WORKING PAPERS

Finite Sample Exact tests for Linear Regressions

Oliver Gossner
Karl Schlag

Februar 2012

Working Paper No: 1201

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

Finite Sample Exact Tests for Linear Regressions

Olivier Gossner and Karl Schlag*

Paris School of Economics, Paris, France
University of Vienna, Department of Economics, Austria,
e-mail: gossner@pse.ens.fr; karl.schlag@univie.ac.at

Abstract: We introduce tests for finite sample multivariate linear regressions with heteroskedastic errors that have mean zero. We assume bounds on endogenous variables but do not make additional assumptions on errors. The tests are exact, i.e., they have guaranteed type I error probabilities. We provide bounds on probability of type II errors, and apply the tests to empirical data.

AMS 2000 subject classifications: Primary 62J05.

Keywords and phrases: Nonparametric linear regression, exact test, heteroskedasticity.

1. Introduction

A common problem in linear regressions is to find a test that guarantees a certain type I error probability when error terms are not normally and identically distributed. Ideally, such a test should guarantee a type I error probability under no assumption the on error terms except for them being independent. It should also be sufficiently powerful to reject the null hypothesis often enough in practice.

This paper introduces two tests for linear regressions and examines their powers. These tests are exact under no assumption but independence, i.e., they guarantee

*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.

type I error probabilities below the level independently of the noise structure. They can be used to derive exact confidence intervals. We also provide bounds on the type II error probabilities of these tests.

The tests require the knowledge of a bounded range for the dependent variable. In practice, such bounded variables are plenty, and include test scores, percentages, as well as indicator variables. The results of Bahadur and Savage (1956) and Dufour (2003) show that without assumptions on the error structure and without such bounds, the only exact tests are trivial.

Starting with White (1980), several asymptotic tests have been proposed (see e.g. MacKinnon and White, 1985; Davidson and MacKinnon, 1993). It has already be pointed out (Greene, 2002, chapter 11) that the use of asymptotic tests for finite samples can be problematic as these tests are not exact. An exact test for finite samples was provided by Schlag (2008a) for simple linear regressions, but their construction remains an open question for general linear regressions.

A branch of literature initiated by Dufour and Hallin (1993) develops exact finite sample tests when errors have median zero (see also Boldin, Simonova, and Tyurin, 1997; Chernozhukov and Jansson, 2009; Coudin and Dufour, 2009; Dufour and Taamouti, 2010). Our work complements this line of research, and is, to the best of our knowledge, the first to develop exact methods in the case of errors with mean zero and more than one non-constant covariate.

Our two tests are referred to as “Non-Standardized” and “Bernoulli”. We briefly summarize their construction. 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 Non-Standardized 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.

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). Each term of the linear combination that constitutes the unbiased estimator is transformed into a Bernoulli random variable. We then test the mean of the obtained family of 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 errors for each of these tests. These bounds can be used to select among the tests, and to choose the free parameters used in the definition of each test.

We investigate the performance of our tests 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. $n = 40$).

We also implement our tests and compute confidence intervals using the empirical data from Duflo, Kremer, and Robinson (2011). The results show that, compared to the standard OLS regresssion analysis, the losses of significance of our exact method are moderate, and the confidence intervals are in most cases inflated by a factor of about 50%. The software used to implement the tests is freely downloadable from the authors' webpages.

The paper is organized as follows. Section 2 introduces the model. Sections 3 and 4 present the Non-Standardized test and the Bernoulli test. In Section 5, we examine their efficiency using numerical examples. Section 6 implements the tests in empirical data. Relaxations of the assumptions on the underlying data generating process are discussed in Section 7. We conclude in Section 8. All proofs are presented in the appendix.

