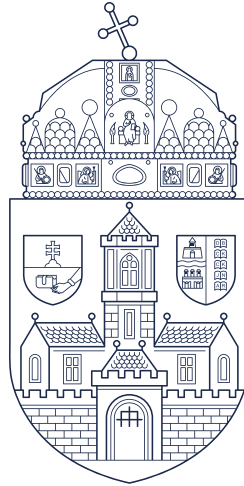


Óbuda University

PhD Thesis



Application Possibilities of Robust and Non-parametric
Statistical Tools in Respect of Economic Data

by

Ferenc Tolner

Supervisors:

Dr. György Eigner

Dr. Balázs Barta

Applied Informatics and Applied Mathematics Doctoral
School

Budapest, 2024

Contents

- 1 Background of the Research** **2**
- 2 Directions and Goals of the Research** **4**
 - 2.1 MFV-based Linear Regression for Regional Economic Convergence 4
 - 2.2 MFV-based Robust Outlier Detection 4
 - 2.3 Economic Resilience of Small and Medium-sized Enterprises 5
- 3 Materials and Methods of the Investigation** **6**
 - 3.1 MFV-based Linear Regression for Regional Economic Convergence 6
 - 3.2 MFV-based Robust Outlier Detection 6
 - 3.3 Economic Resilience of Small and Medium-sized Enterprises 7
- 4 New Scientific Results** **9**
- 5 Discussion and Practical Applicability of the Results** **13**
 - 5.1 Application of robust and non-parametric statistical tools regarding economic convergence 13
 - 5.2 Robust outlier detection based on the MFV concept and its application regarding economic convergence 14
 - 5.3 Resilience investigations via non-parametric hypothesis testing 14
- Index** **16**
- References** **16**

Acknowledgements

The outlined work is not yet the closing word, but an important milestone of the research I have been carrying out throughout my Ph.D. program. It could not have been reached without the support of many devoted people, to whom I would like to express my deepest gratitude. First and foremost to my supervisor, Dr. György Eigner from Óbuda University, who closely followed my pursuit and also witnessed my struggling. His prompt and encouraging guidances provided me the faith not to give up and search for new possibilities and options. I would also express my gratitude to Dr. Sándor Fegyverneki from the Department of Applied Mathematics, from University of Miskolc, for his invaluable help, for the time he devoted and for his professional guidance in the field of statistics. His small remarks and excellent mindset revealed how complicated it can be to interpret seemingly easy statistical problems and can conventional techniques lead researchers astray, furthermore drew my attention towards robust statistics. Both of their guidance has significantly contributed to increase the value of my work that might one day be utilized in further research activities.

Further thanks deserves Dr. Balázs Barta, Attila Joós and colleagues from the Pannon Business Network Association for the practical problem definition, access to necessary data and the provided framework throughout the years. Last but not least I would like to express my appreciation for the endurance and loving words of my wife, Mrs. Zsuzsanna Tolnerné Lőrinc who was always there for me and shared my burden in all walks of life.

Background of the Research

The aim of the present research is to introduce and apply robust- and non-parametric statistical methods regarding the question of economic convergence of regions of the European Union (EU) and the resilient behaviour of Hungarian small and medium-sized enterprises (SMEs) against idiosyncratic economic shocks that they undergo. Although, the presented methods and procedures are universal and can be used for answering various other problems, their application is unique in the selected research fields.

My main motivating goal was to introduce the Most Frequent Value (MFV) concept as a robust statistical alternative regarding economic investigations and furthermore to try to draw practically applicable and novel deductions for economic convergence and resilience from a statistical and data analytical point of view [R1].

Economic convergence of EU regions is a research field that is detailed in-depth in corresponding literature and is of great importance regarding proper resource distribution and corresponding strategic EU-level policies [R2, R3]. On the other hand, economic resilience is a continuously growing research field that took momentum especially in past years and definitely has implications both on local and on regional, national and international level as well [R4, R5].

Data-driven decision-making is essential in these fields, but the typically accessible data often do not support assumptions of conventional statistical procedures. The high level of skewness, mixture of distributions pertaining to different populations, outlying observations even in a multidimensional space are common issues that I also had to face throughout my research work. These attributes of economic data however can lead to non-resistant and non-robust results of conventional statistical analyses [R6, R7]. Therefore, I found extremely important to suggest and apply methods in the investigated fields, that alongside with widely used statistical procedures can potentially highlight distorting effects of the above listed undesirable circumstances and can lead to results that are expected to be of greater robustness and outlier resistance but still promising optimal statistical efficiency.

Thanks to the Management Board of the Pannon Business Network Association located at the West-Danubian Szombathely, Hungary, the current research could be based upon a rich longitudinal dataset of annual balance sheets and income statements together with metadata on locations, fields of activities and employment rates over almost 20 years of numerous members of the Hungarian processing industry. From a scientific point of view this offered a unique possibility on investigating economic resilience, since such a data source is extremely rare in corresponding literature [R8, R9, R10, R11]. Moreover, from a practical point of view, implications on national level policies regarding economic shock reactions could be offered in a data-based and data-driven way.

