

# FINITE-SAMPLE EXACT TESTS FOR LINEAR REGRESSIONS WITH BOUNDED DEPENDENT VARIABLES

OLIVIER GOSSNER AND KARL H. SCHLAG

ABSTRACT. We introduce tests for finite-sample linear regressions with heteroskedastic errors. The tests are exact, i.e., they have guaranteed type I error probabilities when bounds are known on the range of the dependent variable, without any assumptions about the noise structure. We provide upper bounds on probability of type II errors, and apply the tests to empirical data.

**Keywords:** Nonparametric linear regression, exact test, heteroskedasticity

**JEL classification:** C12

## 1. INTRODUCTION

The fundamental goal of hypothesis testing, as set by Neyman and Pearson (1930), is the minimization of both type I and type II error probabilities. To cite Neyman and Pearson (1930, p 100) : *(1) we must be able to reduce the chance of rejecting a true hypothesis to as low a value as desired; (2) the test must be so devised that it will reject the hypothesis tested when it is likely to be false.*

In the usual model of linear regressions, how well are these goals achieved? When error terms are normally distributed and homoskedastic, the classical test<sup>1</sup> has a type I error probability equal to the nominal level of the test. But error terms in real data almost never have a precisely normal distribution, let alone a homoskedastic one. For any given heteroskedastic noise structure, White (1980)'s robust test

---

*Date:* April 2013.

Olivier Gossner: Paris School of Economics, Paris, and London School of Economics, London; email: [ogossner@gmail.com](mailto:ogossner@gmail.com), Karl Schlag: Department of Economics, University of Vienna, Austria; email: [karl.schlag@univie.ac.at](mailto:karl.schlag@univie.ac.at).

Olivier Gossner benefited from the support of the STICERD grant "Exact nonparametric tests for linear regressions - theory and applications" and from the LabEx OSÉ. Karl Schlag gratefully acknowledges financial support from the Ministerio de Educación y Ciencia de España, Grant MEC-SEJ2006-09993 and from the Barcelona Graduate School of Economics. We are grateful to an anonymous referee for useful comments.

<sup>1</sup>In the remaining of the paper, "classical test" refers to the classical OLS  $t$ -test assuming normal and homoskedastic errors.

guarantees a type I error probability that approaches the nominal level when the sample size goes to infinity. But without restrictions on the (unknown) noise structure, and for any sample size, the probability of a type I error resulting from the use of White (1980)'s test can be as large as 1. In fact, this is a consequence of a general impossibility result due to Bahadur and Savage (1956) and Dufour (2003) that shows that no meaningful test can be constructed in which the probability of a type I error is guaranteed to be less than 1.

The use of statistical tools in situations where the underlying distributional assumptions are not satisfied can have catastrophic consequences. Practitioners can be lead to greatly underestimate the probability of certain outcomes, and remain unprepared to those outcomes while thinking they are safe. This is what must have happened to David Viniar, CFO of Goldman Sachs, who declared in August 2007 about the financial crisis: "We were seeing things that were 25-standard deviation moves, several days in a row" (quoted in Larsen, 2007). Since the probability of a 25-standard deviation event under the normal distribution is less than 1 over  $10^{137}$ , we can safely conclude that the distributional assumptions used by Viniar and his colleagues were not satisfied.

In this paper our message is a positive one. We identify an important class of statistical problems where the negative conclusions from Bahadur and Savage (1956) and Dufour (2003) do not apply, and we introduce tests with guaranteed upper bounds on type I and type II errors for this class of problems. The tests are exact in the sense that they guarantee a type I error probability below the nominal level independently of the error structure<sup>2</sup>. We also implement our these tests in practical numerical examples.

The class of problems we consider is the class in which a bound on the dependent variable is known. This condition is satisfied in a large range of applications. For instance, it is warranted by the very nature of the endogenous variable (e.g., proportions, success or failure, test scores) in 43 of the 75 papers using linear models

---

<sup>2</sup>This is the same sense as in e.g. Yates (1934) and Dufour and Torres (2000).

published in 2011 in the *American Economic Review*. It should be noted that even under the boundedness condition, the existence of exact tests was previously an open problem. Previous exact tests were derived by Schlag (2006, 2008a) for the mean of a random variable and for the slope of a simple linear regression. Exact tests for linear regressions under the alternative assumption that error terms have median zero are developed by Dufour and Hallin (1993), Boldin *et al.* (1997), Chernozhukov *et al.* (2009), Coudin and Dufour (2009), Dufour and Taamouti (2010).

We refer to our two tests as the nonstandardized and the Bernoulli tests. We briefly summarize their constructions here. Each test relies on a linear combination of the dependent variables (such as in the OLS method) that is an unbiased estimator of the coefficient to be tested.

The nonstandardized test relies on inequalities due to Cantelli (1910), Hoeffding (1963), and Bhattacharyya (1987), as well as on the Berry-Esseen inequality (Berry, 1941; Esseen, 1942; Shevtsova, 2010), to bound the tail probabilities of the unbiased estimator. One challenge in the construction is to apply the Berry-Esseen inequality even though there is no lower bound on the variance of any of the error terms.

The Bernoulli test generalizes the methodology introduced by Schlag (2006, 2008b) for mean tests. Each term of the linear combination that constitutes the unbiased estimator is probabilistically transformed into a Bernoulli random variable. We then design a test for the mean of the family obtained using Hoeffding (1956)'s bound on the sum of independent Bernoulli random variables. This defines a randomized test, on which we then rely to construct a deterministic test.

We provide bounds on the probabilities of type II error of our tests. These bounds can be used to select – depending on the sample size and the realization of the exogenous variables – which of our tests is most appropriate. We also rely on these bounds to show that these tests have enough power for practical applications.

In two canonical numerical examples involving one covariate in addition to the constant, the bounds on the probability of type II errors show that the tests perform well even for small sample sizes (e.g.,  $n = 40$ ).

We implement our tests and compute confidence intervals using the empirical data from Duflo *et al.* (2011).<sup>3</sup> We compare the results relying on our test with the 95/ ones obtained using either the classical method and White's heteroskedastic robust method. The results show that, compared to the classical test or White's test, the losses of significance of our exact method are moderate, and the confidence intervals are in most cases augmented by a factor of no more than 50 percent.

The paper is organized as follows. Section 2 introduces the model. Sections 3 and 4 present the nonstandardized test and the Bernoulli test. In Section 5, we examine their efficiency using numerical examples. Section 7 shows an application of the tests to empirical data. The underlying data-generating process is discussed and extensions are discussed in Section 8. We conclude in Section 9. All proofs are presented in the appendix.

## 2. LINEAR REGRESSION

We consider the standard linear regression model with random regressors, given by

$$(1) \quad Y_i = X_i\beta + \varepsilon_i, \quad i = 1, \dots, n$$

where  $X_i$  is the  $i$ -th row of a random matrix  $X \in \mathbb{R}^{n \times m}$  of independent variables,  $\beta \in \mathbb{R}^m$  is the vector of unknown coefficients, and  $\varepsilon \in \mathbb{R}^n$  is the random vector of errors. The *fixed regressor case* in which  $X$  is nonrandom and known ex-ante to the statistician is a special case. We assume (i) *strict exogeneity*:  $\mathbf{E}(\varepsilon|X) = 0$  a.s., and (ii) almost surely no *multicollinearity*:  $X$  has rank  $m$  with probability 1. To keep the exposition simple, in most of the paper we also assume (iii) *conditional independence of errors*:  $(\varepsilon_i)_i$  are independent conditional on  $X$ . Finally, we assume (iv) *bounded endogenous variable*:<sup>4</sup> there exist  $\omega$  and  $\omega'$  with  $\omega < \omega'$  such that  $P(Y_i \in [\omega, \omega']) = 1$  for  $i = 1, \dots, n$ . In particular, (iv) implies that  $X_i\beta \in [\omega, \omega']$

<sup>3</sup>The software used to implement the tests is freely downloadable from the authors' webpages.