2. Linear regression

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

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

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 non random and known ex-ante to the statistician is a special case. We assume (i) *strict exogeneity*: $\mathbf{E}(\varepsilon|X) = 0$, (ii) almost surely no *multi-colinearity*: X has rank m with probability 1, and (iii) *conditional independence of errors*: $(\varepsilon_i)_i$ are independent conditional on X. Finally, we assume (iv) *boundedness of the endogenous variable*: there exist ω and ω' with $\omega < \omega'$ such that $P(Y_i \in [\omega, \omega']) = 1$ for $i = 1, ..., n$. In particular, (iv) implies that $X_i\beta \in [\omega, \omega']$ almost surely and ensures existence of all moments of ε_i for $i = 1, ..., n$. We assume wlog. that $\omega' = \omega + 1$, other cases reduce to this one by dividing each side of the linear equation equation (1) by $\omega' - \omega$. We relax (iii) and (iv) in Section 7. We do not make further assumptions on error terms such as $\mathrm{Var}(\varepsilon_i) > 0$ or homoskedasticity.

We present exact tests at the level of significance α for the one-sided hypotheses $H_0 : \beta_j \leq \bar{\beta}_j$ against $H_1 : \beta_j > \bar{\beta}_j$ where $\bar{\beta}_j \in \mathbb{R}$. Tests of $H_0 : \beta_j \geq \bar{\beta}_j$, $H_0 : \beta_j = \bar{\beta}_j$ and confidence intervals can be derived easily. *Exact* means that the probability of a type I error of the test is proven to be at most α for any random vectors (X, ε) that satisfy (i)-(iv). In particular, bounds on the probabilities of type I errors are guaranteed for every given sample size and do not rely on asymptotic theory.

3. The Non-Standardized 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 τ_j, $\hat{\beta}_j = \tau_j' Y$ is an unbiased estimate of β_j. We present a test given τ_j, and later discuss the choice of τ_j. We let $\| \|_\infty$ denote the supremum norm, and $\| \|$ denote the Euclidian norm, and Φ denotes the cumulative normal distribution.

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

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

and

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

The tests use the following bound (see Lemma 1 in the Appendix) on the variance of $\hat{\beta}_j$ as a function of β_j

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

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

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

It is easily checked that φ is continuously decreasing in t and $\lim_{t\to\infty} \varphi(\bar{\sigma}_{0,\beta_j}, t, \tau_j) = 0$. Hence, for $0 < \alpha < 1$, there is minimal value $\bar{t}_N$ such that $\varphi(\bar{\sigma}_{0,\beta_j}, \bar{t}_N, \tau_j) \le \alpha$. We define the Non-Standardized test as the one that rejects the null hypothesis when $\hat{\beta}_j - \bar{\beta}_j \ge \bar{t}_N$.

Theorem 1 *The Non-Standardized test has type I error probability bounded above by α, and type II error probability bounded above by $\varphi\left(\bar{\sigma}_{\beta_j}, \beta_j - \bar{\beta}_j - \bar{t}_N, \tau_j\right)$ for every $\beta_j \geq \bar{\beta}_j - \bar{t}_N$.*

To prove Theorem 1, we use inequalities due to Cantelli (1910), Bhattacharyya (1987), Hoeffding (1963) and Berry-Esseen (Berry, 1941; Esseen, 1942; Shevtsova, 2010) to prove that under the null hypothesis, $P(\hat{\beta}_j \geq \bar{\beta}_j + \bar{t}_N)$ is bounded above by $\varphi_C(\bar{\sigma}_{0,\beta_j}, \bar{t}_N)$,
$\varphi_{Bh}(\bar{\sigma}_{0,\beta_j}, \bar{t}_N, \tau_j)$, $\varphi_H(t, \bar{t}_N)$ and $\varphi_{BE}(\bar{\sigma}_{0,\beta_j}, \bar{t}_N)$ respectively. Combining these results yields the bounds on the probability of type I errors. The bounds on the type II error probability are obtained in a similar manner.

The test is called "Non-Standardized" since it relies on maximal bounds on the deviation of $\hat{\beta}_j$ from its mean and does not try to estimate the variance of $\hat{\beta}_j$ from the data (as the standard OLS test and White's test do).