Corresponding literature often lacked to possess similar annual balance sheet and income statement information, onto I planned to build up my investigations. Consequently,

the accessed materials often misaligned my research setting. On the other hand, additional relevant fields (e.g.: bankruptcy prediction) utilizing similar data showed contradictory, case- and data sensitive results that drew my professional attention towards robust and non-parametric statistical application possibilities [R12, R5, R13]. For this purpose, detailed analysis of the MFV method was carried out, which was originally developed by Steiner et al. in relation to earth scientific investigations [R14, R15, R16, R17, R18] but no application to economic related fields were discovered up to the beginning of my research work. Therefore, throughout my investigations I devoted special interest to robust and non-parametric statistical techniques – in particular to the MFV concept – and their utilization together with other holistic approaches in order to be able to view our disposable economic related information in a novel way.

This work gained further motivation by the Covid-19 pandemic that suddenly hit on several levels ranging from the individual through microeconomic, macroeconomic up to even the social level. It was unknown at the beginning of my research work in 2020 what economic aftermath the pandemic would have, but it could be suspected that not every company would be affected at the same level and reactions to the crisis – with an unknown extent – would be different as well.

Although no further data on SMEs could be accessed after the pandemic, it became clear during the conducted work that literature statements on the ever-increasing exposure to economic turbulences due to globalization effects and growing interconnectedness [R11, R4] were becoming perceivable on the individual's level as well. The rapid financial crisis within several industrial sectors due to the pandemic followed by supply chain problems, raw material shortages, fierce market reactions, political inferences into economic processes on national-, EU- and broader international level finally an outburst of a new war on the edge of Europe. A short summary of events to support how extremely relevant it is to explore the different aspects of entrepreneurial shock reactions and how it can be characterized, measured and predicted based on accessible data in an objective way since it is evident that a possible collapse of individual economic stakeholders can have further negative impacts on other supply chain members, competitors, partners and even on regional or national level.

Directions and Goals of the Research

The theoretical and practical developments of the thesis were organized around the utilization of disposed economic data from different sources for investigating regional level economic convergence on EU level and analysing economic resilience. Data sources were selected in accordance with the main research goal to understand economic resilience in the broadest possible sense and to explore the application possibilities of the MFV method on the research field.

2.1 MFV-based Linear Regression for Regional Economic Convergence

The concept of absolute economic β -convergence is one of the mostly applied and fundamental approaches to investigate economic cohesion tendencies of regions and countries [R19, R20]. The problem can be translated to a linear regression problem and the majority of corresponding literature sources build upon the utilization of the L_2 -norm based Ordinary Least Squares method to approach it [R21, R2]. Accessible data regarding EU NUTS¹ regions show high variability in literature regarding time period, geographical content and applied methodological setup in order to assess whether regions show a converging tendency or not [R22, R23, R24, R25, R26, R27, R28, R29, R30].

Literature results seem to be contradictory that can also highlight the handling challenges of highly skewed data contaminated by outlying observations with conventional statistical procedures [R31, R16, R18]. My primary aim was therefore to collect and investigate accessible data with the robust MFV procedure and apply it in a 2D linear regression setting. Besides robust statistical investigations non-parametric, population-level investigations were also carried out that enabled the rough estimation of possible dates of convergence with respect to Gross Domestic Product (GDP) and Net Disposable Income (NDI) among EU member states and regions connected to the EU before 2004 and afterwards (often referred as West- and East EU states and regions). In the framework of a holistic approach additional datasets were sought regarding Horizon 2020 research activities to further support findings via the analysis of regional-level social network time evolution.

2.2 MFV-based Robust Outlier Detection

Since outliers in economic related data are hard to be considered as faulty observations or members of other populations they have to be treated with care and with special attention

¹Nomenclature of territorial units for statistics

that besides biasing conventional statistical results can carry valuable information. Their careless exclusion – besides altering the apparent type of the distribution of our data or artificially modifying its variance etc. – might lead to the abandonment of potentially important information as well or even lead to biased and false conclusions [R32, R33].

My primary aim was to identify outliers with respect to the constructed robust MFV-based linear regression line when considering EU-level economic β -convergence of NUTS regions and countries, thereby demonstrating the applicability of the MFV concept for robust outlier identification. Moreover, since the MFV-concept enables various applications and generalizations, a further outlier identification approach was elaborated by the utilization of MFV and dihesion values instead of mean values and standard deviations in the well-known Lloyd's algorithm of the widely used k-Means clustering scheme.

2.3 Economic Resilience of Small and Medium-sized Enterprises

SMEs constitute the backbone of the European economy. This sector constitutes approximately 99% of the total enterprises. Therefore, it also contributes with a great share to total employment and overall GDP production [R34]. In case of Hungary, the production sector contributes traditionally the most to the national GDP production [R35, R36]. Literature details numerous concept frameworks to assess economic resilience in several aspects that however mainly incorporate the economic resilience of individual enterprises [R37, R38, R39, R40]. Data-based, comprehensive analysis of whole population of SMEs is barely emerging in corresponding literature that might be due to the complicated access of relevant and satisfactory data.

My primary aim was to analyse accessible data with a special emphasis on reactions against idiosyncratic economic shocks triggered by arbitrary causes and draw general conclusions on resilient behaviour of enterprises on their long-term behaviour and development perspectives. Besides defining, measuring and understanding shock reaction patterns, the prediction of favourable shock reactions would be preferable via an "early warning system" [R8, R9, R10, R11]. Therefore, the elaboration of a predictive model for resilience-behaviour prediction was also among the main goals of my work.

