d$ such that $b = b_0$ where $-\tau_{ij} \leq d_i \leq b_0/n$ if $\tau_{ij} < 0$ and $0 \leq d_i \leq b_0/n - \tau_{ij}$ for $\tau_{ij} \geq 0$, and where $d$ is unique if and only if $|\tau_{ij}| = \left\|\tau_j\right\|_\infty = b_0/n$ for all $i$. It follows that

$$\frac{\bar{\beta}_j + \sum_{i:\tau_{ij}<0}|\tau_{ij}|}{b} \leq p_{\bar{\beta}_j} \leq 1 + \frac{\bar{\beta}_j - \sum_{i:\tau_{ij}\geq 0}\tau_{ij}}{b} =: p^h$$

where any value of $p_{\bar{\beta}_j}$ within this range can be attained for appropriate choice of $d$.

We do not provide a formal analysis of how to choose $d$ and $\tau_j$, instead only discuss some of the tradeoffs involved. It is natural to choose $d$ such that $b = n\,\|\tau_j\|_\infty$ as this means that there is no excessive rescaling of the random variables $\tau_{ij}Y_i$. Lowering $b$ increases the distance $p_{\beta_j} - p_{\bar{\beta}_j}$ between the null hypothesis and the value of $p_{\beta_j}$ at which the type II error probability is evaluated. If $b$ can be lowered while leaving $d^s$, and hence $p_{\bar{\beta}_j}$, unchanged, then this will decrease the type II error probability. However, it may not be possible to lower $b$ without lowering $p_{\bar{\beta}_j}$ when $p_{\bar{\beta}_j} = p^h$, which is the case in our numerical examples.

Note that $b$ is bounded below by $n\,\|\tau_j^*\|_\infty$ where $\tau_j^*$ solves $\min_{\tau_j \in \mathbb{R}^n}\{\|\tau_j\|_\infty : X'\tau_j = e_j\}$. $\tau_j^*$ is obtained as the solution of a linear programming problem, hence is easily computable.[1] For the special case where $X_{i1} = 1$ for all $i$ and $m = j = 2$ we have a closed form solution for $\tau_2^*$. Assume that $n$ is even (the case of $n$ odd is similar) and that $X_{i2}$ is increasing in $i$. Let

$$T = \frac{1}{\sum_{i=n/2+1}^{n} X_{i2} - \sum_{i=1}^{n/2} X_{i2}}. \tag{10}$$

Then $\tau_{i2}^* = T$ for $i > n/2$ and $\tau_{i2}^* = -T$ for $i \leq n/2$ with $\|\tau_2^*\|_\infty = T$.

# 5 Standardized Test

In this section we derive a test that relies on an estimate $s_{\beta_j}^2$ of the variance $\sigma_{\beta_j}^2$ of $\hat{\beta}_j$. The construction of the test is similar to how we proceed in Subsection 3.1.4, with the only major difference that instead of relying on a uniform bound on $\sigma_{\beta_j}^2$ to derive bounds on the probabilities of type I and type II errors, we rely on $s_{\beta_j}^2$.

In order to estimate $\sigma_{\beta_j}^2 = \sum_i \tau_{ij}^2 E(Y_i - X_i\beta)^2$, we rely on an estimator of $\beta$. Thus, we consider any $\tau = (\tau_1, .., \tau_m)$ where for every $k$, $X'\tau_k = e_k$. For such $\tau$, $\hat{\beta} = \tau' Y$ is an unbiased estimator of $\beta$, $E\hat{\beta} = \beta$. Following White (1980), we estimate $\sigma_{\beta_j}^2$ by $s_{\beta_j}^2 = \sum_i \tau_{ij}^2 E(Y_i - X_i\hat{\beta})^2$. We control for the quality of this estimate, using the following lemma proven in the appendix.

[1] $\|\tau_j^*\|_\infty = \min_{\tau_j \in \mathbb{R}^n}\{\|\tau_j\|_\infty : X'\tau_j = e_j\}$ if and only if $\|\tau_j^*\|_\infty = \min_{\tau_j \in \mathbb{R}^n, q \geq 0}\{q : \tau_{ij} \leq q, \tau_{ij} \geq -q, X'\tau_j = e_j\}$.

Let $\|X_i\|_1 = \sum_{k=1}^m |X_{ik}|$, and

$$c_0 = \max_{Y\in[0,1]^n} \sum_i \tau_{ij}^2 \left(|X_i\tau' Y| + 2\right) \|X_i\|_1.$$

**Lemma 7**

$$\begin{aligned} P\left(s_{\beta_j}^2 + a^2 \le \sigma_{\beta_j}^2\right) &\le \min_{\lambda\in[0,1]} \left\{ 2\exp\left(-2\frac{\lambda^2 a^4}{c_0^2\|\tau_k\|^2}\right) + \exp\left(-2\frac{(1-\lambda)^2 a^4}{\sum_i \tau_{ij}^4}\right)\right\} \\ &= : c_3(a). \end{aligned}$$

We choose a test statistic that depends on parameters $a_1$, $w$, $b_1$ with $a_1, w > 0$ and is given by:

$$t_S = \frac{\hat{\beta}_j - \bar{\beta}_j - b_1}{\sqrt{s_{\beta_j}^2 + a_1^2 + w^2}},$$

and define the threshold value $\bar{t}_S$ by

$$\bar{t}_S = \phi^{-1}\left(1 - \left(\alpha - c_3\left(a_1\right)\right)\phi(b_1/w) + AR(w)\right)$$

with the convention that $\bar{t}_S = +\infty$ if $1 - \left(\alpha - c_3\left(a_1\right)\right)\phi(b_1/w) + AR(w) \ge 1$.

Define the Standardized test as the test that rejects $H_0$ when $t_S \ge \bar{t}_S$. The next proposition shows that this is an exact test at the level $\alpha$, and gives a bound on the type II error probability.