In the definition of the Non-Standardized test, τ_j is any vector with the property that $X'\tau_j = e_j$. The bound on type II error probabilities of Theorem 1 can be used to select a vector of weights τ_j. In practice, the system of weights τ_j corresponding to the OLS estimator allows for a good performance of the test, as illustrated in Sections 5 and 6. It has the additional advantage that results are easily comparable to other tests based on the OLS estimate.

4. The Bernoulli test

As the Non-Standardized test, the Bernoulli test is built on a vector $\tau_j \in \mathbb{R}^n$ such that $X'\tau_j = e_j$, so that $\hat{\beta}_j = \tau_j' Y$ is an unbiased estimate of β_j. The test also depends on a 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 $[0, \|\tau_i\|_\infty]$ and on a parameter $\theta \in (0, 1)$. First we present the test for significance level α, then we discuss the choice of τ, d and θ.

As in Schlag (2006, 2008b), we reduce the problem of testing β_j against $\beta_{j,0}$ to testing the mean of a sequence of Bernoulli random variables. More precisely, consider a family (W_i) of independent Bernoulli random variables such that the probability of success of W_i is $(\tau_{ji}Y_i + d_i)/\|\tau_j\|_\infty$, and the conditions imposed on d ensure that these probabilities are in $[0, 1]$. The proportion of successes $\bar{W} = \sum_i W_i/n$ has expectation $p_{\beta_j} = \frac{\beta_j + \sum_i d_i}{n\|\tau_j\|_\infty}$, and $\bar{p} = p_{\bar{\beta}_j}$ is the maximum of p_{β_j} under the null hypothesis.

The Bernoulli test compares the tail distribution of $\bar{W}$ with the tail of the binomial distribution with parameters $(n, \bar{p})$. For $0 < p < 1$ and $k \in \{0, ..., n\}$, we thus let

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

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

$$\lambda P(n\bar{W} \geq \bar{k} - 1) + (1 - \lambda)P(n\bar{W} \geq \bar{k}) \geq \theta.$$

Theorem 2 *The Bernoulli test has type I error probability bounded above by α. If $p_{\beta_j} > \bar{k}/n$, the type II error probability is bounded above by*

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

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 bounds on the probability of type I and type II errors of the randomized test as a function of the binomial distribution with parameter $\bar{p}$.

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 α whenever the probability that the randomized test rejects the null hypothesis, at the significance level $\theta\alpha$, exceeds θ. We then bound the probability of type I and type II errors of the deterministic test.

As in the case of the Non-Standardized test, the bound on type II error probabilities of Theorem 2 can be used to select the parameters τ_j, d, and θ. In practical applications, good performance is attained when τ minimizes $\|\tau_j\|_\infty$, d is given by $d_i = \|\tau_j\|_\infty - \max\{\tau_{ji}\omega, \tau_{ji}(\omega+1)\}$ (note that this choice of d satisfies all required constraints), and θ is computed numerically to minimize the value of β_j for which our bounds guarantee a type II error probability below 0.5, as illustrated in Sections 5 and 6.

5. Numerical examples

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 α, we look for the minimal value of β_2 such that the type II error of the test is guaranteed to fall below 0.5. The tests are implemented with the choice of free parameters explained at the end of Sections 3 and 4.

In the first example, which we call the extreme example, the second covariate X_2 takes only the values -1 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. We assume $Y_i \in [0, 1]$ for every i, which constrains the values of β_2 to belong to $[-1/2, 1/2]$.

Table 1 presents results in the extreme example, and Table 2 presents results in the uniform example. We consider different values of the sample size n, and vary h/n and the significance level α in the extreme example. The column β_2 reports the minimal value of β_2 for which one of our tests is, using the bounds of Theorem 1 and 2, guaranteed to have a type II error probability below 0.5. We then indicate the bound on the value of the type II error at this value of β_2 for the Non-Standardized test, and for the Bernoulli test. For the Non-Standardized test, we also report, in parenthesis, which of the four bounds is binding when determining the threshold $\bar{t}_N$ used for rejecting the null hypothesis and when deriving the type II error bound at β_2: Cantelli (C), Bhattacharyya (Bh), Hoeffding (H) or Berry-Esseen (BE).