Materials and Methods of the Investigation

3.1 MFV-based Linear Regression for Regional Economic Convergence

In contrast to the omnipresent OLS-based approaches of the linear regression problem of the absolute economic β -convergence of EU countries and NUTS regions in literature, robust statistical alternatives have been sought. The MFV-based robust and outlier resistant regression method, introduced first by Steiner et al. has been elaborated and applied for the accessible data [R17]. The implemented algorithm has been compared to other robust regression variants and its convergence properties have also been investigated [T1].

As a first step the background of the algorithm has been detailed and then applied to the economic convergence problem, where it clearly indicated major deviations in the pace provided by OLS-based regression. This observation was further supported by MFV-robustified correlation calculations as well [T2].

Since the existence of economic convergence within the EU on regional level is still a disputed question, population-level analysis of West- and East- European countries and regions has been performed as well, where the diminishing tendency of distribution bimodality regarding certain national- and regional level GDP and NDI curves has been identified. The results, gained based on Mann-Whitney U-tests and Wilcoxon signed rank tests showed evidence in several cases for population level convergence of the two subgroups and rough estimate of date of convergence was also provided via investigating time evolution of calculated decreasing statistic values [T3].

As part of a holistic approach, Horizon 2020 data on R&D projects were also brought into the analysis, where regional-level social network analysis was performed. Albeit results could not be linked to GDP and NDI data, the assessment of the analysis clearly indicated the tendency of greater cardinality of partnership establishments of East-European regions at given financial contribution levels provided by the EU authorities. This can implicate increased speed of uptake in R&D related projects that through innovation can have a direct local consequence on regional development and economic resilience [T4].

3.2 MFV-based Robust Outlier Detection

In my second thesis I focused on the outlier identification possibilities via the robust MFV-method. For this purpose, outliers in case of the linear regression problem of the absolute economic β -convergence of EU countries and regions were considered and isolated

in particular. Robust regression lines have been fit to the corresponding data and outlying observations were categorized compared to the fitted line.

Since no background information on the observations were present, none of the data points could be discarded by considering them to be an outlier of any kind of origin (e.g.: measurement error, member of another population etc.) and should be considered instead as a potential source of valuable information. Therefore, within the framework of the applied regression setup every available data point was kept, incorporated to the conducted analyses and considered to be a valuable information source but with an adequate weight. After identification of outliers compared to the given model, geographical representations were also generated in order to ease field relevant interpretation of the results.

On population level, also the additionally accessed Horizon 2020 R&D project information data were combined to the outlier investigations. Here, robust Mahalanobis distances were utilized to assess R&D related EU contributions compared to regional average GDP values.

As a further application area, the MFV-robustification possibility of the unsupervised k-Means clustering algorithm was done with demonstrations on sample datasets. The possibility for robust outlier detection within the formed clusters via the utilization of robust-Mahalanobis distances has been demonstrated.

3.3 Economic Resilience of Small and Medium-sized Enterprises

For the assessment of economic resilience among SMEs a literature-inspired resilience indicator number has been constructed and "data-tailored" according to practical possibilities. After procuring the necessary data and performing necessary transformations the shock- and shock-reaction history of Hungarian production sector SMEs has been generated.

For the investigation of long-term operation in a quasi experimental setup matched-pair analysis has been concluded. Due to non-normality of the data distributions besides one-tailed 2-sample t-tests, one-tailed Wilcoxon signed-rank tests were used in order to assess the optimistic hypotheses that had been formed in advance for resilient shock reactions of the given entities. The goal was to verify long-term superiority of resilient attributes regarding the occurrence of a second shock and verify that companies can learn from their prior experience and presumably restructure in a way that it contributes to their long-term development or can reassure significant advantages on population level compared to un-shocked control pairs [T5, T6]. Surprisingly, results showed an unexpected but obvious decline of resilient companies in the long-run, that definitely highlight the until now unknown feature of economic resilience.

In spite of the clear, but unexpected findings of the hypothesis tests it is evident, that for short-term survival and for local- and regional level resilience the presence of resilient attribute within economic organisations is of great importance [R41, R38, R4]. Therefore, an "early-warning" system for resilience is desirable, for predicting which enterprises would withstand or fail in case of a future idiosyncratic economic downturn. For this purpose, the labelling of the generated company-year observations was done according to the constructed resilience indicator number, then classification attempts were performed via logistic regression and decision trees on different, by literature suggested [R12, R5, R13, R42] cleaned datasets of company-year observations. Unfortunately, the range of data did not prove satisfactory regarding information content for predicting resilient behaviour of the given economic entities. At this point, my future goal is to introduce new resilience metrics or indices that fit the range of accessible data appropriately, broaden the range of data with transaction information in order to identify supplier-customer relationships

and conclude further machine learning based classification tests for achieving satisfactory predictive power in practice.

4

New Scientific Results

Thesis Group 1: Application of robust and non-parametric statistical tools regarding economic convergence

Thesis 1

I have identified via robust and non-parametric statistical approaches that convergence among EU countries and regions based on the absolute β -convergence regarding GDP per capita is of lesser degree than could be estimated based on the conventional Ordinary Least Squares (OLS) linear regression. I have also ascertained that economic convergence regarding GDP per capita also exists on the whole data-population level between EU member states and regions connected to the EU before and after 2004.