**Proposition 6** *1. If $\beta_j \le \bar{\beta}_j$ and $b_1 \le \hat{\beta}_j - \bar{\beta}_j$ then*

$$P(t_S \ge \bar{t}_S) \le \alpha.$$

*2. For $a_2 > 0$ and $b_2$, let*

$$\bar{u}_S = \bar{t}_S \frac{\sqrt{\sigma_{\beta_j}^2 + w^2 + a_1^2 + a_2^2}}{\sqrt{\sigma_{\beta_j}^2 + w^2}} + \frac{b_1 + b_2 + \bar{\beta}_j - \beta_j}{\sqrt{\sigma_{\beta_j}^2 + w^2}}.$$

*If $\beta_j > \bar{\beta}_j + \bar{t}_S$ then*

$$P(t_S < \bar{t}_S) \le \frac{\phi(\bar{u}_S) + AR(w)}{\phi(b_2/w)} + c_3\left(a_2\right). \tag{11}$$

**Proof.** Using Lemma 7 we obtain:

$$
\begin{aligned}
P&\left(\frac{\hat{\beta}_j-\beta_j-b_1}{\sqrt{s_{\beta_j}^2+w^2+a_1^2}}\geq \bar{t}_S\right) \\
&\leq P\left(\frac{\hat{\beta}_j-\beta_j-b_1}{\sqrt{\sigma_{\beta_j}^2+w^2}}\geq \bar{t}_S, s_{\beta_j}^2+a_1^2\geq\sigma_{\beta_j}^2\right)+P\left(s_{\beta_j}^2+a_1^2<\sigma_{\beta_j}^2\right) \\
&\leq P\left(\frac{\hat{\beta}_j-\beta_j-b_1}{\sqrt{\sigma_{\beta_j}^2+w^2}}\geq \bar{t}_S\right)+c_3\left(a_1\right). \qquad (12)
\end{aligned}
$$

Let $Z\sim\mathcal{N}(0,w^2)$ with $Z$ independent of $Y$. From Lemma 3,

$$
\begin{aligned}
P(Z>-b_1)P\left(\frac{\hat{\beta}_j-\beta_j-b_1}{\sqrt{\sigma_{\beta_j}^2+w^2}}\geq t\right) &\leq P\left(\frac{\hat{\beta}_j-\beta_j+Z}{\sqrt{\sigma_{\beta_j}^2+w^2}}\geq t\right) \\
&\leq 1-\phi\left(t\right)+AR\left(w\right).
\end{aligned}
$$

Hence

$$
P\left(\frac{\hat{\beta}_j-\beta_j-b_1}{\sqrt{\sigma_{\beta_j}^2+w^2}}\geq t\right)\leq\frac{1-\phi(t)+AR(w)}{\phi(b_1/w)}. \qquad (13)
$$

Combining equations (12) and (13), we obtain that for $\beta_j\leq\bar{\beta}_j$:

$$
\begin{aligned}
P\left(\frac{\hat{\beta}_j-\bar{\beta}_j-b_1}{\sqrt{s_{\beta_j}^2+a_1^2+w^2}}>\bar{t}_S\right) &\leq P\left(\frac{\hat{\beta}_j-\beta_j-b_1}{\sqrt{s_{\beta_j}^2+a_1^2+w^2}}\geq\bar{t}_S\right) \\
&\leq\frac{1-\phi(\bar{t}_S)+AR(w)}{\phi(b_1/w)}+c_3\left(a_1\right) \\
&\leq\alpha.
\end{aligned}
$$

which is point 1 of the proposition. For point 2 we first derive

$$
\begin{aligned}
&P\left(\frac{\hat\beta_j-\bar\beta_j-b_1}{\sqrt{s^2_{\beta_j}+a_1^2+w^2}}\le \bar t_S\right)\\
&\quad = P\left(\frac{\hat\beta_j-\beta_j+b_2}{\sqrt{\sigma^2_{\beta_j}+w^2}}\le \bar t_S\frac{\sqrt{s^2_{\beta_j}+a_1^2+w^2}}{\sqrt{\sigma^2_{\beta_j}+w^2}}+\frac{b_1+b_2+\bar\beta_j-\beta_j}{\sqrt{\sigma^2_{\beta_j}+w^2}}\right)\\
&\quad \le P\left(\frac{\hat\beta_j-\beta_j+b_2}{\sqrt{\sigma^2_{\beta_j}+w^2}}\le \bar t_S\frac{\sqrt{\sigma^2_{\beta_j}+w^2+a_1^2+a_2^2}}{\sqrt{\sigma^2_{\beta_j}+w^2}}+\frac{b_1+b_2+\bar\beta_j-\beta_j}{\sqrt{\sigma^2_{\beta_j}+w^2}}, s^2_{\beta_j}\le\sigma^2_{\beta_j}+a_2^2\right)\\
&\quad + P(s^2_{\beta_j}>\sigma^2_{\beta_j}+a_2^2)\\
&\quad \le P\left(\frac{\hat\beta_j-\beta_j+b_2}{\sqrt{\sigma^2_{\beta_j}+w^2}}\le \bar u_S\right)+c_3\left(a_2\right). \qquad (14)
\end{aligned}
$$

We use Lemma 3 and the de-randomization technique of Lemma 4 again. Letting $Z\sim\mathcal{N}(0,w^2)$,

$$
\begin{aligned}
P(Z<b_2)P\left(\frac{\hat\beta_j-\beta_j+b_2}{\sqrt{\sigma^2_{\beta_j}+w^2}}\le \bar u_S\right) &\le P\left(\frac{\hat\beta_j-\beta_j+Z}{\sqrt{\sigma^2_{\beta_j}+w^2}}<\bar u_S\right)\\
&\le \phi(\bar u_S)+AR(w).
\end{aligned}
$$

Hence

$$
P\left(\frac{\hat\beta_j-\beta_j+b_2}{\sqrt{\sigma^2_{\beta_j}+w^2}}\le \bar u_S\right)\le\frac{\phi(\bar u_S)+AR(w)}{\phi(b_2/w)}. \qquad (15)
$$

Combining (14) and (15) we finally obtain:

$$
P\left(\frac{\hat\beta_j-\bar\beta_j-b_1}{\sqrt{s^2_{\beta_j}+a_1^2+w^2}}\le \bar t_S\right)\le\frac{\phi(\bar u)+AR(w)}{\phi(b_2/w)}+c_3\left(a_2\right).
$$

■

The different parameters $a_1$, $a_2$ $b_1$, $b_2$, $w$ and $\tau$ used in the construction of the Standardized test can be chosen in order to minimize the bound (11) on the probability of type II error given $\alpha$.

## 5.1 Asymptotics

The aim of this subsection is to show that, for a particular choice of parameters $a_1$, $a_2$, $w$ and $\tau$, the test statistic, the rejection zone of the Standardized test are asymptotically equivalent to the widely used asymptotic test of White (1980). Furthermore, the bound on the probability of type II errors is asymptotically no worse than using White's test. To prove this last point, we show that for a particular choice of $a_2$, $b_2$, the bound on type II error in point 2 of Proposition 6 is asymptotically equivalent to that of White.