<sup>4</sup>Without any restriction on the support of  $Y$ , the possibility of very small or very large outcomes that occur with very small probability (fat tails) make it impossible to make any inference about  $\mathbf{E}Y$  based on the observed values of  $Y$ , as shown by Bahadur and Savage (1956) when testing for means and by Dufour (2003) in linear regression analysis.

almost surely, and ensures the existence of all moments of  $\varepsilon_i$  for  $i = 1, \dots, n$ . We assume that  $\omega' = \omega + 1$ ; this is without loss of generality since we can reduce other cases to this one by dividing each side of equation (1) by  $\omega' - \omega$ . We relax (iii) and (iv) in Section 8. The assumptions (i) to (iv) are stronger than those of e.g. White (1980), but are sufficient to guarantee the existence of unbiased estimators (not just asymptotically so) of  $\beta$ . We do not make any further assumptions about the error terms, such as  $\text{Var}(\varepsilon_i) > 0$  or homoskedasticity.

We present two exact tests at the level of significance  $\alpha > 0$  for the one-sided hypothesis  $H_0 : \beta_j \leq \bar{\beta}_j$  against  $H_1 : \beta_j > \bar{\beta}_j$ , where  $\bar{\beta}_j \in \mathbb{R}$ .<sup>5</sup> *Exact* means that the probability of a type I error of the test is proven to be at most  $\alpha$  for any random vectors  $(X, \varepsilon)$  that satisfy (i)-(iv). In particular, bounds on the probabilities of type I errors are guaranteed for every given sample size.

Both tests have a type I error probability bounded by the nominal level conditional on the realization of  $X$ . This allows to combine both tests into another exact test according to the following procedure: given  $X$ , select the test for which our bounds guarantee a type II error probability below 0.5 for the largest range of the parameter to be tested. This is the procedure that we implement in our software and in the numerical applications.

### 3. THE NONSTANDARDIZED TEST

Assumption (ii) ensures the existence of  $\tau_j \in \mathbb{R}^n$  such that  $X'\tau_j = e_j$ , where  $e_{jj} = 1$  and  $e_{jk} = 0$  for  $k \neq j$ . For such  $\tau_j$ ,  $\hat{\beta}_j = \tau_j'Y$  is an unbiased estimate of  $\beta_j$ . One example of  $\tau_j$  is the system of coefficients for which  $\tau_j'Y$  is the OLS estimate of  $\beta_j$ . We present a test for a given such vector  $\tau_j$ , and later discuss the choice of  $\tau_j$ . We let  $\|\cdot\|_\infty$  denote the supremum norm, and  $\|\cdot\|$  denote the Euclidian norm, while  $\Phi$  denotes the cumulative normal distribution.

Consider the functions defined for  $\sigma > 0$ ,  $t > 0$ , and  $\tau_j \in \mathbb{R}^n$ :

---

<sup>5</sup>Tests of  $H_0 : \beta_j \geq \bar{\beta}_j$ ,  $H_0 : \beta_j = \bar{\beta}_j$ , and confidence intervals are derived easily, see Section 8.

$$\begin{aligned}
\varphi_C(\sigma, t) &= \frac{\sigma^2}{\sigma^2 + t^2} \\
\varphi_{Bh}(\sigma, t, \tau_j) &= \begin{cases} \frac{3\sigma^4}{4\sigma^4 - 2\sigma^2 t^2 + t^4} & \text{if } \frac{t^2 - t\|\tau_j\|_\infty}{\sigma^2} > 1, \sigma^2 \leq \frac{t^2\|\tau_j\|_\infty}{\|\tau_j\|_\infty + 3t} \\ \frac{(3\sigma^2 - \|\tau_j\|_\infty^2)\sigma^2}{(3\sigma^2 - \|\tau_j\|_\infty^2)(\sigma^2 + t^2) + (t^2 - t\|\tau_j\|_\infty - \sigma^2)^2} & \text{if } \frac{t^2 - t\|\tau_j\|_\infty}{\sigma^2} > 1, \sigma^2 > \frac{t^2\|\tau_j\|_\infty}{\|\tau_j\|_\infty + 3t} \\ 1 & \text{if } \frac{t^2 - t\|\tau_j\|_\infty}{\sigma^2} \leq 1 \end{cases} \\
\varphi_H(t, \tau_j) &= \exp\left(-\frac{2t^2}{\|\tau_j\|^2}\right) \\
\varphi_{BE}(\sigma, t) &= \inf_{w>0, b_1 \in \mathbb{R}} \frac{1 - \Phi\left(\frac{t-b_1}{\sqrt{\sigma^2+w^2}}\right) + A \frac{2\|\tau_j\|_\infty}{\sqrt{27}w}}{\Phi(b_1/w)} \text{ with } A = .56
\end{aligned}$$

and

$$\varphi(\sigma, t, \tau_j) = \min\{\varphi_C(\sigma, t), \varphi_{Bh}(\sigma, t, \tau_j), \varphi_H(t, \tau_j), \varphi_{BE}(\sigma, t)\}.$$

The tests use the following bound (see Lemma 3 in the appendix) on the variance of  $\hat{\beta}_j$  as a function of  $\beta_j$ :

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

and the following bound on the variance of  $\hat{\beta}_j$  under the null hypothesis:

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

It can be easily verified that  $\varphi$  is continuously decreasing in  $t$  and that  $\lim_{t \rightarrow \infty} \varphi(\bar{\sigma}_{0, \beta_j}, t, \tau_j) = 0$ . Hence, for  $0 < \alpha < 1$ , there exists a minimal value  $\bar{t}_N$  such that  $\varphi(\bar{\sigma}_{0, \beta_j}, \bar{t}_N, \tau_j) \leq \alpha$ . We define the nonstandardized test as the one that rejects the null hypothesis when  $\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}_N$ . The choice of  $\bar{t}_N$  maximizes the analytical bound on the power of the test (it has the largest rejection area) while ensuring that the probability of a type I error is bounded by  $\alpha$ .

**Theorem 1.** *The nonstandardized test has a type I error probability bounded above by  $\alpha$ . For each realization of  $X$ , the type II error probability is bounded above by  $\varphi(\bar{\sigma}_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}_N, \tau_j)$  for every  $\beta_j \geq \bar{\beta}_j - \bar{t}_N$ .*

To prove Theorem 1, we rely on probabilistic bounds on the deviation of  $\hat{\beta}_j$  from its mean  $\beta_j$ . Bounds on the probability that  $\hat{\beta}_j \geq \bar{\beta}_j + t$  under the null hypothesis for  $t > 0$  provide bounds on the type I error probability, while bounds on the probability that  $\hat{\beta}_j - \bar{\beta}_j < t$  give bounds on the type II error probability. More precisely, we use inequalities due to Cantelli (1910), Bhattacharyya (1987), and Hoeffding (1963) to prove that under the null hypothesis,  $P(\hat{\beta}_j \geq \bar{\beta}_j + \bar{t}_N)$  is bounded above by  $\varphi_C(\bar{\sigma}_{0,\beta_j}, \bar{t}_N)$ ,  $\varphi_{Bh}(\bar{\sigma}_{0,\beta_j}, \bar{t}_N, \tau_j)$ , and  $\varphi_H(\bar{t}_N, \tau_j)$ . We also use Berry-Esseen's inequality (Berry, 1941; Esseen, 1942) with the constant obtained by Shevtsova (2010) to bound this probability by  $\varphi_{BE}(\bar{\sigma}_{0,\beta_j}, \bar{t}_N)$ . We finally obtain the bounds on the probabilities of type I and type II errors of Theorem 1 by combining these inequalities. Our test statistic as well as the rejection threshold depend on  $X$ . Since all probability bounds we use hold conditional on  $X$ , the type I error probability is bounded above by  $\alpha$  conditional (and therefore also unconditional) on  $X$ .

The test is called “nonstandardized” since it relies on maximal bounds on the deviation of  $\hat{\beta}_j$  from its mean and does not try to estimate this bound from the data (as the classical test and White's test do).