Publications relevant to the theses: [T1, T2, T3, T4, T7].

Thesis 1.1

With robust correlation calculations based on the Most Frequent Value concept, I have shown that there is less correlation among initial values and growth rates of GDP and NDI financial indicators of EU member states and regions, which suggest a slower pace of convergence than is suggested by OLS linear regression based absolute β -convergence theorem.

Publications relevant to the theses: [T2].

Thesis 1.2

With robust linear regression, based on the Most Frequent Value concept, I have shown that convergence of economies of EU member states and regions is of slower pace than provided by OLS linear regression based absolute β -convergence theorem.

Publications relevant to the theses: [T1, T7].

Thesis 1.3

With non-parametric hypothesis testing I have shown the existence of convergence among subgroups of economies of EU member states and regions connected to the EU before and after 2004 regarding GDP per capita on the whole data-population level. Based on the non-parametric approach, I have suggested a method for the estimation of rate of convergence and for the time needed for member regions to catch up.

Publications relevant to the theses: [T3].

Thesis 1.3.1

Having created the regional level social networks of participating organizations of the Horizon 2020 framework, I have shown that regions of EU member states connected after 2004 had generated more connections at the same level of average annual GDP or R&D subsidy given by the EU as member states and regions connected before 2004. This observation further supports the existence of regional convergence.

Publications relevant to the theses: [T4].

Thesis Group 2: Robust outlier detection based on the MFV concept and its application regarding economic convergence

Thesis 2

I have developed a clustering algorithm based on the MFV concept and performed robust outlier detection in case of linear regression problem with application area of the economic convergence of EU regions. I have developed a methodology to identify regions in a robust way that show faster or slower economic convergence than the bulk of the regions suggested by the absolute economic β -convergence theorem.

Publications relevant to the theses: [T1, T7, T8, T9].

Thesis 2.1

I have developed a method to identify outliers based on the MFV concept compared to the MFV-robustified linear regression. Regarding the absolute economic β -convergence of EU member states and regions, I have identified regions converging faster or slower than the bulk of the data suggested by the fitted robust trend line.

Publications relevant to the theses: [T1, T7].

Thesis 2.2

Based on the Minimum Covariant Determinant estimator, I have shown by the usage of robust Mahalanobis distances that regions of EU member states connected to the EU before 2004 and after are forming two separate groups regarding H2020 financial support. Furthermore, I have demonstrated the applicability of the method in identifying outlying regions compared to the bulk of the data when regional GDP, population and R&D funding is considered.

Publications relevant to the theses: [T9].

Thesis 2.3

I have developed a clustering algorithm by using robust location- and scale parameters based on the MFV concept that considers outliers compared to the bulk of the data but attenuates their biasing effect. The developed method can perform similarly or better regarding computation time measured in the number of centroid swaps in case of larger sample sizes simultaneously with more clusters to be identified than the robust k-Medians.

Publications relevant to the theses: [T8].

Thesis Group 3: Resilience investigations via non-parametric hypothesis testing

Thesis 3

I have designed an indicator number for measuring economic resilience of Small and Medium-sized Enterprises (SMEs) based on fluctuations of their annual sales growth, which was used for the classification of their reaction types to idiosyncratic shocks. With non-parametric hypothesis testing I have demonstrated that showing resilient behaviour is only relevant in the short term regarding individual financial development.

Publications relevant to the theses: [T5, T6, T10, T11].

Thesis 3.1

Based on the proposed resilience indicator number via matched pair analysis I have shown, that in short- and medium-term the economic attributes of companies reacting resilient to idiosyncratic economic shocks fall behind of those that had not suffered any kind of previous shocks measured in the setback of annual sales growth.

Publications relevant to the theses: [T5, T6].

Thesis 3.2

Via matched pair analysis, I have shown based on the available balance sheet-, income statement- and other metadata from Hungarian SMEs, that gained experience of short-term successful shock reactions does not influence their survival abilities. The transferability of experiences to the subsequent negative economic events is restricted. However, by enabling longer-term shock reactions the bankruptcy willingness decreases with the length of time enabled to bounce back from the economic shock.

Publications relevant to the theses: [T5, T6].

Thesis 3.3

I have shown that the proposed sales growth based resilience indicator number and corresponding classification of companies can be extended to a continuous scale of economic shocks and shock reactions, measured in the fluctuations of sales growth. The proposed extension offers a possibility for the comparison of individual observations regarding resilient characteristics. Furthermore, can characterize complete industrial branches regarding resiliency within arbitrary time intervals with respect to their shock reactions.

Publications relevant to the theses: [T6, T11].

Discussion and Practical Applicability of the Results

The dissertation presented solutions based on the utilization of robust statistical and non-parametric statistical tools. Throughout the outlined work data analytical considerations led to the selection of the specific practical choices. The outlined smaller developments can be applied on other statistic related problems as well in a general manner and the drawn economic considerations can be used as input for policymakers in order to enhance regional level resource distribution and re-evaluate former conceptions on individual-level economic resilience.