We assume that all regressors are bounded, w.l.o.g., $|X_{i,j}| \leq 1$, and that $det(\frac{X'X}{n})^{-1} > \delta$ for some $\delta > 0$ and $n$ large enough, so that there exists $K$ such that $\| \left(\frac{X'X}{n}\right)^{-1} \|_\infty \leq K$. Since we are in the fixed regressor case, these assumptions are implied by Assumption 2 of White (1980).

We also assume (Assumption 3 b in White (1980)) that the average covariance matrix $\bar{V} = \frac{1}{n}\sum_i E(Y_i - X_i\beta)^2 X_i' X_i$ is such that $det(\bar{V}) > \delta$ for $n$ large enough.

Let $\hat{\beta}_j$ be the OLS estimator of $\beta_j$, hence let $\tau_j = e_j'(X'X)^{-1}X'$. Under the assumptions above, $\|\tau_j\|_\infty \leq \frac{mK}{n}$. We choose the parameters $a_1 = a_1(n)$, $b_1 = b_1(n)$, and $w = w(n)$ such that $b_1 = n^{-0.6}$, $w = a_1 = n^{-2/3}$.

Recall that White's test statistic is $t_W = \frac{\hat{\beta}_j - \bar{\beta}_j}{s_{\beta_j}}$, while our test statistic is

$$t_S = \frac{\hat{\beta}_j - \bar{\beta}_j - b_1}{\sqrt{s^2_{\beta_j} + a_1^2 + w^2}}.$$

Point 1 of Theorem 1 below shows that the two test statistics $t_W$ and $t_S$ are asymptotically equivalent.

Point 2 of Theorem 1 establishes the convergence of the rejection threshold $\bar{t}_S$ of the Standardized test to $\phi^{-1}(1-\alpha)$, the rejection threshold for White's test.

Finally, fix $C_W > 0$ and consider a sequence $(Y_n)_n$, hence implicitly also a sequence of $\beta_j$ and $\sigma_{\beta_j}$, such that along this sequence the probability of type II error computed from White's asymptotic normal approximation equals $C_W$:

$$\phi\left(\phi^{-1}(1-\alpha) + \frac{\bar{\beta}_j - \beta_j}{\sigma_{\beta_j}}\right) = C_W.$$

Set $a_2 = a_1$, $b_2 = b_1$. Along this sequence of underlying parameters, point 2 of Proposition 6 shows that the type II error probability of the Standardized test is bounded above by:

$$C_S = \frac{\phi(\bar{u}_S) + AR(w)}{\phi(b_2/w)} + c_3(a_1)$$

where

$$\bar{u}_S = \bar{t}_S \frac{\sqrt{\sigma^2 + w^2 + a_1^2 + a_2^2}}{\sqrt{\sigma_{\beta_j}^2 + w^2}} + \frac{b_1 + b_2 + \bar{\beta}_j - \beta_j}{\sqrt{\sigma_{\beta_j}^2 + w^2}}$$

Point 3 of Theorem 1 shows the convergence of $C_S$ to $C_W$, so that the two formulas asymptotically give the same power.

**Theorem 1** *When $n \to \infty$,*

1. *for every $\beta_j$*

$$\frac{t_S}{t_W} \to 1 \quad a.s. \ ,$$

2. 

$$\bar{t}_S \to \phi^{-1}(1-\alpha) \ ,$$

3. *for $C_W > 0$*

$$C_S \to C_W \ .$$

# 6 Numerical Comparison

In two numerical examples we compare the performance of the Non-Standardized and the Bernoulli test as well as the different methods used to bound the probability of type I and type II error within the Non-Standardized test. Both examples involve one covariate, plus the constant. The Standardized test is not included, as, unlike the others, its bound on the probability of type II error depends on $\sigma_{\beta_j}$, and hence it does not offer direct comparison with the other tests. Comparison with the test introduced by Schlag (2008b) is not included either, as this test isn't defined beyond a single covariate.[2]

In the extreme example, the covariate only takes two different values and our tests reduce to finding significant difference between two means. In the normal example, the covariate is distributed according to the quantiles of the normal distribution.

[2]We still point out to the reader interested in working with one covariate that the test introduced in Schlag (2008b) performs better than the tests included in the table for small samples.

## 6.1 The Extreme Example

In the extreme example, the first covariate is constant ($X_{i1} = 1$ for every $i$), while the second covariate takes only the values $-1$ and 1: $X_{i2} = 1$ for $i \leq h$ and $X_{i2} = -1$ for $i > h$ for some $1 \leq h \leq n/2$. The value of $h$ characterizes the balancedness of the sample, the sample is perfectly balanced for $h = n/2$, and gets more and more unbalanced as $h$ gets closer to 1. The bound on the outcomes of $Y_i$, given by $Y_i \in [0, 1]$, constrains the values of $\beta_2$ to belong to $[-1/2, 1/2]$.

We wish to test $H_0 : \beta_2 \leq 0$ against $H_1 : \beta_2 > 0$. Since $H_0$ can be rewritten $H_0 : \frac{1}{h}\sum_{i=1}^{h} EY_i \leq \frac{1}{n-h}\sum_{i=h+1}^{n} EY_i$, our problem is equivalent to testing the difference of means of two populations.

An easy computation shows that, subject to $X'\tau_2 = e_2$, the minimizer of $\|\tau_2\|$ is given by

$$\tau_{i2} = \frac{1}{2h} \text{ if } i \leq h, \quad \tau_{i2} = -\frac{1}{2(n-h)} \text{ if } i > h.$$

$\|\tau_2\|_\infty$ has a continuum of minimizers, including the above choice of $\tau_2$ and the solution given in Subsection 4.3. The corresponding norm values are

$$\|\tau_2\|^2 = \frac{n}{4h(n-h)} \text{ and } \|\tau_2\|_\infty = \frac{1}{2h}.$$

For the Bernoulli test it turns out best in this example to use the minimizer of $\|\tau_2\|$ and to choose $\bar{p} = p^h = 1 - h/n$.