In the definition of the nonstandardized test,  $\tau_j$  is any vector with the property that  $X'\tau_j = e_j$ . The bound on type II error probabilities specified in Theorem 1 can be used to select the vector of weights  $\tau_j$  that guarantees the largest range of parameters for which the type II error probability falls below a selected level. In practice, the system of weights  $\tau_j$  corresponding to the OLS estimator allows for a good performance of the test, as illustrated in Sections 5 and 7. It has the additional advantage that results are easily comparable to other tests based on the OLS estimate.

#### 4. THE BERNOULLI TEST

Like the nonstandardized test, the Bernoulli test is built on a vector  $\tau_j \in \mathbb{R}^n$  satisfying  $X'\tau_j = e_j$ , so that  $\hat{\beta}_j = \tau_j'Y$  is an unbiased estimate of  $\beta_j$ . The test also

depends on a parameter  $\theta > 0$  and on vector  $d \in \mathbb{R}^n$  such that for every  $i$ , both  $\tau_{ji}\omega + d_i$  and  $\tau_{ji}(\omega + 1) + d_i$  are in  $[0, \|\tau_i\|_\infty]$ .<sup>6</sup> We present the test for significance level  $\alpha$  first, and then discuss the choice of  $\tau$ ,  $\theta$ , and  $d$ .

As in Schlag (2006, 2008b), we reduce the problem of testing  $\beta_j$  against  $\bar{\beta}_j$  to testing the mean of a sequence of Bernoulli random variables. More precisely, consider a family  $(W_i)_i$  of independent Bernoulli random variables such that the probability of success of  $W_i$  is  $(\tau_{ji}Y_i + d_i)/\|\tau_j\|_\infty$ . The proportion of successes  $\bar{W} = \sum_i W_i/n$  has expected value  $p_{\beta_j} = \frac{\beta_j + \sum_i d_i}{n\|\tau_j\|_\infty}$ , and  $\bar{p} = p_{\bar{\beta}_j}$  is the maximum of  $p_{\beta_j}$  under the null hypothesis.

Since we allow the choice of  $\tau$ ,  $d$ , and  $\theta$  to depend on the realization  $X$ , the distribution of  $W$  depends on  $X$  as well as on  $Y$ , but is fully known given their realizations. Given realizations of  $X$  and  $Y$ , and for  $n \geq k \geq 0$ , we let  $F_{XY}(k)$  denote the probability that  $n\bar{W} \geq \bar{k}$ .

The Bernoulli test compares the tail distribution of  $\bar{W}$  given by  $F_{XY}$  with the tail of the binomial distribution of parameters  $(n, \bar{p})$ . For  $0 < p < 1$  and  $k \in \{0, \dots, n\}$ , it is useful to introduce the notation:

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

Let  $\bar{k}$  be the smallest integer such that  $\bar{k} > n\bar{p} + 1$  and  $B(\bar{k}, \bar{p}) \leq \theta\alpha$ , and let  $\lambda = \frac{\theta\alpha - B(\bar{k}, \bar{p})}{B(\bar{k}-1, \bar{p}) - B(\bar{k}, \bar{p})}$ .

The Bernoulli test rejects the null hypothesis if the following inequality is satisfied:

$$\lambda F_{XY}(\bar{k} - 1) + (1 - \lambda) F_{XY}(\bar{k}) \geq \theta.$$

**Theorem 2.** *The Bernoulli test has a type I error probability bounded above by  $\alpha$ . For each realization of  $X$ , the type II error probability is bounded above by*

$$\frac{1 - \lambda B(\bar{k} - 1, p_{\beta_j}) - (1 - \lambda) B(\bar{k}, p_{\beta_j})}{1 - \theta}.$$

when  $p_{\beta_j} > \bar{k}/n$ .

<sup>6</sup>Note that  $d_i = \|\tau_j\|_\infty - \max\{\tau_{ji}\omega, \tau_{ji}(\omega + 1)\}$  satisfies these constraints.

In a first step to prove Theorem 2, we build a randomized test that, based on a realization of  $(W_i)$ , rejects the null hypothesis for large enough values of  $\bar{W}$ . Recall that under the null hypothesis, the expected value of  $\bar{W}$  is at most  $\bar{p}$ . A theorem by Hoeffding (1956) shows that, for a given value of its expectation, the tail probability of  $\bar{W}$  is maximal when  $(W_i)_i$  is an i.i.d. family of random variables. That theorem yields a bound on the probability of a type I error for the randomized test as a function of the binomial distribution with parameter  $\bar{p}$ . We also obtain bounds on the probability of a type II error for the randomized test.

In a second step, we construct a deterministic test from the randomized test, as in Schlag (2006). This deterministic test rejects the null hypothesis at significance level  $\alpha$  whenever the probability that the randomized test rejects the null hypothesis at a significance level  $\theta\alpha$  exceeds  $\theta$ . We then bound the probability of type I and type II errors of the deterministic test.

The Bernoulli test, as the nonstandardized test, depends on the realization of  $X$ . The type I error probability is bounded by  $\alpha$  for each realization of  $X$ , thus also unconditional on  $X$ .

As in the case of the nonstandardized test, the bound on type II error probabilities specified in Theorem 2 can be used to select the parameters  $\tau_j$ ,  $d$ , and  $\theta$  that guarantee a type II error probability below a certain level for the largest range of parameters. In practical applications, good performance is obtained when  $\tau$  minimizes  $\|\tau_j\|_\infty$ ,  $d$  is given by  $d_i = \|\tau_i\|_\infty - \max\{\tau_{ji}\omega, \tau_{ji}(\omega + 1)\}$ , and  $\theta$  is computed numerically to minimize the value of  $\beta_j$  for which our bounds guarantee a type II error probability below 0.5. We follow this approach in Sections 5 and 7.

## 5. NUMERICAL APPLICATION OF OUR METHOD

We investigate the performance of our tests in two numerical examples. Both examples involve a constant and a second covariate. We test for  $H_0 : \beta_2 \leq 0$  against  $H_1 : \beta_2 > 0$ . For a given sample, and fixing a significance level  $\alpha$ , we look for the test that attains the smallest value of  $\beta_2$  such that the probability of a type II

error for this test is guaranteed to fall below 0.5. We choose the free parameters within each test according to the procedure explained at the end of Sections 3 and 4. In both tests, for computational simplicity, we consider  $\tau$  either to be the OLS estimator or the unbiased estimator that minimizes  $\|\tau_j\|_\infty$ .

In the first example, which we call the step example, the second covariate  $X_2$  takes only the values 0 and 1. The number of times that  $X_2$  takes the value 1 is denoted by  $h$ . The sample is balanced for  $h = n/2$ , and gets more and more unbalanced as  $h$  gets closer to 1. In the second example, which we call the uniform example,  $X_{i2}$  is uniformly distributed on  $[-1, 1]$ :  $X_{i2} = -1 + (2i - 1)/n$  for every  $i$ . In both examples we assume  $Y_i \in [0, 1]$  for every  $i$ , which constrains the values of  $\beta_2$  to belong to  $[0, 1]$  in the step example and to  $[-1/2, 1/2]$  in the step example.

Table 1 presents results in the step example, and Table 2 presents results in the uniform example. We consider different values of the sample size  $n$ , and vary  $h/n$  in the step example. The column  $\alpha$  shows the nominal significance level. The column  $\beta_2''$  reports the minimal value of  $\beta_2$  for which the bound on the probability of a type II error specified by either Theorem 1 or 2 is equal to 0.5. We select the test, reported in the third column, for which this bound is achieved.<sup>7</sup> We report  $B$  for the Bernoulli test, and, if the nonstandardized is selected, we report which one of the four bounds is binding when determining the threshold  $\bar{t}_N$  used for rejecting the null hypothesis, followed by which one is binding when deriving the type II error bound at  $\beta_2$ ,  $C$  for Cantelli,  $Bh$  for Bhattacharyya,  $H$  for Hoeffding, and  $BE$  for Berry-Esseen. In Table 2 we also report in the parentheses whether best performance is achieved by setting  $\tau$  equal to the OLS estimator, indicated by *ols*, or equal to the unbiased estimator that minimizes  $\|\tau_j\|_\infty$ , indicated by *mm*. Such entries are not reported in Table 1, as in the step example, the OLS estimator also minimizes  $\|\tau_j\|_\infty$ . Column “other type II” shows the type II error probability of the test that was not selected when  $\beta_2 \geq \beta_2''$ , as derived from our theorems, with

---

<sup>7</sup>Note that it is valid to make a selection between the two tests based on the observation of  $X$  since our bounds on the type I and type II error probabilities are conditional on this realization.

the symbols in parentheses indicating the binding bounds and the choice of  $\tau_j$  when applicable. Given the method for selecting the test, all these values are at least 0.5.