5.1 Application of robust and non-parametric statistical tools regarding economic convergence

The first thesis group investigated the practical usability of the MFV-based linear regression for the absolute economic β -convergence problem regarding EU countries and regions. Further results have been added by concluding non-parametric Mann-Whitney U-tests and Wilcoxon signed rank tests on population level by observing the time evolution of bimodal distributions of annual national and regional level GDP and NDI data. Furthermore, additional Horizon 2020 datasets on R&D activities on different EU levels have been incorporated in order to further support the previously gained results.

According to the findings, EU-level economic convergence among regions does exist, but its pace is slower than concluded by previous literature sources. Moreover, an approximate date of convergence of West- and East European subpopulations has been provided.

Although the applied methodology of MFV-based linear regression proved to be satisfactory, I have identified several opportunities for further improvement and possible research. The regression task has been implemented for 3-dimensional data as well, but could be done to higher dimensions and the generalisation of the Most Frequent Value for higher dimensions still poses an interesting research area. Furthermore, albeit literature suggests alternatives to interpret R^2 values for robustly fitted regression lines [R43], the question of confidence intervals is still to be elaborated for the MFV-based regression.

From policy point of view the results detailed in the thesis can be of great importance, since they give feedback on prior EU actions regarding population level effect that aimed to promote regional cohesion strivings. In case of further data procurement, results can be refined and effects of further policies could be inserted to an extended evaluation.

5.2 Robust outlier detection based on the MFV concept and its application regarding economic convergence

The second thesis group introduced outlier detection alternatives highly relying on the MFV-concept. For the case of the 2-dimensional linear regression a simple procedure has been proposed for the isolation of outliers in a data dependent way utilizing the problem specifically calculated, data-dependent MFV and dihesion values. Thereby, a naturally arising more robust and resistant outlier identification could be applied to the selected absolute economic β -convergence problem of EU countries and regions. Not only the slower pace of convergence compared by the predictions of OLS-based linear regression has been found but non-trivial outliers as well that can highlight problematic tendencies at certain regions for policymakers.

The k-Means clustering procedure based on the Lloyd's algorithm was updated according to the MFV concept and applied on existing, open data sources in order to gain a more robust clustering algorithm version that relies on a more interpretable background than the k-Medians or k-Modes. Albeit the proposed algorithm in its current form is rather computational costly, it can still serve as a reasonable robust clustering alternative in case of smaller datasets, or for larger datasets with higher cluster cardinality.

Within the framework of a future research activity the proposed methods can be further improved and generalized for higher dimensions. Their applicability for other data sources alongside with conventional statistical procedures is highly advised where non-normality or outlier presence is provided. It would be of great importance to further optimize for the computational cost of the "k-MFVs" algorithm and extend the application possibilities for arbitrary nonspherical data. The examination of these questions were beyond of the framework of the present thesis.

5.3 Resilience investigations via non-parametric hypothesis testing

The third thesis group investigated the resilient behaviour of Hungarian SMEs from the processing industry. The developed and investigated methods were meant to be applied on the disposed data set, however due to the time-consuming relatively late procurement of the data and the possibilities of approaching them for the given scientific goals made it necessary to utilize more data-tailored approaches.

In order to assess economic resilience an appropriate resilience indicator number has been used, based on which the idiosyncratic shock reactions of companies could be interpreted and its time-evolution investigated. It has been revealed via Mann-Whitney U-tests and Wilcoxon signed rank tests that our optimistic mindset on the long-term development perspectives of resiliently behaving economic entities is groundless. Instead, a long-term economic decline compared to non-shocked control pairs is to be expected with a slow uptake. This has a severe message for national and regional level assessment of economic resiliency by highlighting the need of supporting also the resilient companies in order not to lose development momentum on the individual-, regional- and national level.

The applied procedure enables the comparison of enterprises regarding economic shock reactions and offers possibilities for further similar, real-life-suited and data-tailored resilience indicator number or index definitions. As a further step a composite metric, utilizing other financial data beyond sales growth could be investigated, moreover besides one-year shock reactions, the two-years and three-years shock reactions could be viewed and compared by the former results.

In a future research, the from practical point of view relevant "early warning system for

resilience” should be further elaborated and verified based on existing longitudinal financial data records. Unfortunately, the logistic regression- and decision-tree based classification attempts on the labelled company-year observations did not enable an adequate model for the prediction of resilient shock reactions. Therefore, the extension of data is planned with transaction information for the identification of customer-supplier relationships and the investigation of time-propagation of economic shocks and shock reactions as a response within the resulted social network. Combining this investigation with geographical data would be an interesting research, since it could contribute to the understanding of economic shock development on regional level and within supplier-customer social network of the Hungarian SMEs of the processing industry.

