User guide

Febuary 10, 2012

This is the user guide for implementing the tests for a single coefficient of a linear regression as explained in Gossner and Schlag (2011). It assumes no knowledge of the R programming language.

Given Y=X*beta+error where there are no assumptions imposed on the errors, it tests the one sided hypothesis H0: betaj<=betabarj against H1: betaj>betabarj where j is index of coefficient.
It also tests H0: betaj>=betabarj against H1: betaj<betabarj.

One constraint is that a bounded interval containing the range of Y has to be known ex-ante. One can also apply when Y is unbounded by making statements that are conditional on Y falling within some exogenously given range. Note that this range may not be chosen as the maximal and minimal element observed in the data.

Installing R and set-up

1) Install R, which is freely available from the web, for instance http://cran.r-project.org/bin/

2) The following R packages need to be installed (only once):
Rlab
Rglpk (version 0.3-5)
quadprog

These can be installed as follows: in the drop-down menu go to packages -> Install package(s) then choose a mirror following the prompt. Select *Rlab* from the list of packages. Once the *Rlab* file has been downloaded, check that the file is in the library directory, for me this is C:\Program Files\R\R-2.11.1\librar . The same then needs to be done for *Rglpk* and *quadprog.*

On a mac, use the package installer.

3) Each time you open R you have to load the two packages above, in the drop-down menu go to packages -> load packages

On mac, use the package manager to select these two packages.

Then enter this command in the R console
setwd("PATH")
where you replace "PATH" by the path to your folder containing the file
Nonstandardized_and_Bernoullitests_v031.r

Preparing the regression

Two files must be prepared, one for the covariates, and one for the dependent variable.

The file containing the covariates is a tab separated table of the form (X_ij)_ij where the row i is the observation numbers and the column j is the covariate. If you wish to have a constant then you need to add it to this matrix.

The file containing the values for the dependent variable contain one observation per line.

The two files must be in the same directory as the file ExactLinear.r

Setting the regression parameters

The file Nonstandardized_and_Bernoullitests_v031.r contains a number of instructions that set the parameters of the regression. This file must be edited according to the regression you wish to run, then saved.

The values of alpha, j, betabarj, betaj, w1Y, w2Y, Y, X are set by the user.
- j is the index of the covariate to be tested
- alpha is the significance level for the test
- betabarj is the value of the against which the test is run. H0 : betaj ≤ betabatj
- betaj is used by the program to select the most efficient test. More precisely, it selects the test that is guaranteed to have the lowest type II error at the value of betaj. It is recommended to set this value of betaj such that the obtained type II error bound is around 0.2 or 0.5
- w1Y is the minimal value for Y
- w2Y is the maximal value for Y
- Y is the file containing the values of the dependent as described above
- X is the file containing the values of the covariates as described above

Some extra parameters can be optionally set to user values. We recommend novice users to leave them to their default values.

- lambda is a parameter used to select d in the Bernoulli method. The selection of d in Gossner and Schlag (2012) set this parameter to 1.
- monte is the number of Monte Carlo simulations for the Bernoulli test. increasing monte will increase accuracy.

Launching the regression

The regression is launched by typing into the console the command

```
source("Nonstandardized_and_Bernoullitests_v031.r")
```

Reading the output

Let's focus on "Summary Results". As an example we use the following output.

```
      Summary Results, n= 734 m= 28     Y range [ 0 , 1 ]
1   Estimated coefficient of beta 3 :            -0.03091
2       Reject H0: beta 3 >= 0.185 ? NO  at alpha =  0.025
3                      based on:           Bernoulli
4                Estimator (tau):             Minimax
5                         theta:            0.10517
6 Type II bound for beta 3 <= -0.015 :             0.88914
```

The dataset contains 734 independent observations and 28 covariates, the dependent variable is known to belong to [0,1].
1: We are testing coefficient beta_3, for your information (not directly relevant for the further output) its OLS estimate is equal to -0.03091.
2: The null hypothesis H0: beta 3 >= 0.185 is not rejected at level 2.5%.

3,4: The method, Bernoulli or Non-Standardized, is selected as explained in the paper. In this case, it is the Bernoulli test with tau that mimizises its sup norm, and theta=0.10517
5: The bound on type II errirs if beta 3 <= -0.015 is equal to 0.88914.

The suggested methodology is that the user should change betaj1 until the type II bound is approximately equal to 0.5. It is namely this bound that determines which test is selected that determines whether or not the null hypothesis is rejected.

This leads to the following output:

```
            Summary Results, n= 734 m= 28      Y range [ 0 , 1 ]
1         Estimated coefficient of beta 3 :            -0.03091
2           Reject H0: beta 3 >= 0.185 ? YES  at alpha =  0.025
3                       based on:  Non-Standardized test
4                 Estimator (tau):                OLS
5              using H  cutoff (t-bar):           0.20914
6 Type II bound ( BE  ) for beta 3 <= -0.083 :              0.49735
```

In fact, now the null hypothesis is rejected. The test selected is the Non-Standardized test with threshold (or cutoff) determined by Hoeffding inequality (H) and type II bound determined by Berry-Esseen's inequality (BE).