Computation of $\bar{\sigma}_{\beta_j}$ shows, for $\beta_2 \leq n/(4(n-h))$, which is the case in the numerics we consider, that

$$\begin{aligned}\bar{\sigma}_{\beta_j}^2 &= \beta_2 \sum_i \tau_{i2}^2 X_{i2}(1 - \beta_2 X_{i2}) + \frac{(\|\tau_2\|^2 - 2\beta_2 \sum_i \tau_{i2}^2 X_{i2})^2}{4\|\tau_2\|^2} \\ &= \frac{1}{16}\frac{n}{h(n-h)} - \frac{\beta_2^2}{n} = \frac{1}{4}\|\tau_2\|^2 - \frac{\beta_2^2}{n},\end{aligned}$$

Hence, the same $\tau_2$ minimizes $\|\tau_2\|$, $\bar{\sigma}_{\beta_2}$, $\bar{\sigma}_{0,\beta_2}$ and $\|\tau_2\|_\infty$, and hence minimizes the bound on the probability of type II error of the Non-Standardized test as given by (8).

Let $\bar{t}_C$ be the value of $\bar{t}$ derived using Cantelli's inequality, so $\varphi_C(\bar{\sigma}_{0,\beta_j}, \bar{t}) = \alpha$. Similarly, let $\bar{t}_Y$, $\bar{t}_H$ and $\bar{t}_{BE}$ be smallest values of $\bar{t}$ such that the bounds derived using Bhattacharyya's, Hoeffding's and Berry-Esseen's inequality are less or equal to $\alpha$. Following Section 3.3, the Non-Standardized test rejects the null hypothesis when $\hat{\beta}_j \geq \bar{\beta}_j + \bar{t}_N$ where $\bar{t}_N = \min\{\bar{t}_C, \bar{t}_Y, \bar{t}_H, \bar{t}_{BE}\}$.

As stated in Section 3.5, the bound based on Hoeffding's inequality is superior to that based on Cantelli's inequality, i.e., $\bar{t}_H < \bar{t}_C$, when $\alpha < 0.285$. The bound based on Bhattacharyya's inequality can be superior to that based on Hoeffding's inequality. This is the case, for instance, when $\alpha = 0.1$, $n \geq 45$ and $h = n/2$. However, for $\alpha = 0.05$, as assumed in the following tables, we do not encounter such a case. For $\alpha = 0.05$ and $\alpha = 0.01$, and $n \leq 2 \cdot 10^6$, we find that the Berry-Esseen inequality gives higher rejection thresholds than Hoeffding's inequality, i.e., $t_H < t_{BE}$. In our tables below, as $\alpha = 0.05$, $\bar{t}_N = \bar{t}_H$, the cutoff of the Non-Standardized test is determined by Hoeffding's inequality.

In the range of values we use, we find that

$$\bar{\sigma}^2_{\beta_j} > \frac{\left(\beta_j - \bar{\beta}_j - \bar{t}_N\right)^2 \|\tau_j\|_\infty}{\|\tau_j\|_\infty + 2\left(\beta_j - \bar{\beta}_j - \bar{t}_N\right)}$$

which means that, when deriving the upper bound on type II error probability, only the second part of the definition of $\varphi_Y$ in Proposition 2 applies.

Tables 1 and 2 summarize our numerical results. Each row refers to a different specification of the data as identified by the sample size $n$ (first column) and the value of $h$ (second column). The cutoff $\bar{t}_N$ (equal to $\bar{t}_H$) used in the Non-Standardized test is shown in the third column. The fourth column expresses $\bar{k}_b$, the cutoff in Bernoulli test, in the form $\bar{k}_B/n - \bar{p}$ which is a natural measure of how much evidence is needed beyond what is expected in order to reject the null hypothesis.

The fifth column specifies the value of $\beta_2$ guaranteed to provide type II error probabilities below the values shown in the remaining columns. In the first table the value of $\beta_2$ is chosen so that the best bound on the probability of type II error among our tests equals 0.5. In the second table we compare the tests in terms of their ability to guarantee type II error to be below 0.2, and twice, for $n = 500$ and $h = 100, 250$, also in terms of type II error below 0.05.

The last 5 columns show the bounds on the probabilities of type II errors obtained using each of the respective inequalities of Cantelli (C), Bhattacharyya (Y), Hoeffding (H) and Berry-Esseen (BE) in the Non-Standardized test, and in the Bernoulli test (B). The bound on the probability of type II error of the Non-Standardized test given the value of $\beta_2$ in the fifth column is the minimal value of the entries in columns "C", "Y", "H", and "BE".

| $n$ | $h$ | $\bar{t}_N$ | $\bar{k}_B/n - \bar{p}$ | $\beta_2$ | C | Y | H | BE | B |
|---|---|---|---|---|---|---|---|---|---|
| 40 | 20 | 0.194 | 0.225 | 0.198 | 0.997 | 1 | 0.999 | 1 | **0.5** |
| 40 | 10 | 0.225 | 0.175 | 0.301 | **0.5** | 1 | 0.695 | 0.803 | 0.538 |
| 100 | 50 | 0.122 | 0.11 | 0.127 | 0.992 | 1 | 0.996 | 1 | **0.5** |
| 100 | 25 | 0.141 | 0.11 | 0.196 | **0.5** | 1 | 0.642 | 0.653 | 0.586 |
| 500 | 250 | 0.0547 | 0.052 | 0.057 | 0.989 | 1 | 0.995 | 0.762 | **0.5** |
| 500 | 200 | 0.0559 | 0.052 | 0.0713 | 0.682 | 1 | 0.796 | 0.552 | **0.5** |
| 500 | 150 | 0.0597 | 0.048 | 0.0814 | 0.552 | 1 | 0.673 | **0.5** | 0.566 |
| 500 | 100 | 0.0684 | 0.042 | 0.096 | **0.5** | 1 | 0.613 | 0.507 | 0.893 |
| 5000 | 2500 | 0.0173 | 0.0154 | $0.0181 \cdot 10^{-2}$ | 0.989 | 1 | 0.994 | 0.637 | **0.5** |
| $2 \cdot 10^6$ | $10^6$ | $8.66 \cdot 10^{-4}$ | $7.7 \cdot 10^{-4}$ | $9 \cdot 10^{-4}$ | 0.989 | 1 | 0.994 | 0.502 | **0.5** |

Table 1: Comparison of methods in the extreme example for maximal type II error probabilities of 0.5.

We make some observations given these two tables. Overall, each test and each bound has its own region where it adds value to making inference about $\beta_2$.

1. Our tests perform well in small samples. The bound on the probability of type II error of the Bernoulli test is below 0.5 when $n = 40$ and $h = 20$ for $\beta_2 \geq 0.198$, the bound on the probability of type II error of the Non-Standardized test is below 0.5 when $n = 100$ and $h = 25$ for $\beta_2 \geq 0.196$.