Bibliography

- [R1] Ferenc Steiner. “(Editor), Optimum Methods in Statistics”. In: (1997). Akadémiai Kiadó, Budapest, Hungary, 370p., ISBN: 963 05 7439 X.
- [R2] Wojciech Dyba et al. “Regional Development in Central-Eastern European Countries at the Beginning of the 21st Century: Path Dependence and Effects of EU Cohesion Policy”. In: *Quaestiones Geographicae* 37.2 (Feb. 2018), pp. 77–92. DOI: <https://doi.org/10.2478/quageo-2018-0017>.
- [R3] Maciej Smetkowski and Piotr Wójcik. “Regional Convergence in Central and Eastern European Countries: A Multidimensional Approach”. In: *European Planning Studies* 20.6 (June 2012), pp. 1–17. DOI: <https://doi.org/10.1080/09654313.2012.673560>.
- [R4] Javaneh Ramezani and Luís M. Camarinha-Matos. “Approaches for resilience and antifragility in collaborative business ecosystems”. In: *Technological Forecasting and Social Change* 151 (2020), p. 119846. DOI: <https://doi.org/10.1016/j.techfore.2019.119846>.
- [R5] Sami Ben Jabeur. “Bankruptcy prediction using Partial Least Squares Logistic Regression”. In: *Jornal of Reailing and Computer Services* 36.C (Feb. 2017), pp. 197–202. DOI: <https://doi.org/10.1016/j.jretconser.2017.02.005>.
- [R6] Stephan Morgenthaler. “A survey of robust statistics”. In: *Statistical Methods and Applications* 15 (Jan. 2007), pp. 271–293. DOI: <https://doi.org/10.1007/s10260-006-0034-4>.
- [R7] Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. Wiley, New York, Second Edition, Sept. 2011. ISBN: 978-1-118-21033-8.
- [R8] Adina Aldea et al. “Assessing Resilience in Enterprise Architecture: A Systematic Review”. In: *2020 IEEE 24th International Enterprise Distributed Object Computing Conference (EDOC), Eindhoven, The Netherlands, 5–8 October* (Oct. 2020). DOI: <https://doi.org/10.1109/EDOC49727.2020.00011>.
- [R9] Eric Lutters Denzil Kennon Corné S.L. Schutte. “An alternative view to assessing antifragility in an organisation: A case study in a manufacturing SME”. In: *CIRP Annals - Manufacturing Technology* 64.1 (2015), pp. 177–180. DOI: <https://doi.org/10.1016/j.cirp.2015.04.024>.
- [R10] John Johnson and Adrian V. Gheorghe. “Antifragility Analysis and Measurement Framework for Systems of Systems”. In: *International Journal of Disaster Risk Science* 4.4 (2013), pp. 159–168. DOI: <https://doi.org/10.1007/s13753-013-0017-7>.

- [R11] Tjo-Kin Man. “Measuring and Analysing Resilience of Enterprise Architectures”. In: *31th Twente Student Conference on IT, University of Twente, Enschede, Netherlands*, (July 2019).
- [R12] Mohamed Oudgou Youssef Zizi and Abdeslam El Moudden. “Determinants and Predictors of SMEs Financial Failure: A Logistic Regression Approach”. In: *Risks* 8.4 (Oct. 2020), p. 107. DOI: <https://doi.org/10.3390/risks8040107>.
- [R13] Tamás Kristóf and Miklós Virág. “A Comprehensive Review of Corporate Bankruptcy Prediction in Hungary”. In: *Journal of Risk and Financial Management* 13.2 (Feb. 2020), pp. 1–20. DOI: <https://doi.org/10.3390/jrfm13020035>.
- [R14] Ferenc Steiner. “Need and possibilities for increasing the efficiency of geostatistical calculations”. In: *Földtani Közöny, Bull. of the Hungarian Geol. Soc.* 118.2 (1988), pp. 175–183.
- [R15] Ferenc Steiner. “Comparison of the L2-, L1- and P-norm based statistical procedures in respect of their asymptotic efficiencies”. In: *Magyar Geofizika* 41.1 (Feb. 2000).
- [R16] Ferenc Steiner. “New results on the theory of the most frequent value procedures”. In: *Geophysical Transactions* 41.1-2 (1997), pp. 1–21.
- [R17] F. Steiner. “The Bases of Geostatistics (In Hungarian)”. In: (1990). Tankönyvkiadó, Budapest, Hungary, 363p., ISBN: 963 18 2819 0.
- [R18] F. Steiner. “(Editor), The Most Frequent Value. Introduction to a Modern Conception of Statistics”. In: (1991). Akadémiai Kiadó, Budapest, Hungary, 315p., ISBN: 963 05 5687 1.
- [R19] Robert J. Barro and Xavier Sala-i-Mart. “Convergence”. In: *Journal of Political Economy*, 100.2 (Apr. 1992), pp. 223–251. DOI: <https://doi.org/10.1086/261816>.
- [R20] Robert M. Solow. “A Contribution to the Theory of Economic Growth”. In: *The Quarterly Journal of Economics* 70.1 (Feb. 1956), pp. 65–94. DOI: <https://doi.org/10.2307/1884513>.
- [R21] Kateina Dvoroková. “Sigma Versus Beta-convergence in EU28: Do they lead to different results?” In: *WSEAS Transactions on Business and Economics* 11.1 (Jan. 2014), pp. 314–321. URL: <https://www.wseas.org/multimedia/journals/economics/2014/a185707-228.pdf>.
- [R22] Mindaugas Butkus et al. “What Is the Evolution of Convergence in the EU? Decomposing EU Disparities up to NUTS 3 Level”. In: *Sustainability* 10.5 (May 2018), pp. 1–37. DOI: <https://doi.org/10.3390/su10051552>.
- [R23] Roberto Ezcurra and Manuel Rapún. “Regional Dynamics and Convergence Profiles in the Enlarged European Union: A Non-parametric Approach”. In: *Journal of Economic and Human Geography* 98.5 (Dec. 2007), pp. 564–584. DOI: <https://doi.org/10.1111/j.1467-9663.2007.00426.x>.
- [R24] Angelos Lontakis, Christos T. Papadas, and Irene Tzouramani. “Regional Economic Convergence in Greece: A Stochastic Dominance Approach”. In: *50th Congress of the European Regional Science Association: “Sustainable Regional Growth and Development in the Creative Knowledge Economy”* Conference Paper (Aug. 2010). URL: <https://www.econstor.eu/handle/10419/119147>.
- [R25] Nikolay Nenovsky and Kiril Tochkov. “The Distribution Dynamics of Income in Central and Eastern Europe relative to the EU: A Nonparametric Analysis”. In: *William Davidson Institute Working Paper* 1063 (Nov. 2013). DOI: <https://doi.org/10.2139/ssrn.2370625>.