For instance, in the extreme example with $n = 40$, the Bernoulli test is selected for testing $H_0 : \beta_2 \leq 0$ at level 0.05. It guarantees a type II error probability below 0.5 for all $\beta_2 \geq 0.2$. In contrast, the Non-Standardized test can only guarantee a type II error probability below 0.94 for this set of parameters, it does this by deriving the threshold $\bar{t}_N$ with φ_H and the type II error probability bound using φ_{BE}.

n	h/n	α	β_2	NS		B
40	0.50	0.05	0.20	0.94	(H,BE)	**0.50**
40	0.25	0.05	0.30	**0.50**	(H,C)	0.54
40	0.25	0.10	0.11	**0.50**	(H,Bh)	1.0
100	0.50	0.05	0.13	0.84	(H,BE)	**0.50**
100	0.25	0.05	0.20	**0.50**	(H,C)	0.59
5000	0.50	0.05	0.02	0.59	(BE,BE)	**0.50**

TABLE 1
Extreme example

n	α	β_2	NS		B
60	0.05	0.32	0.78	(H,C)	**0.50**
500	0.05	0.11	0.63	(H,BE)	**0.50**
6000	0.05	0.03	**0.50**	(H,BE)	0.51

TABLE 2
Uniform example

Note that since the reported values of β_2 are based on Theorems 1 and 2, they are upper bounds on the minimal value of β_2 for which the type II error probability falls below 0.5. Note also that we make no claims that our selection of parameters is optimal, optimizing on these parameters can further improve the performance of the test.

We make a few observations based on these tables. The methods work sufficiently well to allow to reject the null in a substantial range of values of β_2 even for small samples ($n = 40, 60$). The Bernoulli test performs better than the Non-Standardized test when the covariates are symmetrically distributed around 0 (in the extreme 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 Non-Standardized test is binding for some range of parameters.

6. Empirical application

In this section we apply our methods to regressions from Duflo, Kremer, and Robinson (2011, Table 4 Panel A). Their objective is to understand, by means of a randomized experiment, whether farmers can be induced to use fertilizer with the so-called SAFI program. We apply our tests to investigate the robustness of their analysis to assumptions on errors.

In each of the six 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 variables "safi season 1" indicates whether or not the farmer was offered or not a certain SAFI program, "starter kit" and "demo" indicate whether the farmer received a starter kit or participated in a demonstration plot, and "kit and demo" is the interaction between these two variables. The "household" dummy variable indicates whether the household used fertilizer previous to the treatment. Additional dummy variables control for the 16 possible schools attended. Regressions 2, 4, 6 include a number of controls (non-reported), including the farmer's gender, whether home has mud walls, the number of years of education, and the 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.

We test the significance of parameters, and provide 95% confidence intervals. We use the specification of parameters as in the end of Sections 3 and 4, and rely on the exact test that guarantees 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. We report variable significance and confidence intervals from the OLS, White, and our method.

The OLS method used in Duflo, Kremer, and Robinson (2011) relies on the assumption on homoskedastic errors. However, this assumption is rejected at the 1% level in the data in all six models by the Breusch-Pagan test. White's method is robust to heteroskedastic errors, but is based on asymptotic theory. A Monte Carlo simulation shows that the demo variable, which is found to be significant by White's test at the 1% level in regressions 2-6, is rejected at this level with probability as large as 72% under the null hypothesis.

Our tests confirm the main findings of Duflo, Kremer, and Robinson (2011), which is the significant effect of the SAFI program on fertilizer adoption in the same season. This is a robust conclusion that is not based on any assumption on the error terms.

