Report on Master Thesis

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

Student:	**Ján Palguta**
Advisor:	**Prof. RNDr. Ing. František Turnovec, CSc.**
Title of the thesis:	**Information Complexity of Strategic Voting**

***OVERALL ASSESSMENT** (provided in English, Czech, or Slovak):*

The thesis consists of an introduction to strategic voting theory, game representation of strategic voting and measuring responsiveness of strategic voting to information possessed by voters about other voters' preference profiles. This measurement is achieved by simulating many preference profiles of 2, 3, 5, 7 and 11 voters over 3 or 4 alternatives, which gives in total 10 different situations and hence 10 different results. I am not fully persuaded that simulations can provide general results as long as there is no analytical work behind them telling us exactly what are the impacts of all the parameters including the probability distributions which enter the simulation process, however I find this attempt more or less relevant to at least indicate some answers to the research questions listed on page 9.

I have several questions or comments to the text itself:

1) I would rather use „analytical skills" than „analytic skills" in the list of RESEARCH QUESTIONS on page 9.
2) In the LITERATURE listed on page 10 I would recommend to use the number of issue besides the volume of the journals, where articles were published for records n. 4 (issue No. 1), 9 (issue No. 4), 10 (issue No. 3), 12 (issue No. 1) and 14 (issue No. 2).
3) On page 11 where the Arrow's impossibility theorem is described, ther should be „no voting system for three **or** more alternatives" instead of „no voting system for three **of** more alternatives".
4) In the second line on page 12, there is redundant „that".
5) In the line 11 on page 13 should be „occurs" instead of „occur".
6) In the line 5 of the second paragraph on page 13, there is redundant „voters".
7) In the first line of the last paragraph on page 14, there should be „capsize" and not „capsizes".
8) In the description of the voting environment the set A is not a member of the set U, but member of the set T or subset of U.
9) In the page 16 and 17, the definition statements should be followed by colons.
10) On page 17 is stated: „Let us further denote the elements in the individual preference order by $[r_1 r_2 \ldots r_m]$, where j denotes the jth position of an element in the individual preference ordering.", but I don't see any j in the preference order.
11) In the page 17, the alternatives are denoted not „font-consistently", which can be a source of confusion.
12) On page 18, there should be a comma behind the definition of collective preference profile R.
13) On page 19 in the definition of social preference ordering, is the social preference ordering given by $a_1 R_i a_2 R_i A_3 \ldots a_{m-1} R_i a_m$ for some particular i or for all i in N? I expect the latter, but I miss it in the text.
14) In the line 1 of the second paragraph on page 19, there is redundant „of".
15) In the definition of lottery, there should be a capital K in the expression $p_1 = p_2 = \ldots = p_k = k^{-1}$ instead of small k.
16) In the Example 2.4. on page 21 there should probably be „three canoes" or „the canoes" instead of „thee canoes".
17) **Why the Euclidean metric is used to evaluate the distance between two orderings? Why for example the Kemeny's distance is not used? What is the impact of choosing Euclidean metric to the final results?**
18) On page 23 author states, the utility function is monotonically decreasing in distance and writes down that the partial derivative of the utility function as a function of distance with respect to this distance is nonpositive. But this can hold without having a monotonically decreasing function as the function can be constant. How does the author knows, the utility is differentiable for all relevant distances (in the definition of utility function, there is no mention of differentiability)? However a major objection to the differentiability of utility with respect to the

Report on Master Thesis

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

Student:	Ján Palguta
Advisor:	Prof. RNDr. Ing. František Turnovec, CSc.
Title of the thesis:	Information Complexity of Strategic Voting

distance is that there is a finite number of different orderings on a finite set of alternatives and hence there is also a finite number of attainable distances between two orderings. The utility function is then defined in a finite number of distinct points, not continuum and so the derivative does not exist at all. I would get rid of the expression with derivative.

19) On page 24 I would recommend to use rather „have chosen" than „chose".

20) I do not fully understand what does it mean for the weights to be marginally non-increasing in j (the last paragraph on page 23). Can author provide the definition or he means just non-increasing?

