Advisor's Report on Dissertation Thesis

Institute of economic Studies, Faculty of Social Sciences, Charles University in Prague
Opletalova 26, 1110 00 Praha 1, the Czech Republic
Phone: +420 222 112 330, Fax: +420 222 112 304

Author	Mgr. Eva Michalíková
Advisor	Prof. RNDr. Jan Ámos Víšek, CSc.
Title of the Thesis	Methods of the robust econometrics with aplications on the economic data
Type of Defence	DEFENCE

The content of thesis was described in the Abstract, Foreword of dissertation and Introductions of respective chapters and also in all opponents' reports, hence I will not repeat it. So, let me to do only two things. Firstly, to answer the questions given in the questionary for advisor and secondly, I'll give a response to some suggestions of opponents' reports.

The contribution of the thesis is in two directions, in econometrical and economic basic research. The former is in an original application and generalization of some methods of data processing, namely robust regression analysis and generalized method of moments. The later is described in the reports of opponents, especially those who are specialists for economics and it is much more reliable than my judgement.

The question if the thesis is based on the relevant references seems to me to be a bit smile-evoking. E. g., in the thesis - as it was already mentioned - GMM was applied. When we, I mean our institute, celebrated in the past autumn 20 years of the faculty and of the institute, I looked for references for this method (by the Google) and it found 6 320 000 references. Who is able to say which of them are "relevant" for GMM in this thesis. But now seriously. I read all the papers at the moment when they were prepared to be published and I felt quite comfortable what concerns of introduction into the framework of the topic treated in the papers, partly by the explanation and discussion given there, partly by references on papers. Hence I think that the thesis gives sufficient scale of references. Moreover, I did not notice any case of employment of any result which would not be properly referred to.

Two of three chapters of the thesis were already published, the last one will be surely published (I believe that even if the present author would not be eager to do so, the coauthor, professor Karel Janda will not allow her to leave it unpublished).

Let me to skip the question about some major suggestion what to improve as it will be clear in a minute later.

I recommend the thesis to be now defended because the feedback offered by opponents' reports was taken into account and I believe that also in the final speech at the defence of the thesis will reflect it. What concerns of still unpublished paper it is on the authors to accept (or not) the recommendations but they should (at least) to take them as a feedback, indicating that something should be explained in more details, that something deserves a larger discussion, that something is not so commonly known and hence a few words is to be said about it, etc.

In the report of professor Marian Grendar there are two suggestions or critiques. First of them speaks about not taking into account (when evaluating the significance of the explanatory variables) the proper p-values, derived as conditional probabilities of the corresponding t-statistics. The second, in a very modest way recalls that the main estimating technique, namely LTS was invented more than 25 years ago and that, since then, much more attention was paid to M-estimators. Moreover, the M-estimators are more efficient than LTS estimation. The next discussion enlightens the point of view which influenced the fact how the recommendations were taken into account.

Let us discuss the issues one by one. First of all, let us recall a definition of M-estimator.

Definice 1 *Let $\rho : R \to R^+$ be a function having a derivative almost everywhere (for technical simplification) and denote the corresponding derivative by ψ. Then*

$$\hat{\beta}^{(M,n)}(Y,X) = \underset{\beta \in R^p}{arg\,min} \sum_{i=1}^{n} \rho\left(Y_i - X_i'\beta\right)$$

is called M-estimator of regression coefficients.

Two of (very) desirable properties of the (classical) estimators are equivariance and/or invariance with respect to some family of transformations of data (usually to shift, i.e. to a change of location of data, to multiplication, i.e. to a change of scale of data or to an affine (regression) transformation, i.e. to a rotation of the coordinate system. The classical estimators as the least squares, moments ones or maximum likelihood posses these properties by definition. So, we are are used to them so much that we forget to verify whether the non-traditional estimators have them or not. First of all, let's recall what is scale- and regression-equivariance.