2. The Bernoulli test performs best when the sample is balanced, so when $h = n/2$. This finding is intuitive. The Bernoulli test relies on rescaling variables $n\tau_{i2}Y_i$ into an interval of width $n \|\tau_2\|_\infty$. If $|\tau_{i2}|$ is small then the information contained in $Y_i$ is diluted. When $h = n/2$ then $|\tau_{2i}|$ is independent of $i$ so this dilution does not occur. Once the data has been transformed into 0's or 1's, it is as if we are comparing the number of successes (occurrences of $W_i = 1$) between the two samples $\{W_i, i \leq n/2\}$ to $\{W_i, i > n/2\}$. The Bernoulli test does this very effectively as it relies on the binomial test, its only downside is that the level of the binomial test is chosen to be $\theta\alpha$ to then be able to derive a test with level $\alpha$ that is nonrandomized. However, despite this adjustment, $n = 2 \cdot 10^6$ is not large enough for it to be outperformed by the Non-Standardized test.

| $n$ | $h$ | $\bar{t}_N$ | $\bar{k}_B/n - 1/2$ | $\beta_2$ | C | Y | H | BE | B |
|---|---|---|---|---|---|---|---|---|---|
| 40 | 20 | 0.194 | 0.2 | 0.243 | 0.659 | 1 | 0.821 | 0.763 | **0.2** |
| 100 | 50 | 0.122 | 0.12 | 0.159 | 0.631 | 1 | 0.769 | 0.627 | **0.2** |
| 100 | 25 | 0.141 | 0.1 | 0.233 | 0.247 | **0.2** | 0.28 | 0.418 | 0.3 |
| 500 | 250 | 0.0547 | 0.046 | 0.072 | 0.621 | 1 | 0.741 | 0.276 | **0.2** |
| 500 | 250 | 0.0547 | 0.044 | 0.0872 | 0.316 | 0.27 | 0.349 | 0.265 | **0.05** |
| 500 | 200 | 0.0559 | 0.046 | 0.0869 | 0.344 | 0.314 | 0.396 | 0.311 | **0.2** |
| 500 | 150 | 0.0597 | 0.046 | 0.0998 | 0.264 | **0.2** | 0.26 | 0.266 | 0.284 |
| 500 | 100 | 0.0684 | 0.038 | 0.115 | 0.261 | **0.2** | 0.254 | 0.3 | 0.443 |
| 500 | 100 | 0.0684 | 0.038 | 0.137 | 0.137 | 0.0527 | **0.05** | 0.169 | 0.199 |
| 5000 | 2500 | 0.0173 | 0.0145 | 0.0228 | 0.621 | 1 | 0.737 | 0.371 | **0.2** |
| $2 \cdot 10^6$ | $10^6$ | $8.7 \cdot 10^{-4}$ | $8.2 \cdot 10^{-4}$ | $1.17 \cdot 10^{-3}$ | 0.621 | 1 | 0.737 | 0.255 | **0.2** |

Table 2: Comparison of methods in the extreme example for maximal type II error probabilities of 0.2 and 0.05.

3. The Non-Standardized test outperforms the Bernoulli test when the sample is unbalanced, e.g. when $n = 40$ and $h = 10$. In this case, as $|\tau_{i2}|$ is very different depending on whether $i \leq h$ or $i > h$, too much information on $Y_i$ is lost in the Bernoulli test due to rescaling of $W_i$ for $i > h$. For small samples, the probability of type II error of the Non-Standardized test is guaranteed to be below 0.5 by using Cantelli's inequality and to be below 0.2 by using Bhattacharyya's inequality. Hoeffding's inequality is more valuable for bounding the probability of type II error when concerned with large deviations, such as when ensuring the probability of type II error below 0.05 when $n = 500$ and $h = 100$. The Berry-Esseen inequality is valuable for guaranteeing the probability of type II error below 0.5 in larger samples when the sample is not too balanced nor too unbalanced, e.g. when $n = 500$ and $h = 150$.

## 6.2 The Normal Example

In the extreme example, the covariate takes only two values. We now study another example, in which the distribution of the covariate approximates the normal distribution.

We let $X_{i1} = 1$ for every $i$, and $X_{i2} = \phi^{-1}\left(\frac{i}{n+1}\right)$ for $i = 1, ..., n$. As $Y_i \in [0, 1]$ for all $i$, $\beta_2 \leq 1/(2X_{n2})$. The minimum of $\|\tau_2\|^2$ subject to $X'\tau_2 = e_2$ equals

$$\frac{1}{\sum_k X_{k2}^2},$$

it is minimized when

$$\tau_{i2} = \frac{X_{i2}}{\sum_k X_{k2}^2}.$$

The minimum of $\|\tau_2\|_\infty$ subject to $X'\tau_2 = e_2$ is equal to

$$\frac{1}{2\sum_{j=n/2+1}^{n} X_{j2}},$$

where the unique minimizer satisfies $|\tau_{i2}| = \|\tau_2\|_\infty$ for all $i$.

In this example we find numerically that the Bernoulli test performs better in terms of the bound on the type II error probability when one chooses $\tau_j$ equal to the minimizer of $\|\tau_2\|_\infty$, which means that $d$ is unique and $p_{\bar{\beta}_j} = 1/2$, instead of choosing $\tau_j$ equal to the minimizer of $\|\tau_2\|$ where $p_{\bar{\beta}_j}$ can be chosen much larger. The reason seems to be that the value of $\|\tau_2\|_\infty$ is more than double in the latter case than in the former case.

Analytic computation shows that $\bar{\sigma}_{\beta_j}$ is given by equation (16) in Lemma 8. Unlike in the extreme example, $\bar{\sigma}_{\beta_j}$ is strictly smaller than the bound presented in Lemma 6. For instance, when $n = 60$ then $\bar{\sigma}^2_{\beta_j} = 0.0047 - 4.2 \times 10^{-2}\beta_2^2$ while the bound given in Lemma 6 equals $0.0047 - 1.67 \times 10^{-2}\beta_2^2$. For our calculations below the difference between these two bounds plays less of a role as the sample gets larger. For the value of $\beta_2$ used in the table below, when $n = 500$ then $\bar{\sigma}^2_{\beta_j} = 4.8 \cdot 10^{-4}$ while the bound from Lemma 6 is $5.1 \cdot 10^{-4}$. Hence, relying on Lemma 6 to construct the tests would lead to a slightly less powerful test than relying on the exact value as we do.

As in the extreme example, given $\alpha = 0.05$, $\bar{t}_N = \bar{t}_H$, the cutoff of the Non-Standardized test is determined by the bound derived using Hoeffding's inequality. We find that the bound on the probability of a type II error derived using the Berry-Esseen inequality is sharper when $\tau_2$ is chosen as in the OLS method as compared to when it minimizes $\|\tau_2\|_\infty$.