We also confirm the absence of a significant effect of SAFI on fertilizer adoption in future seasons (regressions 3-6). The loss of significance of parameters using our exact method is very mild compared to the OLS method: two variables found significant at the 1% significance level with OLS are only significant at the 5% with our exact test, other variables have in the same range of significance with OLS and with our method. This loss is somewhat higher compared to White's method, which finds the

variable	model	test of $H_0 : \beta_j = 0$ exact	t test	robust	95% confidence intervals exact	t test	robust
safi season 1	1	**	***	***	$[0.00, 0.23]$	$[0.04, 0.19]$	$[0.04, 0.19]$
starter kit	1	not	not	not	$[-0.07, 0.19]$	$[-0.03, 0.15]$	$[-0.02, 0.14]$
kit and demo	1	not	not	not	$[-0.22, 0.16]$	$[-0.15, 0.10]$	$[-0.14, 0.10]$
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	not	*	*	$[-0.07, 0.23]$	$[-0.01, 0.17]$	$[-0.01, 0.17]$
kit and demo	2	not	not	not	$[-0.27, 0.15]$	$[-0.20, 0.07]$	$[-0.20, 0.07]$
demo	2	not	not	***	$[-0.95, 1.84]$	$[-0.42, 1.30]$	$[0.24, 0.64]$
household	2	***	***	***	$[0.19, 0.43]$	$[0.24, 0.39]$	$[0.24, 0.39]$
safi season 1	3	not	not	not	$[-0.11, 0.13]$	$[-0.08, 0.09]$	$[-0.08, 0.09]$
starter kit	3	not	not	not	$[-0.12, 0.16]$	$[-0.07, 0.12]$	$[-0.07, 0.12]$
kit and demo	3	not	not	not	$[-0.18, 0.23]$	$[-0.11, 0.16]$	$[-0.11, 0.16]$
demo	3	not	not	***	$[-1.02, 1.75]$	$[-0.55, 1.27]$	$[0.18, 0.55]$
household	3	***	***	***	$[0.21, 0.42]$	$[0.25, 0.39]$	$[0.24, 0.40]$
safi season 1	4	not	not	not	$[-0.13, 0.14]$	$[-0.08, 0.10]$	$[-0.08, 0.10]$
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	$[-0.23, 0.22]$	$[-0.16, 0.15]$	$[-0.16, 0.15]$
demo	4	not	not	***	$[-0.94, 1.82]$	$[-0.45, 1.38]$	$[0.23, 0.69]$
household	4	***	***	***	$[0.16, 0.38]$	$[0.20, 0.37]$	$[0.19, 0.37]$
safi season 1	5	not	not	not	$[-0.10, 0.12]$	$[-0.07, 0.08]$	$[-0.07, 0.08]$
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	$[-0.18, 0.19]$	$[-0.13, 0.13]$	$[-0.13, 0.13]$
demo	5	not	not	***	$[-0.63, 1.35]$	$[-0.30, 1.02]$	$[0.10, 0.62]$
household	5	***	***	***	$[0.18, 0.38]$	$[0.21, 0.35]$	$[0.21, 0.35]$
safi season 1	6	not	not	not	$[-0.11, 0.13]$	$[-0.07, 0.09]$	$[-0.07, 0.09]$
starter kit	6	not	not	not	$[-0.17, 0.12]$	$[-0.13, 0.07]$	$[-0.12, 0.07]$
kit and demo	6	not	not	not	$[-0.24, 0.18]$	$[-0.17, 0.11]$	$[-0.17, 0.11]$
demo	6	not	not	***	$[-0.96, 1.81]$	$[-0.48, 1.35]$	$[0.22, 0.65]$
household	6	***	***	***	$[0.14, 0.36]$	$[0.17, 0.33]$	$[0.17, 0.33]$

TABLE 3

*Comparison of tests and confidence intervals: exact for our method, t-test for the standard OLS, robust for White's method. Model indicates the regression number. Significance levels: *** for 1%, ** for 5%, * for 10%, and not for no significance at 10%.*

demo variable to be highly significant in regressions 2-6, while neither our method or the OLS method find this variable to be significant. As mentioned above, Monte Carlo simulations cast doubts on the appropriateness of White's test for this variable.