Definice 2 *The estimator $\hat{\beta}$ is called scale-equivariant*

$$\forall\left(c \in R^+\right) \qquad \hat{\beta}\left(cY,X\right) = c\hat{\beta}\left(Y,X\right)$$

and regression-equivariant

$$\forall\left(\gamma \in R^p\right) \qquad \hat{\beta}\left(Y + X\gamma, X\right) = \hat{\beta}\left(Y,X\right) + \gamma.$$

It is clear that M-estimator (as given in Definition 1) is not scale- and regression-equivariant. To achieve the scale- and regression equivariance, we have to studentize the residuals, $Y_i - X_i'\beta$.

Definice 3 *Let σ denote the standard deviation of disturbances*[1]*. Then*

$$\hat{\beta}^{(M,n)}(Y,X) = \underset{\beta \in R^p, \sigma \in R^+}{arg\,min} \sum_{i=1}^{n} \rho\left(\frac{Y_i - X_i'\beta}{\sigma}\right)$$

is (also) called M-estimator of regression coefficients (sometimes they are called generalized M-estimators).

There are algorithm (see e. g. Marazzi (1992)) which, in an iterative way, are able to evaluate such estimators, but it is (extremely) complicated to treat them theoretically (consistence, asymptotic normality). Hence we employ, when computing estimators as well as studying their theoretical properties, a consistent estimator of the standard deviation of disturbances, say $\hat{\sigma}$. There are plenty such estimators. Then we put

[1] For simplicity, let us assume homoscedasticity.

Definice 4

$$\hat{\beta}^{(M,n)}(Y,X) = \underset{\beta \in R^p}{arg\,min} \sum_{i=1}^{n} \rho\left(\frac{Y_i - X_i'\beta}{\hat{\sigma}}\right).$$

is called again M-estimator of regression coefficients.

It seems simple but it is not. Already in 1975 Peter Bickel (see also Jurečková, Sen (1993)) demonstrated that if we ask for the scale- and regression equivariance of M-estimators, we have to have the corresponding estimator of the standard deviation of disturbances scale-equivariant and regression-invariant, i. e.

$$\forall\left(c \in R^{+}\right) \qquad \hat{\sigma}\left(cY, X\right) = c\hat{\sigma}\left(Y, X\right)$$

and

$$\forall\left(\gamma \in R^{p}\right) \qquad \hat{\sigma}\left(Y + X\gamma, X\right) = \hat{\sigma}\left(Y, X\right).$$

There are (up to my knowledge) only several estimators fulfilling these requirements - see Croux, Rousseeuw (1992), Jurečková, Sen (1993), Víšek (1990), (2010) (by the way, when the last paper was prepared on the invitation of the *Institute of Mathematical Statistics*[2], I made a review of the topic and I found that the results by Bickel and Jurečková with Sen was discreetly forgotten). The estimators of the standard deviation of disturbances which are the scale-equivariant and regression-invariant, is either difficult to use because the computation is intensive (Jurečková, Sen (1993) - the estimator is based on the regression ranks and/or quantiles) or they are, all after, based on a preliminary estimator of regression coefficients with high breakdown point (as LMS, LTS, LWS[3]), having the properties of the scale- and regression equivariance.

So, it seems preferable to use such estimators of regression coefficients which are directly (i.e. without studentization) scale- and regression-equivariant as LMS, LTS, LWS or IWV[4] (moreover as we have nowadays a reliable and quick algorithm(s) for their evaluation, Hawkins (1994), Hawkins, Olive (1999), Víšek (1994), (1996), (2000), (2006c)).

By the way, a high popularity of M-estimators has its roots in the fact that they are easier to treat theoretically and easier to understand (as they resemble the maximum likelihood - after all, their name it indicates). On the other hand, to treat other estimators, especially those which have discontinuous object function (which is used in the definition of the respective estimator as a solution of an extremal problem), need not be very easy. E.g. the consistency of LTS was proved for a long time only for simple regression, Rousseeuw (1994), and only recently the consistency was proved for multiple regression model, Víšek (2006a) (nevertheless in a terrible way). Nowadays, however we can prove consistency of LWS (even under heteroscedasticity) in a bit simpler way, nevertheless even today such a tool as Skorohod's embedding into the Wiener process had to be used.