In column “true type I” we report true type I error probabilities under the additional restriction that  $Y \in \{0, 1\}$  (in which case the error structure is fully specified by the values of  $X$  and  $\beta$ ), rounding up below 0.01.<sup>8</sup> Finally, under “true type II” we report the type II error probabilities of our selected test using Monte-Carlo simulations under the additional assumption  $Y \in \{0, 1\}$ .

For instance, in the step example, the first line of Table 1 shows that with  $n = 40$  and  $h/n = .5$ , the Bernoulli test is selected for testing  $H_0 : \beta_2 \leq 0$  at (nominal) level 0.05. For  $\beta_2 \geq 0.2$  this test guarantees (by Theorem 4) a type II error probability below 0.5. For the same parameters, the nonstandardized test guarantees a type II error probability only below 0.94; it does so by using  $\varphi_H$  to derive the threshold  $\bar{t}_N$ , and by using  $\varphi_{BE}$  to derive the bound on the probability of a type II error. Under the additional restriction that  $Y \in \{0, 1\}$  the true size of our test is 0.01 and the true type II error probability is 0.3.

$n$	$h/n$	$\alpha$	$\beta_2''$	test*	other type II	true type I	true type II
40	0.50	0.05	0.40	B	0.94 (H,BE)	0.01	0.30
40	0.25	0.05	0.60	H,C	0.54 (B)	0.01	0.16
100	0.50	0.05	0.25	B	0.84 (H,BE)	0.01	0.32
100	0.25	0.05	0.39	H,C	0.59 (B)	0.01	0.15
500	0.1	0.05	0.26	H,C	0.80 (B)	0.01	0.14
5000	0.50	0.05	0.03	B	0.59 (BE,BE)	0.01	??

TABLE 1. Step example with  $H_0 : \beta_2 \leq 0$  and  $H_1 : \beta_2 \geq \beta_2''$ . “true type I” and “true type II” errors are obtained by Monte-Carlo simulations under the additional assumption that  $Y \in \{0, 1\}$ .

$n$	$\alpha$	$\beta_2''$	test*	other type II	true type I	true type II
60	0.05	0.32	B,mm	0.78 (H,C,ols)	0.01	0.37
500	0.05	0.11	B,mm	0.63 (H,BE,ols)	0.01	0.35
6000	0.05	0.033	H,BE,ols	0.51 (B,mm)	0.01	0.30

TABLE 2. Uniform example with  $H_0 : \beta_2 \leq 0$  and  $H_1 : \beta_2 \geq \beta_2''$ . “true type I” and “true type II” errors are obtained by Monte-Carlo simulations under the additional assumption that  $Y \in \{0, 1\}$ .

<sup>8</sup>Without with this additional assumption, computing the type I and type II error probabilities would require an optimization procedure over the infinitely dimensional error structure.

The results show that our theoretical bounds allow us to reject the null hypothesis in a substantial range of values of  $\beta_2$ , even for small samples ( $n = 40, 60$ ).

The Bernoulli test performs better than the nonstandardized test when the covariates are symmetrically distributed around zero (in the step example when  $h/n = 0.5$  or in the uniform example) and the sample size is small or moderate. Each of the four probability bounds used in the construction of the nonstandardized test is binding for some range of parameters. This shows that dispensing with one of the tests or of one of the four bounds in the nonstandardized test would weaken our test.

The type I error probabilities obtained by Monte-Carlo simulations fall significantly below the nominal level, and are typically less than 1% instead of 5%. Our results only ensure that these probabilities are below the nominal level. The type II error probabilities obtained by Monte-Carlo simulations are also below the upper bounds of 0.5 that is guaranteed by our results, and range between 0.13 and 0.37 in the tables. This suggests that our methods are more powerful than our theoretical bounds indicates.

## 6. COMPARISON OF SIZE AND POWER OF DIFFERENT METHODS

In this section, in a practical example, we compare the type I error probabilities obtained by Monte-Carlo simulations between our method, the classical test, and White's test.

We use the same data structure as in the step example of section 5. We let  $h/n = 0.15$ , fix  $Y \in \{0, 1\}$  and test for  $H_0 : \beta_2 \leq 0.5$ . Because of the simplicity of the example, we believe that it is relevant to understand phenomena arising in common applications. The results are described in Table 3.

The left part of the table describes  $X$  (by  $n$  and  $h/n$ ), the nominal level of significance  $\alpha$  at which all tests are evaluated, the value  $\beta_2''$  at which our method guarantees a type II error probability of 0.5 as well as the corresponding method

$n$	$h/n$	$\alpha$	$\beta_2''$	test*	true type I error		
					exact	classical	robust
100	0.15	0.05	0.86	B	0.01	0.33	0.31
300	0.15	0.05	0.73	B	0.01	0.28	0.26
1000	0.15	0.05	0.63	B	0.01	0.24	0.23

TABLE 3. Step example with  $Y \in \{0, 1\}$  and  $H_0 : \beta_2 \leq 0.5$ . All values 0.01 under “exact” represent empirical probabilities of less than 1%.

selected. The corresponding rejection probabilities obtained by Monte-Carlo simulations are reported in the right part of the table: “exact” for our test, “classical” for the classical test and “robust” for White’s robust test.

The results show that both the classical and White’s robust test are severely oversized even for sample size as large as 1000, with type I error probabilities of 24% and 23%, to be compared with the nominal level of 5%. On the other hand, our method has a type I error probability of only less than 1%, and is therefore undersized. These findings do not depend on the fact that all variables are binary valued, a continuity argument shows that similar type I error probabilities can arise when both  $X_{i2}$  and  $Y_i$  have full support on  $[0, 1]$ .

For more general error structures where  $Y \in [0, 1]$ , the type I errors probabilities reported in Table 3 are lower lower bounds on the true values. This allows to conclude that both the classical test and White’s robust test are severely oversized in this case as well. On the the other hand, our methods continue to guarantee a type I error probability below the nominal level in this setup as well.

## 7. EMPIRICAL APPLICATION

In this section we apply our methods to regressions from Duflo *et al.* (2011). We test the significance of parameters and provide 95% confidence intervals. When testing for significance, we rely on the exact test that guarantees a type II error probability below 0.5 for the largest range of parameters. Confidence intervals are derived by considering the set of parameters where we cannot reject the null hypothesis with the equi-tailed two-sided test with level 0.05.

In each of the regressions, the dependent variable is a Bernoulli random variable specifying whether or not a farmer has used fertilizers in a given season: season 1 for regressions 1-2, season 2 for regressions 3-4, and season 3 for regressions 5-6. The independent variable “safi season 1” indicates whether the farmer was offered or not a certain SAFI program in season 1; “starter kit” and “demo” indicate respectively whether the farmer received a starter kit or participated in a demonstration plot; “kit and demo” is an interaction variable between these last two variables; and “household” indicates whether the household used fertilizer previous to the treatment. Additional dummy variables control for the sixteen possible schools attended. Regressions 2, 4, and 6 include a number of controls (not reported here), including the farmer’s gender, whether the farmer’s home has mud walls, the number of years of education, and the farmer’s income in the past month.

The number of observations ranges from 626 to 902, the number of variables is 21 for regressions without extra control variables and 28 for those with them.

In all the tests of  $H_0 : \beta_j = 0$ , the nonstandardized test is selected. The same test is also selected for all confidence intervals, except for the household covariate in regressions 2 and 6 where the Bernoulli test is used.