21) In the second paragraph on page 24 there should probably be „weights attached" not „weights attaches".

22) In the last paragraph on page 24, there is redundant „the" in the last sentence.

23) In the definition of average distance on page 25, there should probably be „Let D_{iSk} represent the k-th potential distance" instead of „Let D_{iSk} represent k potential distances". Also in the expression defining the average distance of voter i individual preference ordering from the social preference ordering should be the last summand equal to $p_K D_{iSK}$ instead of $p_k D_{iSK}$.

24) In the figure 1 on page 25 the x-axis (horizontal) represents the average distance or something else? If it represents the average distance, how it can attain utilitiy values for negative distances, when the Euclidean distance can never be negative?

25) In the first paragraph on page 26 author writes about utility function being convex on the negative domain, but this feature of utility function was not mentioned before. And again the utility function was defined on a finite set of alternatives or on distances, so I do not understand what does the negative domain stands for.

26) In the last paragraph on page 27, there should rather be „other voters' preferences" than „other voters' preference" as long as the other voters have not a single preference.

27) The statement about no strategic behavior of the other voters is crucial and author mentions arising Nash equilibria. Such an analysis would be probably more relevant. Even there is a part where author introduces the situations with more than a single strategic voter, he assumes it is the same as when there is a lack of information of their preferences, but what if they know each other's preferences, but also know about each other that they are sophisticated? Is it really the same as a lack of information?

28) In the first line of the second paragraph on page 29, there should probably be „will be crucial" instead of „will we crucial".

29) In the last line of the third paragraph on page 31, there should rather be „other voters' preferences" than „other voter's preferences".

30) In the second line of the first paragraph on page 32, there should be rather „preference profiles of all other players" than „preference profile of all other players".

31) I do not fully understand the fundamental part of the analysis. Why author uses 100 000 iterrations for all the combinations of 2, 3, 5, 7, 11 voters and 3 or 4 competing alternatives when for up to 3 voters the total number of collective preference profiles is below 100 000 and for 5 voters and 3 alternatives it is also below 100 000? Wouldn't it be enough to check just all the possible collective profiles?

32) How has author chosen the numbers of voters and numbers of alternatives? Does the analysis of such a small part of the possible number of combinations of the number of alternatives and number of voters say anything about the general behavior of the strategic voting and its responsiveness to level of information possessed by the voters? Are these combinations representative in some way?

33) In the fourth paragraph on page 45, there is missing „be" in the text „some voters may not able to construct".

34) I would welcome to have some explicitly given bound on deviation of the simulated results from the results that would be in case of taking into account all the possible collective preferences (using for example the Höeffding inequality). How fast is the convergence of the

Report on Master Thesis

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

Student:	**Ján Palguta**
Advisor:	**Prof. RNDr. Ing. František Turnovec, CSc.**
Title of the thesis:	**Information Complexity of Strategic Voting**

simulated values to the real values that would be computed from all the possible collective preference profiles?

35) In the summary statistics in tables 19 and 20, what does the sample variance of the values tells us? They depend on the number of selected aggregation rules, so the absolute value of the sample variance does not tell us much and for comparisons between the different numbers of voters it is also problematic as the number of not missing probabilities is not the same in the case of just 2 voters and additionaly when different aggregation rules were selected, the variances will be different. For example taking into account only Plurality voting, Hare's and Coomb's procedures, the sample variances would be for 2, 3, 5, 7 and 11 voters and 3 alternatives: (0.083, 0.000, 0.011, 0.042, 0.18) compared to what is in the table: (0.019, 0.016, 0.019, 0.010, 0.010). This indicates that the variances are more or less given by arbitrary selection of a group of agregating rules instead of describing some feature of the strategic voting.