The question about the p-values requires a much larger discussion - moreover it is a question of a philosophical approach. Professor Marian Grendar refers to the paper by Welsh and Ronchetti (2002) but the paper is about the one-step M-estimators. The situation with the M-estimators is a bit different from the situation of employment of LTS by Eva Michalíková. Nevertheless, we should frankly admit that we only started some considerations how to take into account the fact

[2]IMS is a "highest body" in statistics, publishing the "extra-league" journals as Annals of Statistics, Annals of Probability and also some aperiodics as IMS Collections or IMS Lecture Notes.

[3]See Víšek (2011)

[4]Instrumental weighted variables, Víšek (2006b)

that the processing of "cleaned" data is a second step in the processing and may be that it should be taken into account, or at least it should be shown that it does not matter. Anyway, it is a nice inspiration for a further research (it seems nowadays that well prepared and carefully realized simulations may be the most efficient way how to solve the problem). But as I said, it is the topic for a much larger discussion than we have the space here.

References

Bickel, P. J. (1975): One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc. 70, 428–433.*

Croux C., P. J. Rousseeuw (1992): A class of high-breakdown scale estimators based on subranges. *Communications in Statistics - Theory and Methods 21, 1935 - 1951.*

Hawkins, D. M. (1994): The feasible solution algorithm for least trimmed squares regression. Computational Statistics and Data Analysis 17, 185 - 196.

Hawkins, D. M., D. J. Olive (1999): Improved feasible solution algorithms for breakdown estimation. *Computational Statistics & Data Analysis, Volume 30, Number 1, 1 - 12.*

Jurečková, J., P. K. Sen (1993): Regression rank scores scale statistics and studentization in linear models. *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, Physica Verlag, 111-121.*

Marazzi, A. (1992): *Algorithms, Routines and S Functions for Robust Statistics.* Wadsworth & Brooks/Cole Publishing Company, Belmont, California, 1992.

Rousseeuw, P.J. (1994): Unconventional features of positive-breakdown estimators. *Statistical & Probability letters* 19, 417 - 431.

Víšek, J. Á. (1990): Empirical study of estimators of coefficients of linear regression model. *Technical report of Institute of Information Theory and Automation, Czechoslovak Academy of Sciences (1990), number 1699.*

__________ (1994): A cautionary note on the method of Least Median of Squares reconsidered, *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Lachout, P., Víšek, J.Á. (eds), Academy of Sciences of the Czech Republic, Prague, 1994, pp. 254 - 259.*

__________ (1996): On high breakdown point estimation. *Computational Statistics (1996) 11:137-146, Berlin.*

__________ (2000): On the diversity of estimates. *Computational Statistics and Data Analysis 34, (2000) 67 - 89.*

__________ (2006a): The least trimmed squares. Part I - Consistency. *Kybernetika (2006), vol 42, 1 - 36.*

__________ (2006b): Instrumental weighted variables. *Austrian Journal of Statistics, 35(2006), No. 2& 3, 379 - 387.*

__________ (2006c): Instrumental Weighted Variables - algorithm. *Proceedings of the COMPSTAT 2006, eds. A. Rizzi & M. Vichi, Physica-Verlag, Heidelberg 2006, 777-786.*

__________ (2010): Robust error-term-scale estimate. *IMS Collections. Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: Festschrift for Jana Jurečková, 7(2010), 254 - 267.*

__________ (2011): Consistency of the least weighted squares under heteroscedasticity. *Kybernetika 47 , 179-206.*

Welsh, A. H., E. Ronchetti (2002): A journey in single steps: robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference 103, 287 - 310.*

Date	March 30, 2012
Advisor's signature	
Advisor's Affiliation	Prof. RNDr. JanÁmos Víšek, CSc. Institute of economic Studies, Faculty of Social Sciences, Charles University in Prague

