



UNIVERSITATEA BABEȘ-BOLYAI
TRADIȚIE ȘI EXCELENȚĂ



UBBFSEGA
Universitatea Babeș-Bolyai | Facultatea de Științe Economice și Gestionarea Afacerilor



Str. Teodor Mihali nr. 58-60
Cluj-Napoca, RO-400591
Tel.: 0264-41.86.52-5
Fax: 0264-41.25.70
econ@econ.ubbcluj.ro
www.econ.ubbcluj.ro

HABILITATION THESIS

Theoretical contributions in ROC curve and applied econometrics in insurance and scientometrics

Cristian Mihai DRAGOȘ

Cluj Napoca
2019

SUMMARY

The present habilitation thesis resumes my research activity following the successful PHD thesis defence in 2004 at the Babeş-Bolyai University from Cluj-Napoca, Romania, in the domain of Economic Cybernetics, Statistics and Informatics.

The first part of the thesis includes some of the theoretical and applied contributions, grouped in three sections: **Section A:** Theoretical contributions and applications of the ROC curve for discrete choice multinomial models; **Section B:** Applied multinomial discrete choice models in insurance; **Section C:** Econometric approach of factors affecting the scientific productivity of countries.

Section A includes theoretical contribution regarding Receiver Operating Characteristic curve for discrete choice multinomial models (categorical and ordered) and applications in financial scoring and automobile industry.

The first part of section A is entitled *Applied binary ROC curve in financial scoring*. A survey of economic literature about the statistical methods has been made, in order to compare their relative efficacy and the potential applications. A comparison of discriminant analysis, econometrics of discrete regressions (logit and probit models), nonparametric methods, artificial intelligence algorithms (neural networks and genetic algorithms) has been made. An experimental study realized through simulations pursues the selection of the model with the highest forecasting accuracy. The results, evaluated through binary ROC curve, show that an increase in the number of variables from the scoring model impacts significantly the growth of the average predictability and the decrease of the forecasting error's variance in the tested samples.

The second part of section A is entitled *Defining a two-dimensional ROC curve for categorical discrete choice models*. The study makes a review of the applications using categorical discrete choice models. The necessity of an adequate instrument of goodness of fit is emphasized, for a comparative analysis of the econometric models from a predictive point of view. We define the Receiver Operating Characteristics (ROC) curve for discrete choice models, both for an alternative of choice and globally,

for the whole set of alternatives. The global sensitivity can be also written as a weighted average of the sensitivities of the M alternatives. In the sample, the weights are the relative frequencies of the individuals that choose those alternatives. Similar with the global sensitivity, the global specificity can be also written as a weighted average of the specificities of the M alternatives. In the sample, the weights are the relative frequencies of the individuals that didn't choose those alternatives. One application on Romanian car industry depicts the differences that may occur between the two models based on the new defined ROC curve.

The third part of section A is entitled *Defining a two-dimensional ROC curve for ordinal discrete choice models*. Receiver Operating Characteristic (ROC) curve has been, for more than seven decades, a very popular instrument of goodness of fit for *binary* discrete choice models. Some studies have generalized this instrument in multidimensional space for ordered and categorical models. Starting from a set of principles, we define a two dimensional ROC for *ordered* discrete choice models: 2D-ROC-ODCM. In order to assess the *specificity*, the new instrument considers the distance between each alternative predicted by the model and the alternative in which the observation is. We employ simulations, evaluating the 2D-ROC-ODCM behaviour in certain situations: different correlations between the endogenous variable and the regressors, the variation of the number of regressors and the empirical variance in relation with the volume of the testing sample.

Section B includes applications of the categorical multinomial models of discrete choice and corresponding ROC curve in estimating the human behaviour in life and automobile insurance.

The first part of section B is entitled *Estimating consumers' behaviour in auto insurance using discrete choice models*. The study models the voluntary motor damage insurance consumer behaviour using discrete choice models, hypothesizing a hierarchical and a non-hierarchical decision. The sample consists of 311 car owners from the Cluj County, Romania. The econometric estimations use binary logit, multinomial logit and nested logit models. The predictive power of these models is compared by means of the Receiver Operating Characteristic curve for discrete choice models. The results reveal that the main factors affecting the purchase of a voluntary motor insurance policy are risk preference/aversion, the distance travelled by car, the

driver's education level and the ratio between the driver's income and the car price. In contrast to previous studies that estimated the risk profile only through proxy variables without accounting for any behavioural aspects, our study has successfully integrated the risk profile of the policyholders as a self-standing explanatory variable.

