Report on Master's Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague

Student:	**Magdaléna Raušová**
Advisor:	**Jan Ámos Víšek**
Title of the thesis:	**The least weighted squares and its asymptotics**

OVERALL ASSESSMENT

The topic of thesis represents one of complicated problems from the theoretical point of view and very laborious in the simulations. Nevertheless, it had to be solved as the results opened a way for constructing modifications of diagnostic tools, we are used to employ for OLS as Durbin-Watson statistic, White test, Hausman specification test, etc - modifications for the least weighted squares (LWS). It is sufficient to realize two things: Firstly that diagnostic tools play a key role when we try to justify the estimate of the underlaying regression model. Secondly, practically all classical diagnostic tools are tailored (quite naturally) to given esstimator, typically to the OLS (or ML) and the residuals – as an approxiation of disturbances (about which we have no direct information) – play the key role in their construction. Due to the fact that we have for LWS no formula for regression coefficients we cannot give formula for residuals. Then the representation of residuals which we obtain when we compute residuals employing the asymptotic representation of estimates of the regression coefficients, can be used as „proxies" for hypothetical residuals we would obtain if we could use some formula for the LWS-estimates of regression coefficients.

Although the asymptotic representation of LWS estimator was established earlier within the homoscedastic framework, its generalization for the heteroscedastic situation (which is of course much more realistic framework for economic problems) was not in any case simple task. By the way, even to learn and understand new mathematical notions and methods for the student of economics (not student of mathematics) - to be able to understand the proofs for the homoscedastic disturbances - represented quite large portion of endeavour. To make an idea how technically complicated task Magdaléna Raušova solved it is sufficient to look into the chapter 3, especially on pages 30 up to 75. Even today when the requirement on the knowledge of students of the economics in the region of mathematics has increased, the scale of new things which Magdaléne had to learn is extraordinary. After learning them she proved - by employing them - that she did not only learn them but that she understand their underlying ideas. I would bet that it will help her in the next study so much that she will be one of best. I would add that I am sure that even this theoretical part of work could be sufficient to fulfill requirement on master's thesis. On the other hand, from the point of view of complexity of the research the theoretical results of asymptotic character require always to carry out some simulations to find out how tight approximation the asymptotic formulas give and from such sample size we can rely on them.

That is why she wrote programs in MATLAB for making the simulations and then carried out the Monte Carlo study. Collecting the results into the tables is not very easy and entertaining work. Prior to saying a few words about the results of simulations, I would like to recall one thing. Although we (I mean all researchers in robust statistics and econometrics) construct the robust methods in order to cope with a contamination of data, we prove consistency, root of n consistency, etc. under the "null hypothesis", i. e. in the framework which assumes "noncontaminated data" (the converted commas mean that "noncontaminated data" in this case means that the data are distributed according to some d. f. which fulfills some assumptions while the contamination is assumed to be something in data sets

of finite extent which we cannot reasonably statistically “catch”). That is why we should – strictly taken – make the simulations for asymptotic representation of LWS only for noncontaminated data. For this case it may appear that the asymptotic representation can work better (i. e. they can have a d. f. with the smaller variance) than the results of simulations. That is easy to see why: The symptotic representation in fact takes into account the hypothetical data of infinite size, i. e. data which are distributed exactly according to the assumed d.f.. In other words, exaggerating a bit we have data set so ”dense” that the empirical d.f. of them coinsides with the theoretical one. Contrary to it, simulated data have finite size and due to the fact that – even if we don’t include any contamination – there can (and in fact among e.g. 500 simulated samples, do) appear in some data sets some atypical (outlying) observations. Then the variance of estimators of individual regression coefficients is larger then the asymptotic variance and hence the d.f. has a bit fatter tails than the d.f. of asymptotic representations. Of course, from time to time, it may appear also an opposite case, results of simulations are random !!

On the other hand, when Magdaléna prepared codes for these simulations, it was quite easy to answer interesting question how much the d.f. of estimates obtained for contaminated data differ from d.f. for asymptotic representation, i.e. how quickly (in some sense) the estimates of regression coefficients deteriorate with increasing level of contamination. That was the reason to include also these results.

At the end of supervisor’s report I cannot say anything else what I said already at the end of my report on Magdalena’s bachelor theses: I appreciate her entusiasm with which she learnt so much new things, I appreciate her intuition when we discussed something both in theoretical region as well as in programing in MATLAB and although she met with the things in question the first time and she needed to think about it once again at home, she was able to catch it generally although vaguely and to make about it some useful idea. The old teacher- as I am – is aware of the fact that relation between teacher and student is mutual – not only student learns something from teacher but also teacher learns some useful experience from the work with student, from her approach to solving the problems, etc. I am grateful to Magdaléna for it.

I assume that the quality of results, of the text as well as the amount of work deserve some special appreciation. I believe that something like compliment by dean of faculty can be easy justfiable.

SUMMARY OF POINTS AWARDED

CATEGORY		**POINTS**
Literature	*(max. 20 points)*	20
Methods	*(max. 30 points)*	30
Contribution	*(max. 30 points)*	30
Manuscript Form	*(max. 20 points)*	18
TOTAL POINTS	*(max. 100 points)*	**98**
GRADE	**(1 – 2 – 3 – 4)**	**1**

NAME OF THE REFEREE: Jan Ámos Víšek

DATE OF EVALUATION: September 9, 2016

Referee Signature