We summarize our results in Tables 3 and 4.

The Non-Standardized test is best for guaranteeing type II error below 0.2 in small samples and for guaranteeing it to be below 0.5 in large samples. In these cases too

| $n$ | $\bar{t}_N$ | $\bar{k}_B/n - \bar{p}$ | $\beta_2$ | C | Y | H | BE | B |
|---|---|---|---|---|---|---|---|---|
| 60 | 0.168 | 0.17 | 0.212 | 0.587 | 1 | 0.811 | 0.912 | **0.5** |
| 100 | 0.127 | 0.12 | 0.162 | 0.622 | 1 | 0.798 | 0.851 | **0.5** |
| 500 | 0.0553 | 0.054 | 0.0711 | 0.659 | 1 | 0.784 | 0.695 | **0.5** |
| 4000 | 0.0194 | 0.016 | 0.0253 | 0.637 | 1 | 0.754 | 0.524 | **0.5** |
| 6000 | 0.0158 | 0.013 | 0.0207 | 0.637 | 1 | 0.753 | **0.5** | **0.5** |
| 8000 | 0.0137 | 0.012 | 0.0177 | 0.661 | 1 | 0.774 | **0.5** | 0.516 |

Table 3: Comparison of methods in the normal example for maximal type II error probabilities of 0.5.

| $n$ | $\bar{t}_N$ | $\bar{k}_B/n - \bar{p}$ | $\beta_2$ | C | Y | H | BE | B |
|---|---|---|---|---|---|---|---|---|
| 60 | 0.168 | 0.15 | 0.253 | 0.217 | **0.2** | 0.465 | 0.622 | 0.264 |
| 100 | 0.127 | 0.12 | 0.201 | 0.232 | **0.2** | 0.367 | 0.526 | 0.218 |
| 500 | 0.0553 | 0.054 | 0.0908 | 0.269 | 0.222 | 0.292 | 0.376 | **0.2** |
| 500 | 0.0553 | 0.046 | 0.11 | 0.14 | 0.0584 | 0.0539 | 0.221 | **0.05** |
| 4000 | 0.0194 | 0.016 | 0.0261 | 0.28 | 0.22 | 0.28 | 0.255 | **0.2** |
| 6000 | 0.0158 | 0.013 | 0.0261 | 0.28 | 0.218 | 0.28 | 0.237 | **0.2** |

Table 4: Comparison of methods in the normal example for maximal type II error probabilities of 0.2 and 0.05.

much information is lost due to the rescaling of variables within the Bernoulli test. Otherwise the Bernoulli test performs best.

# 7 Conclusion

The question of testing and building confidence intervals for parameters of a linear regression in the presence of heteroskedasticity is a long standing one in Econometrics. White (1980) introduced an asymptotic solution to this problem. This paper introduces several finite sample methods that are exact in the sense that they do not rely on assumptions on the noise terms beyond independence.

The tests rely on a known bound on range the dependent variable. Such bounds are known in most practical cases, and as shown by Bahadur and Savage (1956), no finite sample exact methods exist if this assumption is relaxed. Until now, one had to apply asymptotic solutions in the analysis of finite sample data, without any control of the rate of convergence of the finite sample test statistics distribution to the asymptotic one. Note also that, in the fixed regressor case, White's asymptotic approach requires a bound on the range of the covariates, and the rate of convergence of the finite test to the asymptotic test necessarily relies on an assumption such as a bound on the range of the dependent variable, or, alternatively, its variance.

The tests are easy to implement. In some cases they contain free parameters that require fine tuning, in other cases we can directly present the formula, such as when the cutoff under the Non-Standardized test is derived using Hoeffding's inequality (see (9)). Similarly, the proofs are straightforward. In most cases their construction builds on existing inequalities.

The general methods we follow to construct these tests can be extended. For instance, improvements on the tail inequalities presented in the Non-Standardized naturally lead to improvements the of Non-Standardized test, and, similarly, improvements on the Shiganov bound of the Berry-Esseen inequality would improve the power of the Non-Standardized and Standardized tests.

Evaluating the type II error probabilities numerically, we nd that our tests perform well even in small sample sizes (n=40,60), for which there is a strong doubt on the reliability of asymptotic methods.

# References

Bahadur, R. R., and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *The Annals of Mathematical Statistics*, 27, 1115–1122.

Berry, A. C. (1941): "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, 49, 122–136.

Bhattacharyya, B. B. (1987): "One sided Chebyshev inequality when the first four moments are known," *Communications in Statistics - Theory and Methods*, pp. 2789–2791.

Bienaymé, I. (1853): "Considrations lappui de la dcouverte de Laplace sur la loi de probabilit dans la mthode des moindres carrs," *Comptes Rendus de l'Acadmie des Sciences*, 37, 309324.

Cantelli, F. (1910): "Intorno ad un teorema fondamentale della teoria del rischio," *Bolletino dell' Associazione degli Attuari Italiani*.

Davidson, R., and J. MacKinnon (1993): *Estimation and Inference in Econometrics*. Oxford University Press, Oxford, UK.

Dufour, J.-M. (2003): "Identification, Weak Instruments, and Statistical Inference in Econometrics," *The Canadian Journal of Economics / Revue canadienne d'Economique*, 36, 767–808.

Dufour, J.-M., and M. Hallin (1993): "Improved Eaton Bounds for Linear Combinations of Bounded Radom Variables, With Statistical Applications," *Journal of the American Statistical Association*, 88, 1026–1033.

Esseen, C.-G. (1942): "On the Liaponnoff limit of error in the theory of probability," *Arkiv för matematik, astronomi och fysik*, A28, 1–19.

Greene, W. H. (2002): *Econometric Analysis*. Pearson Education, Upper Saddle River, NJ, USA, fifth edn.

Hoeffding, W. (1956): "On the Distribution of the Number of Successes in Independent Trials," *The Annals of Mathematical Statistics*, 27, 713–721.

——— (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13–30.

Lehmann, E., and J. Romano (2005): *Testing Statistical Hypotheses*. Springer, New-York, USA, third edn.

MacKinnon, J., and H. White (1985): "Some Heteroscedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 19, 305–325.

Pratt, J. W. (1961): "Length of confidence intervals," *Journal of the American Statistical Association*, 56, 549–567.