Advisor's Report on Dissertation Thesis

Institute of economic Studies, Faculty of Social Sciences, Charles University in Prague
Opletalova 26, 1110 00 Praha 1, the Czech Republic
Phone: +420 222 112 330, Fax: +420 222 112 304

Author	Eva Michalíková
Advisor	Prof. RNDr. Jan Ámos Víšek, CSc.
Title of the Thesis	Methods of the robust econometrics with aplications on the economic data
Type of Defence	PRE-DEFENCE

The content of thesis was described in the Abstract, Foreword of dissertation and Introductions of respective chapters and also in all opponents' reports, hence I will not repeat it. So, let me to do only two things. Firstly, to answer the questions given in the questionary for advisor and secondly, I'll give a response to some suggestions of opponents' reports.

The contribution of the thesis is in two directions, in econometrical and economic basic research. The former is in an original application and generalization of some methods of data processing, namely robust regression analysis and generalized method of moments. The later is described in the reports of opponents, especially those who are specialists for economics and it is much more reliable than my judgement.

The question if the thesis is based on the relevant references seems to me to be a bit smile-evoking. E. g., in the thesis - as it was already mentioned - GMM was applied. When we, I mean our institute, celebrated in the past autumn 20 years of the faculty and of the institute, I looked for references for this method (by the Google) and it found 6 320 000 references. Who is able to say which of them are "relevant" for GMM in this thesis. But now seriously. I read all the papers at the moment when they were prepared to be published and I felt quite comfortable what concerns of introduction into the framework of the topic treated in the papers, partly by the explanation and discussion given there, partly by references on papers. Hence I think that the thesis gives sufficient scale of references. Moreover, I did not notice any case of employment of any result which would not be properly referred to.

Two of three chapters of the thesis were already published, the last one will be surely published (I believe that even if the present author would not be eager to do so, the coauthor, professor Karel Janda will not allow her to leave it unpublished).

Let me to skip the question about some major suggestion what to improve as it will be clear in a minute later.

I recommend the thesis to be defended without major changes - but on the other hand, the feedback offered by opponents' reports is to be used at least in the still unpublished paper and may be also in the final speech at the defence of the thesis. The present authors of still unpublished paper need not accept the recommendations but they should (at least) to take them as a feedback, indicating that something should be explained in more details, that something deserves a larger discussion, that something is not so commonly known and hence a few words is to be said about it, etc.

In the report of professor Marian Grendar there are two suggestions or critiques. First of them speaks about not taking into account (when evaluating the significance of the explanatory variables) the proper p-values, derived as conditional probabilities of the corresponding t-statistics. The second, in a very modest way recalls that the main estimating technique, namely LTS was invented more than 25 years ago and that, since then, much more attention was paid to M-estimators. Moreover, the M-estimators are more efficient than LTS estimation. Both may be, but also need not be necessarily true.

Let us discuss the issues one by one. First of all, let us recall a definition of M-estimator.

Definice 1 *Let $\rho : R \rightarrow R^+$ be a function having a derivative almost everywhere (for technical simplification) and denote the corresponding derivative by ψ. Then*

$$\hat{\beta}^{(M,n)}(Y, X) = \underset{\beta \in R^p}{arg\,min} \sum_{i=1}^{n} \rho\left(Y_i - X_i'\beta\right)$$

is called M-estimator of regression coefficients.

Two of (very) desirable properties of the estimators are equivariance and/or invariance with respect to some family of transformations of data (usually to shift, i.e. to a change of location of data, to multiplication, i.e. to a change of scale of data or to an affine (regression) transformation, i.e. to a rotation of the coordinate system. The classical estimators as the least squares, moments ones or maximum likelihood posses these properties by definition. So, we are are used to them so much that we forget to verify whether the non-traditional estimators have them or not. First of all, let's recall what is scale- and regression-equivariance.