Our results confirm the main findings of Duflo *et al.* (2011), which is that the SAFI program had a significant effect on fertilizer adoption in the same season. We also confirm the absence of a significant effect of SAFI on fertilizer adoption in future seasons (regressions 3-6). Significant parameters using the classical test remain significant with our exact methods, albeit slightly less so: two variables found to be significant at the 1% significance level with the classical test are significant only at the 5% level with our exact test, while other variables have the same range of significance with the classical test and with our method. The classical OLS t-test is not appropriate as errors are by assumption not homoskedastic. Not suprisingly, homoskedasticity is rejected at the 1% level. White’s method leads to stronger significance than the classical method or ours. Using White’s method, the demo variable is found to be highly significant in regressions 2-6; while in contrast neither

variable	model	test of $H_0 : \beta_j = 0$			95% confidence intervals		
		exact	classical	robust	exact	classical	robust
safi season 1	1	**	***	***	[0.00, 0.23]	[0.04, 0.19]	[0.04, 0.19]
starter kit	1	not	not	not	[-0.08, 0.20]	[-0.03, 0.15]	[-0.02, 0.14]
kit and demo	1	not	not	not	[-0.22, 0.17]	[-0.15, 0.10]	[-0.14, 0.10]
demo	1	not	not	not	[-0.99, 1.00]	[-0.61, 0.63]	[-0.46, 0.48]
household	1	***	***	***	[0.27, 0.47]	[0.30, 0.43]	[0.30, 0.44]
safi season 1	2	**	***	***	[0.02, 0.27]	[0.06, 0.22]	[0.06, 0.22]
starter kit	2	*	*	*	[-0.07, 0.23]	[-0.01, 0.17]	[-0.01, 0.17]
kit and demo	2	not	not	not	[-0.28, 0.16]	[-0.20, 0.07]	[-0.20, 0.07]
demo	2	not	not	***	[-0.96, 1.84]	[-0.42, 1.30]	[0.24, 0.64]
household	2	***	***	***	[0.19, 0.47]	[0.24, 0.39]	[0.24, 0.39]
safi season 1	3	not	not	not	[-0.12, 0.13]	[-0.08, 0.09]	[-0.08, 0.09]
starter kit	3	not	not	not	[-0.12, 0.17]	[-0.07, 0.12]	[-0.07, 0.12]
kit and demo	3	not	not	not	[-0.19, 0.23]	[-0.11, 0.16]	[-0.11, 0.16]
demo	3	not	not	***	[-1.03, 1.75]	[-0.55, 1.27]	[0.18, 0.55]
household	3	***	***	***	[0.21, 0.43]	[0.25, 0.39]	[0.24, 0.40]
safi season 1	4	not	not	not	[-0.13, 0.15]	[-0.08, 0.10]	[-0.08, 0.10]
starter kit	4	not	not	not	[-0.15, 0.16]	[-0.10, 0.11]	[-0.10, 0.10]
kit and demo	4	not	not	not	[-0.24, 0.23]	[-0.16, 0.15]	[-0.16, 0.15]
demo	4	not	not	***	[-0.94, 1.87]	[-0.45, 1.38]	[0.23, 0.69]
household	4	***	***	***	[0.16, 0.41]	[0.20, 0.37]	[0.19, 0.37]
safi season 1	5	not	not	not	[-0.11, 0.12]	[-0.07, 0.08]	[-0.07, 0.08]
starter kit	5	not	not	not	[-0.14, 0.12]	[-0.10, 0.08]	[-0.10, 0.08]
kit and demo	5	not	not	not	[-0.19, 0.20]	[-0.13, 0.13]	[-0.13, 0.13]
demo	5	not	not	***	[-0.63, 1.36]	[-0.30, 1.02]	[0.10, 0.62]
household	5	***	***	***	[0.18, 0.38]	[0.21, 0.35]	[0.21, 0.35]
safi season 1	6	not	not	not	[-0.12, 0.14]	[-0.07, 0.09]	[-0.07, 0.09]
starter kit	6	not	not	not	[-0.18, 0.12]	[-0.13, 0.07]	[-0.12, 0.07]
kit and demo	6	not	not	not	[-0.25, 0.18]	[-0.17, 0.11]	[-0.17, 0.11]
demo	6	not	not	***	[-0.96, 1.84]	[-0.48, 1.35]	[0.22, 0.65]
household	6	***	***	***	[0.13, 0.38]	[0.17, 0.33]	[0.17, 0.33]

TABLE 4. Comparison of tests and confidence intervals: exact for our method, classical for the classical test, robust for White’s robust method. Model indicates the regression number. Significance levels: \*\*\* for 1%, \*\* for 5%, \* for 10%, and *not* for no significance at 10%.

our method nor the classical method finds this variable to be significant. However, a Monte-Carlo simulation conducted separately shows that White’s test applied to this variable reports significance at the 1% level with probability as high as 62% under the null hypothesis.<sup>9</sup> This casts a severe doubt on the appropriateness of the use of White’s method on this dataset.

The median increase in the size of confidence intervals using our exact method is about 50 percent compared to the classical or White’s method. This seems a moderate price to pay for exactness.

<sup>9</sup>The procedure looks for the vector  $\beta$  of parameters compatible with  $H_0$  that maximizes the probability of rejection. For every  $\beta$ , the noise structure is entirely specified by the fact that the dependent variable takes only the values 0 or 1.

## 8. RELAXING ASSUMPTIONS AND FURTHER TESTS

**8.1. Relaxing Assumptions on Errors.** We now discuss some relaxations of assumptions (iii) and (iv) in Section 2.

Assumption (iii) states that errors are independent conditionally on  $X$ . For the bound based on Cantelli (1910)'s inequality, the weaker condition of pairwise orthogonality  $\mathbf{E}(\varepsilon_i \varepsilon_j | X) = 0$  for  $i \neq j$  is sufficient. The inequality of Bhattacharyya (1987) relies on fourth moments of  $\hat{\beta}_j$ ; in order to use it, we need only to require that  $\mathbf{E}(\varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l | X) = 0$  if  $i \notin \{j, k, l\}$ . Hoeffding's inequality holds for Markov chains (Hoeffding, 1963, p.18); we can use it if we assume only that  $\mathbf{E}(\varepsilon_{j+1} | \varepsilon_1, \dots, \varepsilon_j, X) = 0$  for  $j = 1, \dots, n-1$ . We cannot, however, relax conditional independence when using the Berry-Esseen inequality or in the derivation of the error bounds for the Bernoulli test. Indeed, both the Berry-Esseen inequality and of the result of Hoeffding (1956) explicitly require independent random variables.

The assumption (iv) that the dependent variables are bounded (i.e.,  $\Pr(Y_i \in [\omega, \omega']) = 1$ ) can be relaxed in several ways. The methods presented can be adapted to the case in which the bounds depend both on  $X$  and on  $i$ , i.e., for every  $X$ , there exists  $(\omega_{1i})_i$  and  $(\omega_{2i})_i$  such that  $\Pr(\varepsilon_i \in [\omega_{1i}, \omega_{2i}] | X = x) = 1$  holds for every  $i$ . Alternatively, one can assume a bound on the variance of the noise terms. One can easily adapt the nonstandardized test to this case using the inequalities from Cantelli (1910), Hoeffding (1963), and Bhattacharyya (1987).

**8.2. Further tests.** The tests we have introduced are one-tailed. It is straightforward to construct from them a exact two-tailed tests from the one-tailed tests: the two-tailed test rejects  $H_0 : \beta_j = \beta_0$  at level  $\alpha$  if either of the two one-tailed tests  $H_0 : \beta_j \geq \beta_0$  or  $H_0 : \beta_j \leq \beta_0$  is rejected at level  $\alpha/2$ .

To construct a  $100 \cdot (1 - \alpha)$  % confidence interval, we consider of all those values of  $\bar{\beta}_j$  such that the null hypothesis  $H_0 : \beta_j = \bar{\beta}_j$  cannot be rejected at level  $\alpha$ .