Schlag, K. (2006): "Designing Non-Parametric Estimates and Tests for Means," Economics Working Papers ECO2006/26, European University Institute.

——— (2008a): "Exact Tests for Correlation and for the Slope in Simple Linear Regressions without Making Assumptions," Universitat Pompeu Fabra working paper 1097.

——— (2008b): "A New Method for Constructing Exact Tests without Making any Assumptions," Universitat Pompeu Fabra working paper 1109.

Shiganov, I. S. (1986): "Refinement of the upper bound of the constant in the central limit theorem," *Journal of Mathematical Sciences*, 35, 2545–2550.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

# A Upper Bound on $\sigma_{\beta_j}$

**Lemma 8** *(i)*

$$\sigma_{\beta_j}^2 \leq \frac{1}{4}\left\|\tau_j\right\|^2 - \frac{1}{\#\left\{i : \tau_{ij} \neq 0\right\}}\left(\beta_j - \frac{1}{2}\sum_i \tau_{ij}\right)^2 ,$$

*(ii) if $X_{i1} = 1$ for all $i$ and $m = j = 2$ then*

$$\sigma_{\beta_2}^2 \leq \beta_2 \sum_i \tau_{i2}^2 X_{i2}\left(1 - \beta_2 X_{i2}\right) + \frac{\left(\left\|\tau_2\right\|^2 - 2\beta_2 \sum_i \tau_{i2}^2 X_{i2}\right)^2}{4\left\|\tau_2\right\|^2} \tag{16}$$

*where this bound is tight when*

$$\left|\beta_2\left(X_{i2} - \frac{\sum_k \tau_{k2}^2 X_{k2}}{\left\|\tau_2\right\|^2}\right)\right| \leq \frac{1}{2} \text{ for all } i.$$

**Proof of Lemma 8.** It is a simple excersize to show that

$$\frac{1}{4}\left\|\tau_j\right\|^2 - \frac{1}{n_1}\left(\beta_j - \frac{1}{2}\sum_i \tau_{ij}\right)^2 = \max_{\mu \in \mathbb{R}^n}\left\{\sum_i \tau_{ij}^2 \mu_i\left(1 - \mu_i\right) : \sum_i \tau_{ij}\mu_i = \beta_j\right\}$$

where the maximum is attained when

$$\mu_i = \frac{1}{2} + \frac{1}{n\tau_{ij}}\left(\beta_j - \frac{1}{2}\sum_i \tau_{ij}\right) \text{ for } \tau_{ij} \neq 0.$$

A better, possibly strictly lower bound is obtained if the constraint $\mu \in [0,1]^n$ is included in the above maximization.

For the best bound, using the fact that $VarY_i \leq EY_iE(1-Y_i)$, one needs to solve

$$\max_{z\in\mathbb{R}^m}\left\{\sum_i \tau_{ij}^2 X_i z(1-X_i z) : z_j = \beta_j,\ Xz \in [0,1]^n\right\}.$$

For the special case where $X_{i1} = 1$ for all $i$ and $m = j = 2$ we obtain, when ignoring the constraint $\beta_1 + \beta_2 X_{i2} \in [0,1]$ for all $i$,

$$\sigma_{\beta_2}^2 \leq \beta_2 \sum_i \tau_{i2}^2 X_{i2}(1-\beta_2 X_{i2}) + \frac{\left(\|\tau_2\|^2 - 2\beta_2\sum_i \tau_{i2}^2 X_{i2}\right)^2}{4\|\tau_2\|^2},$$

the value of $\beta_1$ used to attain this maximum is given by

$$\beta_1 = \frac{1}{2} - \beta_2 \frac{\sum_i \tau_{i2}^2 X_{i2}}{\|\tau_2\|^2}.$$

Hence, the above bound is tight if

$$\frac{1}{2} - \beta_2 \frac{\sum_k \tau_{k2}^2 X_{k2}}{\|\tau_2\|^2} + \beta_2 X_{i2} \in [0,1] \text{ for all } i.$$

■

# B Proof of Lemma 7

Let $S_0 = \sum_i \tau_{ij}^2 (Y_i - X_i\beta)^2$.

**Lemma 9** $\left|s_{\beta_j}^2 - S_0\right| \leq c_0 \cdot \left\|\hat{\beta} - \beta\right\|_\infty$ *where*

$$c_0 := \max_{Y\in[0,1]^n} \sum_i \tau_{ij}^2 \left(\left|X_i\tau^T Y\right| + 2\right)\|X_i\|_1.$$

**Proof.** Using the quadratic formula $(y^2 - x^2) = (y + x)(y - x)$ we derive

$$\begin{aligned}
\left|s^2_{\beta_j} - S_0\right| &= \left|\sum_i \tau^2_{ij}\left(\left(X_i\hat{\beta}\right)^2 - (X_i\beta)^2 - 2Y_i e^t_i X\left(\hat{\beta} - \beta\right)\right)\right| \\
&= \left|\sum_i \tau^2_{ij}\left(X_i\left(\hat{\beta} + \beta\right) - 2Y_i\right) X_i\left(\hat{\beta} - \beta\right)\right| \\
&\leq \left\|\hat{\beta} - \beta\right\|_\infty \cdot \left|\sum_{i,k=1}^{n} \tau^2_{ij}\left(X_i\left(\hat{\beta} + \beta\right) - 2Y_i\right) X_{ik}\right| \\
&= \left\|\hat{\beta} - \beta\right\|_\infty \cdot \left|\sum_i \tau^2_{ij}\left(X_i\left(\hat{\beta} + \beta\right) - 2Y_i\right) \sum_{k=1}^{m} X_{ik}\right| \\
&\leq \left\|\hat{\beta} - \beta\right\|_\infty \cdot \sum_i \tau^2_{ij}\left(\left|X_i\hat{\beta}\right| + |Y_i| + |X_i\beta - Y_i|\right)\|X_i\|_1 \\
&\leq \left\|\hat{\beta} - \beta\right\|_\infty \cdot \sum_i \tau^2_{ij}\left(\left|X_i\hat{\beta}\right| + 2\right)\|X_i\|_1 .
\end{aligned}$$

■

**Lemma 10** $P\left(\left|\hat{\beta}_k - \beta_k\right| \geq a\right) \leq 2\exp\left(-2\frac{a^2}{\|\tau_k\|^2}\right) =: c_1(a)$.

This follows directly from Hoeffding's inequality (5). The only difference to our analysis is Section 3.1.3 is here the factor 2 which is due to the fact that the inequality we approximate above is two-sided.