The size of the confidence intervals using the exact method is typically inflated by 50% compared to OLS or White's method. This seems a moderate price to pay for exactness under no assumptions on error terms.

7. Relaxing assumptions on errors

We now discuss some relaxations of assumptions (iii) and (iv) in Section 2.

Assumption (iii) states that errors conditional on X are independent. For the bound based on Cantelli's inequality, the classic pairwise orthogonality condition $\mathbf{E}\left(\varepsilon_i \varepsilon_j | X\right) = 0$ for $i \neq j$ is sufficient. The inequality of Bhattacharyya relies on fourth moments of $\hat{\beta}_j$, accordingly we need to impose that $\mathbf{E}\left(\varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l | X\right) = 0$ if $i \notin \{j, k, l\}$. Hoeffding's inequality holds for Markov chains (Hoeffding, 1963, p. 18), the relevant condition here is that $\mathbf{E}\left(\varepsilon_{j+1} | \varepsilon_1, \ldots, \varepsilon_j, X\right) = 0$ for $j = 1, ..., n-1$. We however cannot relax conditional independence when using the Berry-Esseen inequality when deriving the Bernoulli test. The inequality of Berry-Esseen and the result of Hoeffding (1956) explicitly require independence of the random variables.

Assumption (iv) that the dependent variables are bounded, i.e. $P(Y_i \in [\omega, \omega']) = 1$ 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 $P(\varepsilon_i \in [\omega_{1i}, \omega_{2i}] \, | X = x) = 1$ holds every i. Alternatively, one can assume a bound on the variance of the noise terms. One can easily adapt the Non-Standardized test to this case using Cantelli, Hoeffding, and Bhattacharyya's inequalities. Note that 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.

8. Conclusion

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

The Non-Standardized test relies on a selection of probabilistic bounds. Improvements of these bounds would lead to an improved test. A thorough, yet non-exhaustive, examination of bounds derived from a series of known inequalities did not allow for any improvement over the ones used in this paper for the construction of one-tailed tests.

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.

Boldin, M., G. Simonova, and Y. Tyurin (1997): "Sign-based methods in linear statistical models," in *Translations of Mathematical Monographs*, vol. 162. American Mathematical Society.

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

Chernozhukov, V., C. H., and M. Jansson (2009): "Finite sample inference for quantile regression models," *Journal of Econometrics*, 152, 93–103.

Coudin, E., and J.-M. Dufour (2009): "Finite-sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form," *Econometrics Journal*, 12, S19–S49.

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

Duflo, E., M. Kremer, and J. Robinson (2011): "Nudging farmers to use fertilizer: theory and experimental evidence from Kenya," *American Economic Review*, 101, 2350–2390.

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.

Dufour, J.-M., and A. Taamouti (2010): "Exact optimal inference in regression models under heteroskedasticity and non-normality of unknown form," *Computational Statistics and Data Analysis*, 54, 2532–2553.

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.

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.

Shevtsova, I. G. (2010): "An improvement of convergence rate estimates in the Lyapunov Theorem," *Doklady Mathematics*, 82, 862–864.

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

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 1) with bounds on the deviation of $\hat{\beta}_j$ from its mean provided by Propositions 1, 2, 3 and 4.

A.1. Bound on the variance of $\hat{\beta}_j$

Lemma 1

$$\mathrm{Var}(\hat{\beta}_j) \leq \bar{\sigma}^2_{\beta_j}$$

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

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

Since $\mathrm{Var}(Y) = \sum \tau^2_{ji}\mathrm{Var}(Y_i)$,

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

■

A.2. Cantelli

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

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

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

Proposition 1 *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 $\sigma, t > 0$, φ_C is increasing in σ and 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

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

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

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

which is point 2. Point 3 is immediate. ■

A.3. Bhattacharyya

Consider a random variable Z with $\mathbf{E}Z = 0$. Let $\sigma^2 = \mathrm{Var}(Z)$, $\gamma_1 = \frac{\mathbf{E}Z^3}{\sigma^3}$, 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

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

Relying on this inequality we derive:

Proposition 2 *1. For $\bar{t} > 0$ and $\beta_j \leq \bar{\beta}_j$,*

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

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. φ_{Bh} is increasing in σ and decreasing in t.