Finally, it is desirable in several applications to test for multiple linear restrictions. Although simple tests exact for multiple restrictions can be easily derived from our one and two-tailed tests, such procedures can lack power. We leave the

question of constructing appropriate and powerful enough exact test for multiple restrictions to further research.

## 9. CONCLUSION

This paper introduces finite-sample methods that are exact in the sense that they do not rely on assumptions about the noise terms beyond independence. These tests perform well even in small sample sizes ( $n = 40, 60$ ). They are powerful enough to allow practical conclusions to be drawn when they are applied to independently collected empirical data.

The nonstandardized test relies on a selection of probabilistic bounds. Improvements of these bounds would lead to an improved test. We note, however, that when we conducted a thorough, albeit nonexhaustive, examination of bounds derived from a series of known inequalities such as those from Benktus (2004), Bercu and Touati (2008), Bernstein (1946), Pinelis (2007), and Xia (2008), these bounds did not result in any improvement of our method.

## REFERENCES

- R. R. Bahadur and L. J. Savage. The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, 27:1115–1122, 1956.
- V. Benktus. On hoeffding’s inequalities. *Annals of Probability*, 32:1650–1673, 2004.
- B. Bercu and T. Touati. Exponential inequalities for self-normalized martingales with applications. *The Annals of Applied Probability*, 18:1848–1869, 2008.
- S.N. Bernstein. *The Theory of Probabilities*. Gastehizdat Publishing House, Moscow, 1946.
- A. C. Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49:122–136, 1941.
- B. B. Bhattacharyya. One sided Chebyshev inequality when the first four moments are known. *Communications in Statistics - Theory and Methods*, pages 2789–2791, 1987.

- M.V. Boldin, G. Simonova, and Y.N. Tyurin. Sign-based methods in linear statistical models. In *Translations of Mathematical Monographs*, volume 162. American Mathematical Society, 1997.
- F.P. Cantelli. Intorno ad un teorema fondamentale della teoria del rischio. *Bolletino dell' Associazione degli Attuari Italiani*, 1910.
- V. Chernozhukov, C. Hansen, and M. Jansson. Finite sample inference for quantile regression models. *Journal of Econometrics*, 152:93–103, 2009.
- E. Coudin and J.-M. Dufour. Finite-sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form. *Econometrics Journal*, 12:S19–S49, 2009.
- E. Duflo, M. Kremer, and J. Robinson. Nudging farmers to use fertilizer: theory and experimental evidence from kenya. *American Economic Review*, 101:2350–2390, 2011.
- J.-M. Dufour and M. Hallin. Improved eaton bounds for linear combinations of bounded random variables, with statistical applications. *Journal of the American Statistical Association*, 88:1026–1033, 1993.
- J.-M. Dufour and A. Taamouti. Exact optimal inference in regression models under heteroskedasticity and non-normality of unknown form. *Computational Statistics and Data Analysis*, 54:2532–2553, 2010.
- J.-M. Dufour and O. Torres. Markovian processes, two-sided autoregressions and finite-sample inference for stationary and nonstationary autoregressive processes. *Journal of Econometrics*, 99:255–289, 2000.
- J.-M. Dufour. Identification, weak instruments, and statistical inference in econometrics. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 36:767–808, 2003.
- C.-G. Esseen. On the liapounoff limit of error in the theory of probability. *Arkiv för Matematik, Astronomi och Fysik*, A28:1–19, 1942.
- W. Hoeffding. On the distribution of the number of successes in independent trials. *The Annals of Mathematical Statistics*, 27:713–721, 1956.

- W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- P. T. Larsen. Goldman pays the price of being big. *Financial Times*, August 13, 2007.
- E.P. Lehmann and J.P. Romano. *Testing statistical hypotheses*. Springer, New-York, USA, third edition, 2005.
- J. Neyman and E.S Pearson. On the problem of two samples. In *Joint Statistical Papers*. Cambridge University Press, 1930.
- I. Pinelis. Exact inequalities for sums of asymmetric random variables, with applications. *Probability Theory and Related Fields*, 139:605–635, 2007.
- K. Schlag. Designing non-parametric estimates and tests for means. Economics Working Papers ECO 2006/26, European University Institute, Florence, 2006.
- K. Schlag. Exact tests for correlation and for the slope in simple linear regressions without making assumptions. working paper 1097, Universitat Pompeu Fabra, Barcelona, 2008.
- K. Schlag. A new method for constructing exact tests without making any assumptions. working paper 1109, Universitat Pompeu Fabra, Barcelona, 2008.
- I. G. Shevtsova. An improvement of convergence rate estimates in the lyapunov theorem. *Doklady Mathematics*, 82:862–864, 2010.
- H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–838, 1980.
- Y. Xia. Two refinements of the chernoff bound for the sum of nonidentical bernoulli random variables. *Statistics and Probability Letters*, 78:1557–1559, 2008.
- F. Yates. Contingency tables involving small numbers and the  $\chi^2$  test. *Supplement to the Journal of the Royal Statistical Society*, 1:217–235, 1934.

## APPENDIX A. PROOF OF THEOREM 1

The proof of Theorem 1 is obtained by combining a bound on the variance of  $\hat{\beta}_j$  (Lemma 3) with bounds on the deviation of  $\hat{\beta}_j$  from its mean provided by Propositions 4, 5, 7 and 8.

**A.1. Bound on the variance of  $\hat{\beta}_j$ .** We let  $\sigma_{\beta_j}^2 = \text{Var}(\hat{\beta}_j)$  represent the variance of  $\hat{\beta}_j$ .

**Lemma 3.**

$$\sigma_{\beta_j}^2 \leq \bar{\sigma}_{\beta_j}^2$$

*Proof.* For a given mean of  $Y_i$ ,  $\text{Var}(Y_i)$  is maximized when  $Y_i$  is a Bernoulli random variable taking the values  $\omega$  and  $\omega + 1$ :

$$\text{Var}(Y_i) \leq \mathbf{E}(Y_i - \omega)\mathbf{E}(\omega + 1 - Y_i) = (X_i\beta - \omega)(\omega + 1 - X_i\beta).$$

Since  $\sigma_{\beta_j}^2 = \sum \tau_{ji}^2 \text{Var}(Y_i)$ ,

$$\begin{aligned} \sigma_{\beta_j}^2 &\leq \sum_i \tau_{ji}^2 (X_i\beta - \omega)(\omega + 1 - X_i\beta) \\ &\leq \max_{z \in \mathbb{R}^m} \left\{ \sum_i \tau_{ji}^2 (X_i z - \omega)(\omega + 1 - X_i z) : z_j = \beta_j, Xz \in [\omega, \omega + 1]^n \right\} \\ &= \bar{\sigma}_{\beta_j}^2. \end{aligned}$$

■

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

$$P(Z - \mathbf{E}Z \geq k\sigma) \leq \frac{1}{1 + k^2}.$$

We rely on Cantelli's inequality to bound  $P(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t})$  using  $\varphi_C$ .

**Proposition 4.** (1) For  $\bar{t} > 0$  and  $\beta_j \leq \bar{\beta}_j$ ,

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

(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  $\beta_j \leq \bar{\beta}_j$ , by applying Cantelli's inequality to  $\hat{\beta}$  we obtain

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

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

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

which is point 2. Point 3 is immediate. ■

**A.3. Bhattacharyya.** Consider a random variable  $Z$  with  $\mathbf{E}Z = 0$ , let  $\sigma^2 = \text{Var}(Z)$ ,  $\gamma_1 = \frac{\mathbf{E}Z^3}{\sigma^3}$ , and  $\gamma_2 = \frac{\mathbf{E}Z^4}{\sigma^4}$ . Bhattacharyya (1987)'s inequality states that if  $k^2 - k\gamma_1 - 1 > 0$  then

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

Relying on this inequality we derive:

**Proposition 5.** (1) For  $\bar{t} > 0$  and  $\beta_j \leq \bar{\beta}_j$ ,

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

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

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

(3)  $\varphi_{Bh}$  is increasing in  $\sigma$  and decreasing in  $t$ .