36) **The results seem to depend substantialy also on the weights assignments using Borda's count. Can author provide some explanation of why the results should be resistent to the selection of distance function and the selection of Borda's count for the weights assignment to particular alternatives? I am not persuaded that the results are general and tells anything more than what seems to be true for the arbitrarily selected strict assumptions. I think some robustness check should be applied to confirm the results.**

37) According to the regression table 21 on page 57: I am little bit scared of the regression used for this purpose as technically there is 84 observations and 12 parameters to be estimated. I would also like to know, why there are just 84 observations, when there are 94 entries in the table 18 (maybe it is somewhere explained, but I haven't found it). Are there some entries ommitted? If so, why? If I put up with this analysis, I would have to ask about the normality of error terms, which should moreover be of zero mean and homoscedastic. If the errors are not like this, the F-test and confidential intervals of the estimated parameters as well as the t-tests for their significance will not work properly as the estimation of the parameters and the estimation of the variance of the error term are not independent any more. When I redo the regression analysis for all 94 observations from table 18, I obtain the following:

Estimate		Std. Error	t-value	Pr(>|t|)	
method1	0.164203	0.034602	4.745	8.69e-06	***
method2	-0.017206	0.030895	-0.557	0.579104	
method3	0.389703	0.034602	11.262	< 2e-16	***
method4	0.207578	0.034602	5.999	5.15e-08	***
method5	0.346294	0.030895	11.209	< 2e-16	***
method6	0.123494	0.030895	3.997	0.000139	***
method7	0.102594	0.030895	3.321	0.001342	**
method8	0.098994	0.030895	3.204	0.001930	**
method9	0.344894	0.030895	11.163	< 2e-16	***
method10	-0.020406	0.030895	-0.660	0.510791	
n	-0.005037	0.002670	-1.887	0.062741	.
m4	0.224830	0.016767	13.409	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08128 on 82 degrees of freedom
 (6 observations deleted due to missingness)
Multiple R-squared: 0.9433,   Adjusted R-squared: 0.935
F-statistic: 113.6 on 12 and 82 DF, p-value: < 2.2e-16
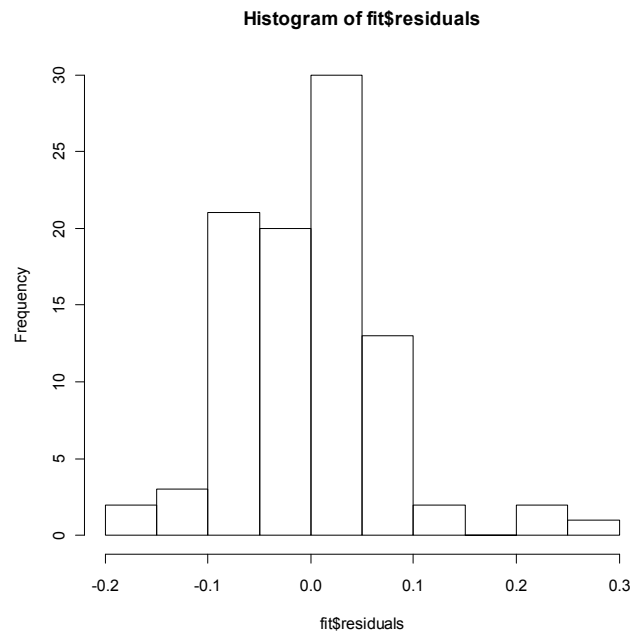With the following QQ-Plot of residuals:

Report on Master Thesis

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

Student:	**Ján Palguta**
Advisor:	**Prof. RNDr. Ing. František Turnovec, CSc.**
Title of the thesis:	**Information Complexity of Strategic Voting**
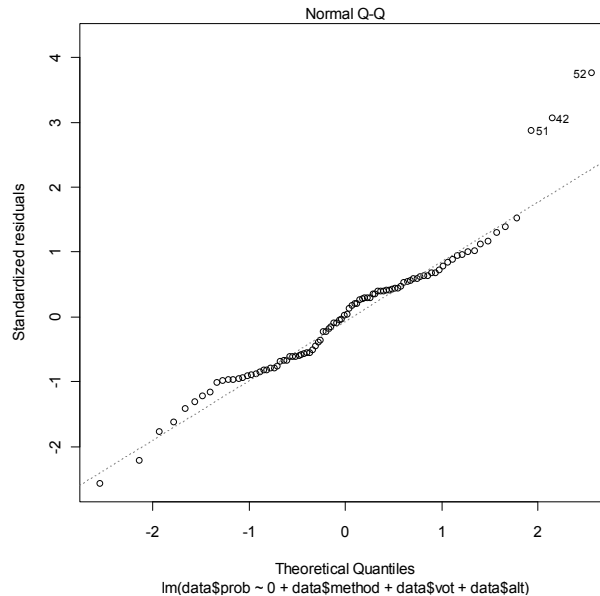
and histogram of residuals:

The Shapiro-Wilk test rejects the hypothesis of normality and hence we can not rely on the statistical tests of parameters' significance and on the F-statistic. In my opinion it is redundant to use regression analysis in this situation as there is no need to support the results with some sophisticated statistics although it looks nice and cute at the first glimps.
The same objections about the regression hold for the analyses performed in Regression table 30 and Regression table 31.

38) On page 57, the vector $proced_i$ is not of 10 x 10 (it would not be a vector, but a matrix), but it is a vector of 10 x 1. Otherwise we would have to sum vector with scalars. It is simply a vector

Report on Master Thesis

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

Student:	**Ján Palguta**
Advisor:	**Prof. RNDr. Ing. František Turnovec, CSc.**
Title of the thesis:	**Information Complexity of Strategic Voting**

of zeros, with a single entry of 1, which is in the place corresponding to the aggregation rule procedure used for the i-th observation.

39) In the second paragraph on page 59 the zero hypothesis is once denoted H_0 and once H_o. Moreover the analysis of the manipulation susceptibility level depending on the number of sincere voters presumes the normality of error term as author uses the confidential intervals derived from the Student distribution which needs to have the nominator and denominator random variables in its definition independent.

40) In the Bibliography list, author sometimes do not use the issue numbers for articles.

CONCLUSION:

I find the attempt to deal with the strategic voting the way it is presented in this work an interesting piece of work with several auspicious ideas and a good starting point for further research. Author shows to be open minded and bring original ideas, however some work should be done in the area of formal explanations and deriving results analytically and not just computationally. I am persuaded there is a big space for an analytical work studying the topic without simulations. Making simulations is relatively easy, but also very dangerous technique which is moreover very badly verifiable and heavily depending on usually a large set of exogenously given parameters, probability distributions and their relationships. Author seems to know this. I also find the assumption of expecting just one voter to behave strategically to be very strong and I very much missed the most interesting situation from my point of view-situation where there are more strategic voters, they know about each other's preferences, but they also know about each other that they behave strategically. If author satisfactorily answers the most fundamental comments above, I would recommend **"výborně" (excellent, 1)** as most of the objections are minor. For further publication, the thesis would need some additional work. Anyway, if the work is proposed to an award I will support it to encourage the author to continue working in this interesting field of study.

SUMMARY OF POINTS AWARDED (for details, see below):

CATEGORY		POINTS
Literature	*(max. 20 points)*	20
Methods	*(max. 30 points)*	20
Contribution	*(max. 30 points)*	27
Manuscript Form	*(max. 20 points)*	15
TOTAL POINTS	*(max. 100 points)*	**82**
GRADE	**(1 – 2 – 3 – 4)**	**1**

NAME OF THE REFEREE: Pavel Doležel

DATE OF EVALUATION: 11.6.2010

_____
                    Referee Signature

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

Strong	*Average*	*Weak*
20	*10*	*0*

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

Strong	*Average*	*Weak*
30	*15*	*0*

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

Strong	*Average*	*Weak*
30	*15*	*0*

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Strong	*Average*	*Weak*
20	*10*	*0*

Overall grading:

TOTAL POINTS	GRADE		
81 – 100	1	= excellent	= výborně
61 – 80	2	= good	= velmi dobře
41 – 60	3	= satisfactory	= dobře
0 – 40	4	= fail	= nedoporučuji k obhajobě