The second part of section B is entitled *Estimating consumers' probabilistic choice among products of life insurance using discrete choice models*. The study focuses on the Romanian life insurance market which is in full expansion. There exists competition between insurance companies as well as between different products of the same company. Using data that we collected from clients of a Romanian insurance company, we have observed two types of variables: attributes of the insurance products (e.g., profitability, risk), as well as characteristics of the individuals (e.g., gender, age, income). Based on elements of economic theory and a general multinomial logit model we explain the behaviour of the life insureds. We estimate the variations in the market shares of life insurance products using marginal effects. The variations are due to possible changes in the values of some attributes or characteristics.

The third part of section B is entitled *Econometric estimation of buying life insurance and private pensions*. We test several hypotheses about the influence of behavioural and socio-demographic factors on buying a life insurance policy or a private pension. Data was gathered from a sample of 1579 individuals, representative for the Romanian consumers. We designed a questionnaire which emphasizes four distinct categories of determinants: socio-demographic factors, general behavioural factors and specific behavioural factors, as well as a self-constructed index of insurance knowledge. Through logit regression models we highlight a different mechanism which distinguishes between intention and decision for both life insurance and private pensions. We show that specific behavioural factors and insurance knowledge are highly significant for the decision, but not significant for the intention to buy both types of products.

Section C includes econometric applications in Scientometrics, with emphasis on determining factors of countries' scientific productivity.

The first part of section C is entitled *Econometric approach of factors affecting scientific productivity in environmental sciences and ecology*. Different academic

bibliometric studies have measured the influence of economic, political and linguistic factors in the academic output of countries. Separate analysis in different fields can reveal specific incentive factors. Our study proves that the Environmental Performance Index, computed by Yale University, is highly significant ($p < 0.01$) for the productivity of research and development activities in Environmental Sciences and Ecology. The control variables like education financing, publishing of ISI Thomson domestic journals and the English language are also significant. The methodology uses Ordinary Least Squares multiple regressions with convincing results ($R^2 = 0.752$). The relative positions of the 92 countries in the sample are also discussed. We draw up a ranking of the countries' concern for the environment, considering evenly the scientific productivity and the environment quality. We notice huge differences concerning the number of inhabitants and population income between the countries that dominate the classification and those occupying the last positions.

The second part of section C is entitled *Econometric approach of productivity in scholarly economics and business*. Some scientometric studies attempt to explain the factors affecting a country's scientific output which is usually measured by proxy variables such as the number of articles and citations in internationally-renowned journals. This paper highlights the main drivers for scientific output in Economics and Business, namely, financing of education and research, population size, the number of scholarly journals and English as the official language. We use multiple OLS regressions and data provided by *Web of Knowledge* and the *World Bank* covering 56 nations. The study also highlights the relationship between scientific output and the efficiency in using the research funding. The rankings of sample countries show that there is a learning process at national level, the output being doubled by efficiency.

The third part of section C is entitled *Scientific productivity versus efficiency of R&D financing: bibliometric analysis of African countries*. Commonly, bibliometric analysis at a national level measures the scientific performance by the complete production of articles and citations, and by productivity, related to the number of inhabitants. In our study we also report on efficiency, understood as the scientific output obtained, against the available financing. The results indicate that only two African countries (South Africa and Tunisia) have undergone a learning process and

become mature entities in the R&D process, productivity being accompanied by efficiency.

The second part of the thesis reviews my career evolution plan, emphasizing my orientation in research activities. My principal area of research in Econometrics and in Economics and Business Statistics considers the following domains of expertise: (1) goodness of fit methods for multinomial discrete choice models; (2) specific statistical methods in Scientometrics; (3) applied discrete regression and qualitative choice models in Economics and Business. Future researches will include mainly theoretical and applied developments on those three domains already mentioned. I will insist on identifying the statistical inference for the theoretical existing models, on econometric evolution of the statistical methods already employed in Scientometrics and on identifying some novel explanatory variables (even coming from psychology or other sides of the human behaviour) for some economic phenomena already studies.

The third part of the thesis includes the references list.