Before applying Bhattacharyya's inequality to  $Z = \hat{\beta}_j - \beta_j$  we bound the corresponding values of  $\gamma_1 = \frac{\mathbf{E}(\hat{\beta}_j - \beta_j)^3}{\sigma_{\beta_j}^3}$  and  $\gamma_2 = \frac{\mathbf{E}(\hat{\beta}_j - \beta_j)^4}{\sigma_{\beta_j}^4}$ .

**Lemma 6.**

$$(3) \quad \frac{\mathbf{E}(\hat{\beta}_j - \beta_j)^3}{\sigma_{\beta_j}^3} \leq \frac{\|\tau_j\|_\infty}{\sigma_{\beta_j}}$$

and

$$(4) \quad \frac{\mathbf{E}(\hat{\beta}_j - \beta_j)^4}{\sigma_{\beta_j}^4} \leq 4.$$

*Proof.* Using the polynomial expansion, and  $\mathbf{E}(\varepsilon_i) = 0$  for every  $i$ , we obtain

$$\mathbf{E}(\hat{\beta}_j - \beta_j)^3 = \mathbf{E}\left(\sum_i \tau_{ji}(Y_i - X_i\beta)\right)^3 = \sum_i \tau_{ji}^3 \mathbf{E}(\varepsilon_i^3).$$

Since  $|\varepsilon_i| \leq 1$ , we have

$$\gamma_1 = \frac{\mathbf{E}(\hat{\beta}_j - \beta_j)^3}{\sigma_{\beta_j}^3} = \frac{\sum_i \tau_{ji}^3 \mathbf{E}(\varepsilon_i^3)}{\sigma_{\beta_j}^3} \leq \frac{\|\tau_j\|_\infty \sum_i \tau_{ji}^2 \mathbf{E}(\varepsilon_i^2)}{\sigma_{\beta_j}^3} = \frac{\|\tau_j\|_\infty}{\sigma_{\beta_j}}.$$

Using the polynomial expansion again, we get

$$\mathbf{E}(\hat{\beta}_j - \beta_j)^4 = \sum_i \tau_{ji}^4 \mathbf{E}(\varepsilon_i^4) + 3 \sum_{i \neq k} \tau_{ji}^2 \mathbf{E}(\varepsilon_i^2) \tau_{jk}^2 \mathbf{E}(\varepsilon_k^2)$$

and

$$\left(\sum_i \tau_{ji}^2 \mathbf{E}(\varepsilon_i^2)\right)^2 = \sum_i \tau_{ji}^4 \mathbf{E}(\varepsilon_i^2)^2 + \sum_{i \neq k} \tau_{ji}^2 \mathbf{E}(\varepsilon_i^2) \tau_{jk}^2 \mathbf{E}(\varepsilon_k^2).$$

From this we derive

$$\mathbf{E}(\hat{\beta}_j - \beta_j)^4 = 3 \left(\sum_i \tau_{ji}^2 \mathbf{E}(\varepsilon_i^2)\right)^2 + \sum_i \tau_{ji}^4 \mathbf{E}(\varepsilon_i^4) - 3 \sum_i \tau_{ji}^4 \mathbf{E}(\varepsilon_i^2)^2.$$

Using the Cauchy-Schwarz inequality twice we obtain

$$\begin{aligned} \sum_i \tau_{ji}^4 \mathbf{E}(\varepsilon_i^4) &= \int \sum_i \tau_{ji}^4 \varepsilon_i^4 dP \leq \int \left( \sum_i \tau_{ji}^2 \varepsilon_i^2 \right)^2 dP \\ &\leq \left( \int \left( \sum_i \tau_{ji}^2 \varepsilon_i^2 \right) dP \right)^2 = \left( \sum_i \tau_{ji}^2 \mathbf{E} \varepsilon_i^2 \right)^2 \end{aligned}$$

and hence

$$\gamma_2 = \frac{\mathbf{E}(\hat{\beta}_j - \beta_j)^4}{\sigma_{\beta_j}^4} \leq 4.$$

■

*Proof of Proposition 5.* For the proof of point 1, we need only to consider the case where  $\frac{\bar{t}^2}{\sigma_{\beta_j}^2} - \frac{\bar{t}\|\tau_j\|_\infty}{\sigma_{\beta_j}^2} - 1 > 0$ , in which we can apply Bhattacharyya's inequality to  $\hat{\beta}_j - \beta_j$  and use (4):

$$\begin{aligned} P(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}) &\leq P(\hat{\beta}_j - \beta_j \geq \bar{t}) \\ &\leq \frac{\gamma_2 - \gamma_1^2 - 1}{(\gamma_2 - \gamma_1^2 - 1) \left( 1 + \left( \frac{\bar{t}}{\sigma_{\beta_j}} \right)^2 \right) + \left( \left( \frac{\bar{t}}{\sigma_{\beta_j}} \right)^2 - \left( \frac{\bar{t}}{\sigma_{\beta_j}} \right) \gamma_1 - 1 \right)^2} \\ (5) \quad &\leq \frac{3 - \gamma_1^2}{(3 - \gamma_1^2) \left( 1 + \frac{\bar{t}^2}{\sigma_{\beta_j}^2} \right) + \left( \frac{\bar{t}^2}{\sigma_{\beta_j}^2} - \frac{\bar{t}}{\sigma_{\beta_j}} \gamma_1 - 1 \right)^2}. \end{aligned}$$

We then obtain point 1 by maximizing (5), 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 using (3). The proof of point 2 is similar, and point 3 comes from the fact that both functionals defining  $\varphi_{Bh}$  when  $\frac{t^2}{\sigma^2} - \frac{t\|\tau_j\|_\infty}{\sigma^2} - 1 > 0$  are increasing in  $\sigma$  and decreasing in  $t$ . ■

**A.4. Hoeffding.** We recall an inequality due to Hoeffding (1963, Theorem 2). Let  $(Z_i)_{i=1}^n$  be independent random variables with  $Z_i \in [a_i, b_i]$ , and  $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$ . For  $\bar{t} > 0$ ,

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

Relying on Hoeffding's inequality we show:

**Proposition 7.** (1) For  $\bar{t} > 0$  and  $\beta_j \leq \bar{\beta}_j$ ,

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

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

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

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

*Proof.* We apply Hoeffding's inequality to  $(Z_i)_i$  where  $Z_i = n\tau_{ji}Y_i$ . So  $Z_i \in [n\tau_{ji}\omega, n\tau_{ji}(\omega + 1)]$  for  $\tau_{ji} \geq 0$  and  $Z_i \in [n\tau_{ji}(\omega + 1), \omega]$  for  $\tau_{ji} < 0$ . For  $\beta_j \leq \bar{\beta}_j$ :

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

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

**A.5. Berry-Esseen.** We recall the Berry-Esseen inequality (Berry, 1941; Esseen, 1942) with the constant as derived by Shevtsova (2010). Let  $(Z_i)_{1 \leq i \leq N}$  be a family of independent random variables with  $\text{Var}(Z_i) = \sigma_i^2$ . For  $\bar{u} \in \mathbb{R}$ ,

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

where  $A = 0.56$ . Using the Berry-Esseen inequality, we show the following proposition:

**Proposition 8.** (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 Proposition 8 is to apply Berry-Esseen's inequality to the random variables  $Z_i = \tau_{ji}Y_i$ . However, a difficulty arises from the fact that the right hand side of Berry-Esseen's inequality is unbounded as 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  $(Z_i)_{1 \leq i \leq N}$  to guarantee such a lower bound. We eliminate this noise in a later step.