- [R26] Danny T. Quah. “Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs”. In: *Journal of Economic Growth* 2.1 (1997), pp. 27–59. DOI: <https://doi.org/10.1023/A:1009781613339>.
- [R27] Gordon Anderson. “Making Inferences about the Polarization, Welfare and Poverty of Nations: A Study of 101 Countries 1970-1995”. In: *Journal of Applied Economics* 19.5 (May 2004), pp. 537–550. DOI: <https://doi.org/10.1002/jae.750>.
- [R28] Danny Quah. “Galtons Fallacy and Tests of the Convergence Hypothesis”. In: *Scandinavian Journal of Economics* 95.4 (Dec. 1993), pp. 427–443. DOI: <https://doi.org/10.2307/3440905>.
- [R29] Mihály Tamás Borsi and Norbert Metiu. “The evolution of economic convergence in the European Union”. In: *Empirical Economics* 38.2 (Mar. 2014). DOI: <https://doi.org/10.1007/s00181-014-0801-2>.
- [R30] Monica Rileanu Szeles. “Exploring the Economic Convergence in the EU’s new Member States by using non-parametric models”. In: *Romanian Journal on Economic Forecasting* 14.1 (2011), pp. 20–40. URL: https://ipe.ro/rjef/rjef1_11/rjef1_2011p20-40.pdf.
- [R31] Béla Hajagos and Ferenc Steiner. “Investigations Concerning Resistance - Importance of the Choice of the Formula Determining the Scale Parameter”. In: *Geophysical Transactions* 38.4 (1993), pp. 211–230.
- [R32] Pál Kerékfy. “About Robust Estimates”. In: *Alkalmazott Matematikai Lapok* 4 (1978), pp. 327–357.
- [R33] Ricardo A. Maronna, R. Douglas Martin, and Víctor J. Yohai. “Robust Statistics. Theory and Methods”. In: (2006). John Wiley & Sons Ltd, England, 403p., ISBN: 0-470-01092-4.
- [R34] Andrea Némethné Gál. “Competitiveness of Small and Medium-sized Enterprises - PhD Thesis (in Hungarian)”. In: (2009). István Széchenyi University, Doctoral School of Regional- and Business Administration, Győr, Available online: <https://rgdi.sze.hu/files/Ertekezések,%20tezisek/Magyar%20Tezis%20NGA.pdf> (accessed on 1 June 2023).
- [R35] Hungarian Central Statistical Office (KSH). “The amount of Gross Value Added and its distribution within the Hungarian NACE branches.” In: (2020). URL: https://www.ksh.hu/docs/hun/xstadat/xstadat%5C_eves/i%5C_qpt002d.html.
- [R36] Hungarian Central Statistical Office (KSH). “Characteristics of small and medium-sized enterprises”. In: *Web.*: <https://www.ksh.hu/docs/hun/xftp/idoszaki/pdf/kkv18.pdf> (2018).
- [R37] A. Aleksy et al. “An assessment of organizational resilience potential in SMEs of the process industry, a fuzzy approach”. In: *Journal of Loss Prevention in the Process Industries* 26.6 (Nov. 2013), pp. 1238–1245. DOI: <https://doi.org/10.1016/j.jlp.2013.06.004>.
- [R38] Cristina Ruiz-Martin, Adolfo López-Paredes, and Gabriel Wainer. “What we know and do not know about organizational resilience”. In: *International Journal of Production Management and Engineering* 6.1 (Dec. 2017), pp. 11–28. DOI: <https://doi.org/10.4995/ijpme.2018.7898>.
- [R39] Adam Rose and Elisabeth Krausmann. “An economic framework for the development of a resilience index for business recovery”. In: *International Journal of Disaster Risk Reduction* 5 (Sept. 2013), pp. 73–83. DOI: <https://doi.org/10.1016/j.ijdr.2013.08.003>.