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

Lemma 2

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

and

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

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

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

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

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

Using the polynomial expansion again, we get

$$\mathbf{E}\left(\hat{\beta}_j - \beta_j\right)^4 = \sum_i \tau^4_{ji}\mathbf{E}\left(\varepsilon^4_i\right) + 3\sum_{i \neq k} \tau^2_{ji}\mathbf{E}\left(\varepsilon^2_i\right)\tau^2_{jk}\mathbf{E}\left(\varepsilon^2_k\right)$$

and

$$\left(\sum_i \tau^2_{ji}\mathbf{E}\left(\varepsilon^2_i\right)\right)^2 = \sum_i \tau^4_{ji}\mathbf{E}\left(\varepsilon^2_i\right)^2 + \sum_{i \neq k} \tau^2_{ji}\mathbf{E}\left(\varepsilon^2_i\right)\tau^2_{jk}\mathbf{E}\left(\varepsilon^2_k\right).$$

From this we derive

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

Using the Cauchy-Schwarz inequality twice we obtain

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

and hence

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

■

Proof of Proposition 2. 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):

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

We then obtain point 1 by maximizing (5), which is concave in γ_1 over all $\gamma_1 \leq \frac{\|\tau_j\|_\infty}{\sigma_{\beta_j}}$, holding σ_{β_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 φ_{Bh} when $\frac{t^2}{\sigma^2} - \frac{t\|\tau_j\|_\infty}{\sigma^2} - 1 > 0$ are increasing in σ 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 $\bar{t} > 0$,

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

Relying on Hoeffding's inequality we show:

Proposition 3 *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 $t > 0$, φ_H is decreasing in t.

Proof. We apply Hoeffding's inequality to $(Z_i)_i$ where $Z_i = n\tau_{ji}Y_i$. So $Z_i \in [0, n\tau_{ji}]$ for $\tau_{ji} \geq 0$ and $Z_i \in [n\tau_{ji}, 0]$ 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 $\mathrm{Var}(Z_i) = \sigma_i^2$. For $\bar{u} \in \mathbb{R}$,

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

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

Proposition 4 *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 $\sigma, t > 0$, φ_{BE} is increasing in σ and decreasing in t.

The idea of the proof of Proposition 4 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 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 $(Z_i)_{1 \leq i \leq N}$ 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*

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

we have

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

Proof. We apply Berry-Esseen's inequality to the family of independent random variables $Z_1, ..., Z_{n+N}$ 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 $\sum_{t=1}^{n+N} Z_i$. Let $\delta \sim \mathcal{N}(0,1)$. The Berry-Esseen inequality applied to $\sum_{t=1}^{n+N} Z_i$ shows:

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

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

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

Lemma 4

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

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 b_1, $\bar{u}$ and independent random variables W_1 and W_2. In our case, we write:

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

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

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

Lemma 5

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

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

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

we derive

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

■

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

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

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

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

Point 3 is immediate. ■

Appendix B: Proofs for Section 4

Proposition 5 *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

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

The interpretation of the proposition is as follows: Consider a randomized test that rejects the null hypothesis with probability equal to 1 if $n\bar{W} \geq \bar{k}$, equal to λ if $n\bar{W} = \bar{k} - 1$, and equal to 0 if $n\bar{W} < k - 1$. 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. Note that this randomized test is the uniformly 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.

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

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

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

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

which is point 2. ■

Proof of Theorem 2. Let $\beta_j \leq \bar{\beta}_j$. From point 1 of Proposition 5, 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 5 implies the stated bound on type II error probabilities. ■