**Lemma 11** $P\left(\sigma^2_{\beta_j} - S_0 \geq a\right) \leq \exp\left(-2\frac{a^2}{\sum_i \tau^4_{ij}}\right) =: c_2(a)$.

Again this follows again directly from (5). In contrast to Lemma 10 we do not need the factor 2 as the approximation of the error is one-sided.

**Proof of Lemma 7.** Let $\lambda \in [0, 1]$. Following Lemmata (9) and (10) we obtain

$$P\left(\left|s^2_{\beta_j} - S_0\right| \geq \lambda a^2\right) \leq P\left(\left\|\hat{\beta} - \beta\right\|_\infty \geq \lambda\frac{a^2}{c_0}\right) \leq c_1\left(\lambda\frac{a^2}{c_0}\right).$$

Following Lemma (11),

$$P\left(\sigma^2_{\beta_j} - S_0 \geq (1 - \lambda) a^2\right) \leq c_2\left((1 - \lambda) a^2\right).$$

Since

$$\left\{\sigma^2_{\beta_j} - s^2_{\beta_j} \geq a^2\right\} \subset \left\{\left|S_0 - s^2_{\beta_j}\right| \geq \lambda a^2\right\} \cup \left\{\sigma^2_{\beta_j} - S_0 \geq (1 - \lambda) a^2\right\}$$

we obtain

$$\begin{aligned} P\left(\sigma^2_{\beta_j} - s^2_{\beta_j} \geq a^2\right) &\leq P\left(\left|S_0 - s^2_{\beta_j}\right| \geq \lambda a^2\right) + P\left(\sigma^2_{\beta_j} - S_0 \geq (1-\lambda)\, a^2\right) \\ &\leq c_1\left(\lambda \frac{a^2}{c_0}\right) + c_2\left((1-\lambda)\, a^2\right). \end{aligned}$$

■

# C Proof of Theorem 1

**Lemma 12** *$c_0 \leq \frac{K_1}{n}$, with $K_1 = m^3 K^2 (m^2 K^2 + 2)$.*

**Proof.**

$$\begin{aligned} c_0 &= \max_{Y \in [0,1]^n} \sum_i \tau^2_{ij} \left(\left|X_i \left(X'X\right)^{-1} X'Y\right| + 2\right) \|X_i\|_1 \\ &\leq \left(\sum_i \tau^2_{ij}\right) \left(\left\|X \left(\frac{X'X}{n}\right)^{-1} X'\right\|_\infty + 2\right) m \\ &\leq \frac{m^2 K^2}{n} \left(m^2 K + 2\right) m. \end{aligned}$$

■

**Lemma 13** *$c_3(an^{-\frac{1}{20}}) \leq 1/n^2$ for $n$ large enough, $R(w) \to 0$, $\sigma_{\beta_j}/w \to \infty$, $s_{\beta_j}/w \to \infty$ a.s. as $n \to \infty$.*

**Proof.** First, $c_3(an^{-\frac{1}{20}})$:

$$\begin{aligned} c_3(an^{-\frac{1}{20}}) &\leq c_1\left(\frac{a^2 n^{-.1}}{2c_0}\right) + c_2\left(\frac{a^2 n^{-.1}}{2}\right) \\ &\leq 2\exp\left(-\frac{a^4 n^{-.2}}{2c_0^2 \left\|\tau_{*j}\right\|^2}\right) + \exp\left(-\frac{a^4 n^{-.2}}{2\sum_i \tau^4_{ij}}\right) \\ &\leq 2\exp\left(-\frac{a^4 n^{2.8}}{2K_1^2 K^2}\right) + \exp\left(-\frac{a^4 n^{2.8}}{2K^4}\right). \end{aligned}$$

It follows straight from the definition of $R(w)$ that

$$R(w) \leq \frac{K^3}{n^2 w^3}.$$

Letting $\Omega = diag(E\varepsilon_i^2)$,

$$\begin{aligned}\sigma_{\beta_j}^2 &= e_j'(X'X)^{-1}X\Omega X'(X'X)^{-1}e_j \\ &= \frac{1}{n}e_j'\left(\frac{X'X}{n}\right)^{-1}\bar{V}\left(\frac{X'X}{n}\right)^{-1}e_j \\ &= \frac{1}{n}\left\|\bar{V}^{\frac{1}{2}}\left(\frac{X'X}{n}\right)^{-1}e_j\right\|^2.\end{aligned}$$

Since $\bar{V}^{\frac{1}{2}}\left(\frac{X'X}{n}\right)^{-1}$ has bounded terms and its determinant is bounded away from 0, $n\sigma_{\beta_j}^2$ is bounded away from 0, which implies that $\sigma_{\beta_j}/w \to \infty$.

Finally, $s_{\beta_j}^2/w^2 \geq \sigma_{\beta_j}^2/w^2 - |\sigma_{\beta_j}^2 - s_{\beta_j}^2|/w^2$, and $c_3(wn^{-\frac{1}{20}}) = c_3(an^{-\frac{1}{20}}) \leq 1/n^2$ implies that $|\sigma_{\beta_j}^2 - s_{\beta_j}^2|/w^2 \to 0$ a.s., hence $s_{\beta_j}^2/w^2 \to \infty$ a.s.. ■

**Proof of Theorem 1.** Point (1) is a direct consequence of $b_1 \to 0$, $a_1 \ll s_{\beta_j}$ a.s. (cf. Lemma 13). For point (2), it is enough to see that $c_3(a_1)\phi(b_1/w)+AR(w) \to 0$, which is straightforward from Lemma 13. For (3), since $c_3(\bar{a}), AR(w) \to 0$, $b_2/w \to \infty$, it is enough to establish that $\bar{u} \to \phi^{-1}(1-\alpha) + \frac{\bar{\beta}_j - \beta_j}{\sigma_{\beta_j}}$:

$$\bar{u} = \bar{t}_S\sqrt{\frac{\sigma_{\beta_j}^2 + 3w^2}{\sigma_{\beta_j}^2 + w^2}} + \frac{2b_1}{\sqrt{\sigma_{\beta_j}^2 + w^2}} + \frac{\bar{\beta}_j - \beta_j}{\sigma_{\beta_j}}\frac{\sigma_{\beta_j}}{\sqrt{\sigma_{\beta_j}^2 + w^2}}$$

and the result follows since $\bar{t}_S \to \phi^{-1}(1-\alpha)$, $b_1, w \ll \sigma_{\beta_j}$, $\frac{\bar{\beta}_j - \beta_j}{\sigma_{\beta_j}}$ constant. ■