Definice 2 *The estimator $\hat{\beta}$ is called scale-equivariant*

$$\forall \left(c \in R^+\right) \qquad \hat{\beta}\left(cY, X\right) = c\hat{\beta}\left(Y, X\right)$$

and regression-equivariant

$$\forall \left(\gamma \in R^p\right) \qquad \hat{\beta}\left(Y + X\gamma, X\right) = \hat{\beta}\left(Y, X\right) + \gamma.$$

It is clear that M-estimator (as given in Definition 1) is not scale- and regression-equivariant. To achieve the scale- and regression equivariance, we have to studentize the residuals, $Y_i - X_i'\beta$.

Definice 3 *Let σ denote the standard deviation of disturbances*[1]*. Then*

$$\hat{\beta}^{(M,n)}(Y, X) = \underset{\beta \in R^p, \sigma \in R^+}{arg\,min} \sum_{i=1}^{n} \rho\left(\frac{Y_i - X_i'\beta}{\sigma}\right)$$

is (also) called M-estimator of regression coefficients (sometimes they are called generalized M-estimators).

There are algorithm (see e. g. Marazzi (1992)) which, in an iterative way, are able to evaluate such estimators, but it is (extremely) complicated to treat them theoretically (consistence, asymptotic normality). Hence we employ, when computing estimators as well as studying their theoretical properties, a consistent estimator of the standard deviation of disturbances, say $\hat{\sigma}$. There are plenty such estimators. Then we put

[1] For simplicity, let us assume homoscedasticity.

Definice 4

$$\hat{\beta}^{(M,n)}(Y,X) = \underset{\beta \in R^p}{arg\,min} \sum_{i=1}^{n} \rho\left(\frac{Y_i - X_i'\beta}{\hat{\sigma}}\right).$$

is called again M-estimator of regression coefficients.

It seems simple but it is not. Already in 1975 Peter Bickel (see also Jurečková, Sen (1993)) demonstrated that if we ask for the scale- and regression equivariance of M-estimators, we have to have the corresponding estimator of the standard deviation of disturbances scale-equivariant and regression-invariant, i. e.

$$\forall\left(c \in R^{+}\right) \qquad \hat{\sigma}\left(cY, X\right) = c\hat{\sigma}\left(Y, X\right)$$

and

$$\forall\left(\gamma \in R^{p}\right) \qquad \hat{\sigma}\left(Y + X\gamma, X\right) = \hat{\sigma}\left(Y, X\right).$$

There are (up to my knowledge) only several estimators fulfilling these requirements - see Croux, Rousseeuw (1992), Jurečková, Sen (1993), Víšek (1990), (2010) (by the way, when the last paper was prepared on the invitation of the *Institute of Mathematical Statistics*[2], I made a review of the topic and I found that the results by Bickel and Jurečková with Sen was discreetly forgotten). The estimators of the standard deviation of disturbances which are the scale-equivariant and regression-invariant, is either difficult what concerns the computation (Jurečková, Sen (1993) - the estimator is based on the regression ranks and/or quantiles) or they are, all after, based on a preliminary estimator of regression coefficients with high breakdown point (as LMS, LTS, LWS[3]), having the properties of the scale- and regression equivariance.

So, it seems preferable to use such estimators of regression coefficients which are directly (i.e. without studentization) scale- and regression-equivariant as LMS, LTS, LWS or IWV[4] (moreover as we have nowadays a reliable and quick algorithm(s) for their evaluation, Hawkins (1994), Hawkins, Olive (1999), Víšek (1994), (1996), (2000), (2006c)).

By the way, a high popularity of M-estimators has its roots in the fact that they are easier to treat theoretically and easier to understand (as they resemble the maximum likelihood - after all, their name it indicates). On the other hand, to treat other estimators, especially those which have discontinuous object function (which is used in the definition of the respective estimator as a solution of an extremal problem), need not be very easy. E.g. the consistency of LTS was proved for a long time only for simple regression, Rousseeuw (1994), and only recently the consistency was proved for multiple regression model, Víšek (2006a) (nevertheless in a terrible way). Nowadays, however we can prove consistency of LWS (even under heteroscedasticity) in a bit simpler way, nevertheless even today such a tool as Skorohod's embedding into the Wiener process had to be used.