**Lemma 9.** *Let  $w > 0, \bar{u} \in \mathbb{R}$ . With  $Z \sim \mathcal{N}(0, w^2)$  independent of  $(Y_i)_i$ , and*

$$R(w) = \frac{\sum_i |\tau_{ji}|^3 \mathbf{E} |Y_i - \mathbf{E}Y_i|^3}{(\sigma_{\beta_j}^2 + w^2)^{3/2}},$$

we have

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

*Proof.* We apply Berry-Esseen's inequality to the family of independent random variables  $Z_1, \dots, Z_{n+N}$  where  $Z_i = \tau_{ji}Y_i$  for  $i \leq n$  and  $Z_i \sim \mathcal{N}\left(0, \frac{w^2}{N}\right)$  for  $n+1 \leq i \leq n+N$ . We note that  $Z$  has the same distribution as  $\sum_{t=1}^{n+N} Z_t$ . Let  $\delta \sim \mathcal{N}(0, 1)$ . The Berry-Esseen inequality applied to  $\sum_{t=1}^{n+N} Z_t$  shows:

$$\begin{aligned} P\left(\frac{\hat{\beta}_j - \beta_j + Z}{\sqrt{\sigma_{\beta_j}^2 + w^2}} \geq \bar{u}\right) &= 1 - P\left(\frac{\sum_{i=1}^{n+N} (Z_i - \mathbf{E}Z_i)}{\sqrt{\sum_{i=1}^{n+N} \sigma_{Z_i}^2}} \leq \bar{u}\right) \\ &\leq 1 - \Phi(\bar{u}) + A \frac{\sum_{i=1}^n |\tau_{ji}|^3 \mathbf{E} |Y_i - \mathbf{E}Z_i|^3 + N \left(\frac{w}{\sqrt{N}}\right)^3 \mathbf{E} |\delta|^3}{(\sigma_{\beta_j}^2 + w^2)^{3/2}}. \end{aligned}$$

As  $N \rightarrow \infty$  the right term decreases and converges to  $1 - \Phi(\bar{u}) + AR(w)$ , and the claim follows. ■

Next we use Lemma 9 to obtain a bound on the upper tail of  $\hat{\beta}_j - \beta_j$ .

**Lemma 10.** *For every  $b_1 \in \mathbb{R}$  and  $w > 0$ ,*

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

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

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

Applying this to  $\bar{u} = \frac{\bar{t} - b_1}{\sqrt{\sigma_{\beta_j}^2 + w^2}}$  and combining with Lemma 9 yields the result. ■

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

**Lemma 11.**

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

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

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

we derive

$$R(w) = \frac{\sum_i |\tau_{ji}|^3 \mathbf{E}|Y_i - \mathbf{E}Y_i|^3}{\left(\sum_i \tau_{ji}^2 \mathbf{E}(Y_i - \mathbf{E}Y_i)^2 + w^2\right)^{3/2}} \leq \frac{\|\tau_j\|_\infty \sum_i |\tau_{ji}|^2 \mathbf{E}(Y_i - \mathbf{E}Y_i)^2}{\left(\sum_i \tau_{ji}^2 \mathbf{E}(Y_i - \mathbf{E}Y_i)^2 + w^2\right)^{3/2}} \leq \frac{2\|\tau_j\|_\infty}{\sqrt{27}w}.$$

■

*Proof of Proposition 8.* Using Lemmata 10 and 11, we obtain that for  $\beta_j \leq \bar{\beta}_j$  and for every  $b_1 \in \mathbb{R}, w > 0$ ,

$$\begin{aligned} P\left(\hat{\beta}_j - \bar{\beta}_j \geq \bar{t}\right) &\leq P\left(\hat{\beta}_j - \beta_j \geq \bar{t}\right) \\ &\leq \frac{1 - \Phi\left(\frac{\bar{t} - b_1}{\sqrt{\sigma_{\beta_j}^2 + w^2}}\right) + A \frac{2\|\tau_j\|_\infty}{\sqrt{27}w}}{\Phi(b_1/w)}. \end{aligned}$$

Therefore,

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

which is point 1. For point 2, we apply point 1 to  $Y' = (\omega + 1)1_n - Y$  where  $1_n \in \mathbb{R}^n$  is such that  $1_{n,i} = 1$  for every  $i$ . For  $\beta_j$  such that  $\beta_j > \bar{\beta}_j + \bar{t}$ ,

$$\begin{aligned} P(\hat{\beta}_j - \bar{\beta}_j < \bar{t}) &\leq P(\tau'_j Y - \bar{\beta}_j \leq \bar{t}) \\ &= P(\tau'_j((\omega + 1)1_n - Y) - (\tau'_j(\omega + 1)1_n - \beta_j) \geq \beta_j - \bar{\beta}_j - \bar{t}) \\ &\leq \varphi_{BE}(\sigma_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}). \end{aligned}$$

Point 3 is immediate. ■

#### APPENDIX B. PROOFS FOR SECTION 4

**Proposition 12.** (1) If  $\beta_j \leq \bar{\beta}_j$  then  $\lambda P(n\bar{W} \geq \bar{k} - 1) + (1 - \lambda)P(n\bar{W} \geq \bar{k}) \leq \theta\alpha$ .

(2) If  $p_{\beta_j} > \bar{k}/n$  then

$$\lambda P(n\bar{W} \geq \bar{k} - 1) + (1 - \lambda)P(n\bar{W} \geq \bar{k}) \geq \lambda B(\bar{k} - 1, p_{\beta_j}) + (1 - \lambda)B(\bar{k}, p_{\beta_j}).$$

Consider a randomized test that rejects the null hypothesis with probability equal to 1 if  $n\bar{W} \geq \bar{k}$ , equal to  $\lambda$  if  $n\bar{W} = \bar{k} - 1$ , and equal to 0 if  $n\bar{W} < \bar{k} - 1$ . Point 1 shows that the type I error probability of this test is bounded by  $\theta\alpha$ . A bound on the type II error probability is given by point 2.<sup>10</sup>

*Proof of Proposition 12.* Theorem 5 in Hoeffding (1956) shows that, if  $\bar{k} \geq n\mathbf{E}\bar{W}$ , then  $P(n\bar{W} \geq \bar{k}) \leq B(\bar{k}, \mathbf{E}\bar{W})$ . Similarly, if  $\bar{k} < n\mathbf{E}\bar{W}$ , then  $P(n\bar{W} \geq \bar{k}) \geq B(\bar{k}, \mathbf{E}\bar{W})$ . Since  $\bar{k} - 1 > n\bar{p} \geq n\mathbf{E}\bar{W}$ , we have

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

<sup>10</sup>Note that this randomized test is the most powerful test (see, e.g., Lehmann and Romano, 2005, Example 3.4.2) for testing  $p \leq \bar{p}$  against  $p > \bar{p}$  at level  $\theta\alpha$  given  $n$  i.i.d. observations.

where the first inequality comes from Hoeffding (1956)'s result, and the second one from the fact that  $B(\bar{k}, \bar{p})$  is increasing in  $\bar{p}$ . Hence point 1. Since  $\mathbf{E}\bar{W} > \bar{k}/n$ , we also have

$$\lambda P(n\bar{W} \geq \bar{k} - 1) + (1 - \lambda)P(n\bar{W} \geq \bar{k}) \geq \lambda B(\bar{k} - 1, \mathbf{E}\bar{W}) + (1 - \lambda)B(\bar{k}, \mathbf{E}\bar{W}),$$

which is point 2. ■

*Proof of Theorem 2.* Let  $\beta_j \leq \bar{\beta}_j$ . From point 1 of Proposition 12, the expectation of the non-negative random variable  $R = \lambda \mathbf{1}_{n\bar{W} \geq k-1} + (1 - \lambda) \mathbf{1}_{n\bar{W} \geq k}$  is bounded by  $\theta\alpha$ . Markov's inequality shows

$$\lambda P(n\bar{W} \geq k - 1) + (1 - \lambda)P(n\bar{W} \geq k) = P(R \geq \theta) \leq \frac{\mathbf{E}R}{\theta} \leq \alpha,$$

which is the desired bound on the type I error probability. We now apply Markov's inequality to  $1 - R$ :

$$P(R < \theta) = P(1 - R > 1 - \theta) \leq \frac{1 - \mathbf{E}R}{1 - \theta},$$

which together with point 2 of Proposition 12 implies the stated bound on type II error probabilities. ■