

# Report on Bachelor Thesis

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

<b>Student:</b>	<b>Magdaléna Raušová</b>
<b>Advisor:</b>	<b>Jan Ámos Víšek</b>
<b>Title of the thesis:</b>	<b>Robustification of regression model with the fixed and random effects</b>

## OVERALL ASSESSMENT

The topic of thesis represents an actual problem, solution of which can be applied in many econometric models. E.g. model of gravity are more and more estimated as model with fixed and/or random effects.

On the other hand, the ideas of robustness penetrate into the applied econometric research already for couple of years. There are several reason for it and we are not going to discuss them. Let us however recall that some robust methods are nowadays already available in the commercially supplied packages and researchers are more and more aware that even very carefully prepared data can (and typically do) contain some contamination. But there are (at least) two obstacles for the wider and reliable (or even accountable) application of the methods. The first one is the fact that the producers of packages included methods which are – from the traditional point of view of researchers – easier to accept philosophically. One such method is e.g. quantile regression which is able to cope with outliers but not with the leverage points (which are however much more influential than the outliers). The other obstacle is a low level of understanding the methods by researchers. The latter one is more dangerous because – as any sophisticated tool – the application of a (highly) robust method can give completely misleading findings if employed without understanding.

By writing the thesis Magdaléna Raušová proved that she has overcome both these points. She enlarged her knowledge of both – classical regression as well as of robust methods. In the region of the classical regression analysis she studied the models with fixed and random effects and the traditional ways how to estimate the coefficients of models. At a first glance it seems not to be a large work but to think out the differences of the estimating methods needs some time, efforts and patience. Moreover, there are also some a bit tricky issues as e.g. a trade off between the magnitude of improvement of efficiency of estimation in the case of applying the corrections on autoregression in random effect model (especially when the correcting factor is close to one and hence the correction is “near” to fixed effect model estimation) versus possibility of maybe weak dependence between explanatory variables and effects which need not be discovered by tests.

Much larger work was done in the region of robust methods. This is the topic which is taught only on a level of a brief information in the extent of a half of lecture. She read two basic books – famous theoretical book by Frank Hampel and his students, Peter Rousseeuw, Elvezio Ronchetti and Werner Stahel, and the book which is much more oriented on the practical applications by Peter Rousseeuw and Annick Leroy. The former one is nice and full of new ideas but the mathematics is done only as some hints to be finished by reader, moreover assuming knowledge e.g. of derivative of functional (e.g. Fréchet or Hadamard) in the rather complicated (vector) space of all distribution functions (i.e. in the space which is uncountably dimensional). The latter requires not only to read the text but to sit down to PC and to try to follow the authors in calculating a lot of examples of real (historical, and nowadays “benchmark”) data. She also studied a series of papers and also a couple of my papers

from the last two decades. She became acquainted with the new approach of the robust statistics to the evaluation of quality of estimators, i.e. with new characteristics of their properties and also with problems with computing these estimators (which were never met in the classical statistics).

Last but not least 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. The results show that the robustification of the traditional estimation methods for the fixed and random effects model is possible and gives acceptably good results if the contamination is up to some level, say 5, maximally 10%. For large contamination, due to the fact that we have to estimate from the first step (i.e. from the "naïve" model without taking into account effects and their influence on the estimation) the improvement is small, if any, and for extremely large contamination (above 15%) it is probably better to use just a naïve (of course, robust) estimation.

The thesis answered positively the basic question whether the robust methods, here the least weighted squares were employed, can be modified for the framework with effects. The thesis is valuable by opening new questions and ideas what to study in the next step. E. g. a straightforward idea can be that the estimation in the case of large contamination is to be iterative to reach an improvement in estimation the "mean value" of the fixed effects and/or a better estimation of variance of random effects and of disturbances. An opened question is also the optimality of the weight function – the topic which will be extremely difficult to treat theoretically but can be diagnosed by simulation studies.

The thesis is the result of more than the whole year work and the result corresponds to it. It represents a lot of work in all directions – study of new topics, generalizing the previous results, writing new software and finally collecting the results in tables. Moreover, writing the „body“ of dissertation which explains to reader the models with effects as well as the ideas and tools of the robust regression is surely a new experience for Magdalena. I do believe that this experience will be useful in future.

I appreciate her enthusiasm with which she learnt so much new things, I appreciate her intuition when we discussed something both in theoretical region as well as in programming in MATLAB and although she met with the thing 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. 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 justifiable.

### **SUMMARY OF POINTS AWARDED**

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

**NAME OF THE REFEREE: Jan Ámos Víšek**

**DATE OF EVALUATION: July 4, 2014**



**Referee Signature**