

Report on Master Thesis

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

Student:	Bc. Jakub Černý
Advisor:	PhDr. Lenka Št'astná, Ph.D.
Title of the thesis:	Spatial Analysis of Czech Parliamentary Election: Comparison of Spatial Econometrics and Machine Learning

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

Short summary

The thesis builds spatial models of electoral outcomes of Czech political parties in the 2021 elections to the Chamber of Deputies. This topic is important for political economists who need to understand if socio-economic differences are still the major driving determinants of voting behavior (as was until 2010) or not, and therefore if changes in economic policies can be predicted primarily by socio-economic changes. In the 2021 elections, this issue was especially interesting as two large coalitions (Spolu and Pirati+STAN) were formed and members in each coalition are to some extent heterogenous.

The thesis consists of two parts. First, it analyzes spatial trends within region-level data using established spatial econometric framework. The aim is to reflect spatial patterns in voting behavior, in particular by assuming that the error terms in the voting regressions are spatially dependent (and also that determinants have a spatial lag). Second, it alternatively analyzes the data using several ML algorithms, especially a random-forest method applied to the spatial context (GRF). There is also a modest attempt to compare the two approaches.

Contribution

There is a contribution specifically in the spatial econometrics part where the author finds support for the Spatial Durbin Error Model.

Another contribution rests in employing ML algorithms to spatial data which is a very new approach. The thesis therefore represents a novel, albeit purely applied, contribution to this area of growing interest. At the same time, if I am correct, out-of-sample predictions of these particular ML models probably cannot be given as coming elections are of different nature and valence of policies has changed significantly over the last year.

Methods

The author had to master spatial econometrics tools which are not part of our curriculum. In addition, he had to get familiar with the state of the art in electoral studies. Also, he had to apply ML methods which are increasingly more popular in our program, but have not yet become part of the standard course offer. He extensively used Python (data processing), R (both spatial econometrics and ML) and Geoda (geographical visualisation).

Literature

In addition to papers in (political) economics, a very lengthy overview of studies from several specialized fields is provided, including electoral studies, regional studies, and literature from geography and artificial intelligence.

Manuscript form

The thesis is clearly written and well structured.

Report on Master Thesis

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

Student:	Bc. Jakub Černý
Advisor:	PhDr. Lenka Šťastná, Ph.D.
Title of the thesis:	Spatial Analysis of Czech Parliamentary Election: Comparison of Spatial Econometrics and Machine Learning

Overall evaluation and suggested questions for the discussion during the defense

I rank this thesis highly. The author got familiar with a collection of econometrics tools that are not a standard part of the program; he also had to work with non-standard (electoral) data. The author carefully executed each step in spatial analysis (see, e.g., construction and selection from spatial weight matrices). Results of the methods are compared.

Although the thesis is of applied nature, the comparison of the results across methods can contribute to broader discussions on properties of ML algorithms.

My questions for the defense are as follows:

- When you speak about lower “interpretability“ of ML results (e.g., p. 10), what do you mean exactly? If I get it correctly, random forest trees differ from classic econometrics only by the way the underlying data generating process is modeled and by complexity of the process, but the process can be unambiguously interpreted like in classic econometrics (is not a black box).
- Share of Roma people: What is exactly the (so-called economic) interpretation of the coefficient of this variable in the baseline regression? (It is said that the effect is largest, but the size of the coefficient obviously depends on construction of the variable.) In other words, what is the effect of an increase by 1%, what is the effect of an increase from a minimal to a maximal value, and what is the effect of an increase by a standard deviation?

The results of the Urkund analysis do not indicate significant text similarity with other available sources.

In my view, the thesis fulfills the requirements for a Master thesis at IES, Faculty of Social Sciences, Charles University, I recommend it for the defense and suggest a grade A.

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

CATEGORY	POINTS
<i>Contribution</i> (max. 30 points)	28
<i>Methods</i> (max. 30 points)	28
<i>Literature</i> (max. 20 points)	20
<i>Manuscript Form</i> (max. 20 points)	20
TOTAL POINTS (max. 100 points)	96
GRADE (A – B – C – D – E – F)	A

NAME OF THE REFEREE: *Martin Gregor*

DATE OF EVALUATION:

Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

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.*

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.*

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

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.*

Overall grading:

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F