The question about the p-values requires a much larger discussion - moreover it is a question of a philosophical approach. Professor Marian Grendar refers to the paper by Welsh and Ronchetti (2002) but the paper is about the one-step M-estimators. The situation with the M-estimators is a bit different from the situation of employment of LTS by Eva Michalíková. Nevertheless, we should frankly admit that we never carried out considerations how to take into account the fact

[2]IMS is a "highest body" in statistics, publishing the "extra-league" journals as Annals of Statistics, Annals of Probability and also some aperiodics as IMS Collections or IMS Lecture Notes.

[3]See Víšek (2011)

[4]Instrumental weighted variables, Víšek (2006b)

that the processing of “cleaned” data is a second step in the processing and may be that it should be taken into account, or at least it should be shown that it does not matter. Anyway, it is a nice inspiration for a further research. But as I said, it is the topic for a much larger discussion than we have the space here. So, we shall cope with it in the modified version of thesis which will be prepared for the “sharp” defence (if any, I hope pretty sure that Yes).

References

Bickel, P. J. (1975): One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc. 70, 428–433.*

Croux C., P. J. Rousseeuw (1992): A class of high-breakdown scale estimators based on subranges. *Communications in Statistics - Theory and Methods 21, 1935 - 1951.*

Hawkins, D. M. (1994): The feasible solution algorithm for least trimmed squares regression. Computational Statistics and Data Analysis 17, 185 - 196.

Hawkins, D. M., D. J. Olive (1999): Improved feasible solution algorithms for breakdown estimation. *Computational Statistics & Data Analysis, Volume 30, Number 1, 1 - 12.*

Jurečková, J., P. K. Sen (1993): Regression rank scores scale statistics and studentization in linear models. *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, Physica Verlag, 111-121.*

Marazzi, A. (1992): *Algorithms, Routines and S Functions for Robust Statistics.* Wadsworth & Brooks/Cole Publishing Company, Belmont, California, 1992.

Rousseeuw, P.J. (1994): Unconventional features of positive-breakdown estimators. *Statistical & Probability letters* 19, 417 - 431.

Víšek, J. Á. (1990): Empirical study of estimators of coefficients of linear regression model. *Technical report of Institute of Information Theory and Automation, Czechoslovak Academy of Sciences (1990), number 1699.*

__________ (1994): A cautionary note on the method of Least Median of Squares reconsidered, *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Lachout, P., Víšek, J.Á. (eds), Academy of Sciences of the Czech Republic, Prague, 1994, pp. 254 - 259.*

__________ (1996): On high breakdown point estimation. *Computational Statistics (1996) 11:137-146, Berlin.*

__________ (2000): On the diversity of estimates. *Computational Statistics and Data Analysis 34, (2000) 67 - 89.*

__________ (2006a): The least trimmed squares. Part I - Consistency. *Kybernetika (2006), vol 42, 1 - 36.*

__________ (2006b): Instrumental weighted variables. *Austrian Journal of Statistics, 35(2006), No. 2& 3, 379 - 387.*

__________ (2006c): Instrumental Weighted Variables - algorithm. *Proceedings of the COMPSTAT 2006, eds. A. Rizzi & M. Vichi, Physica-Verlag, Heidelberg 2006, 777-786.*

__________ (2010): Robust error-term-scale estimate. *IMS Collections. Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: Festschrift for Jana Jurečková, 7(2010), 254 - 267.*

__________ (2011): Consistency of the least weighted squares under heteroscedasticity. To appear in *Kybernetika.*

Welsh, A. H., E. Ronchetti (2002): A journey in single steps: robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference 103, 287 - 310.*

Date	March 16, 2011
Advisor’s siniture	
Advisor’s Affiliation	Prof. RNDr. JanÁmos Víšek, CSc. Institute of economic Studies, Faculty of Social Sciences, Charles University in Prague