- [R40] Muhammedamin Hussen saad et al. “Conceptualization of SMEs business resilience: A systematic literature review”. In: *Cogent Business & Management* 8.1 (2021), p. 1938347. DOI: <https://doi.org/10.1080/23311975.2021.1938347>.
- [R41] Paula Graça and Luís M. Camarinha-Matos. “Performance indicators for collaborative business ecosystems Literature review and trends”. In: *Technological Forecasting & Social Change* 116 (Mar. 2017), pp. 237–255. DOI: <https://doi.org/10.1016/j.techfore.2016.10.012>.
- [R42] Mario Hernandez Tinoco and Nick Wilson. “Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables”. In: *International Review of Financial Analysis* 30.C (2013), pp. 391–419. DOI: <https://doi.org/10.1016/j.irfa.2013.02.013>.
- [R43] Peter J. Rousseeuw and Annick M. Leroy. *Robust Regression and Outlier Detection*. John Wiley Sons, Inc., Oct. 1987. ISBN: 9780471852339. DOI: <https://doi.org/10.1002/0471725382>.

Own Publications Pertaining to Theses

- [T1] Ferenc Tolner, Balázs Barta, and György Eigner. “Comparison of Newtons and Broydens Method as Nonlinear Solver in the Implementation of MFV-robustified Linear Regression”. In: *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, Oct. 2022. DOI: <https://doi.org/10.1109/SMC53654.2022.9945222>. URL: <https://ieeexplore.ieee.org/document/9945222>.
- [T2] Ferenc Tolner et al. “Application of MFV-robustified Correlation Coefficient for the Investigation of the Strength of Beta-convergence of EU NUTS regions”. In: *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, Mar. 2022. DOI: <https://doi.org/10.1109/SAMI54271.2022.9780675>. URL: <https://ieeexplore.ieee.org/document/9780675>.
- [T3] Ferenc Tolner, Balázs Barta, and György Eigner. “Economic Cohesion Perspectives of the EU member Regions: A Non-parametric Approach”. In: *2022 IEEE 16th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. IEEE, May 2022. DOI: <https://doi.org/10.1109/SACI55618.2022.9919533>. URL: <https://ieeexplore.ieee.org/document/9919533>.
- [T4] Ferenc Tolner, Balázs Barta, and György Eigner. “Regional Level Investigation of EU-Funded H2020 Collaboration via Social Network Analysis”. In: *2023 IEEE 23th International Symposium on Computational Intelligence and Informatics (CINTI)*. IEEE, Nov. 2023. DOI: <https://doi.org/10.1109/CINTI59972.2023.10381903>. URL: <https://ieeexplore.ieee.org/document/10381903>.
- [T5] Ferenc Tolner et al. “Long-term Development Perspectives of Resilient Companies”. In: *2023 IEEE 21th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl’any, Slovakia*. IEEE, Jan. 2023. DOI: <https://doi.org/10.1109/SAMI58000.2023.10044504>. URL: <https://ieeexplore.ieee.org/document/10044504>.
- [T6] Ferenc Tolner, Balázs Barta, and György Eigner. “Reaction to Idiosyncratic Economic Shocks - Economic Resilience of Small and Medium-Sized Enterprises”. In: *Sustainability* 16.13 (2024), p. 5470. DOI: <https://doi.org/10.3390/su16135470>. URL: <https://www.mdpi.com/2071-1050/16/13/5470>.
- [T7] Ferenc Tolner, Balázs Barta, and György Eigner. “Outlier Identification with MFV-robustified Linear Regression in Case of Economic Convergence of EU NUTS Regions”. In: *ACTA POLYTECHNICA HUNGARICA* 21.8 (2023), pp. 47–66. DOI:

- <https://doi.org/10.12700/APH.21.8.2024.8.3>. URL:
https://acta.uni-obuda.hu/Tolner_Barta_Eigner_148.pdf.
- [T8] Ferenc Tolner et al. “Robust clustering based on the most frequent value method”. In: *Multidisciplinary Sciences* 13.1 (2023), pp. 141–153. DOI: <https://doi.org/10.35925/j.multi.2023.1.11>. URL: <https://ojs.uni-miskolc.hu/index.php/multi/article/view/2202>.
- [T9] Ferenc Tolner, Balázs Barta, and György Eigner. “Comprehensive Analysis of H2020 Funding Participation Based on LDA Topic Modeling and Robust Outlier Identification”. In: *2023 IEEE 21th International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, Sept. 2023. DOI: <https://doi.org/10.1109/SISY60376.2023.10417969>. URL: <https://ieeexplore.ieee.org/document/10417969>.
- [T10] Ferenc Tolner, György Eigner, and Balázs Barta. “Resilience Interpretations of Small and Medium-sized Enterprises and its Analytical Approaches - Literature Review”. In: *2021 IEEE 19th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, Jan. 2021. DOI: <https://doi.org/10.1109/SAMI50585.2021.9378637>. URL: <https://ieeexplore.ieee.org/document/9378637>.
- [T11] Ferenc Tolner, Balázs Barta, and György Eigner. “Economic Resilience and Antifragility: Classification of SME’s Shock Reactions based on Balance Sheet and Income Statement Data”. In: *2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania*. IEEE, May 2022. DOI: <https://doi.org/10.1109/SACI58269.2023.10158644>. URL: <https://ieeexplore.ieee.org/document/10158644>.

Own Publications Not Pertaining to Theses

- [Tx1] Ferenc Tolner et al. “Enhancing Cross-border Co-operation of Business Organizations based on the Investigation of Textual- and Categorical Information”. In: *ACTA POLYTECHNICA HUNGARICA* 19.5 (2022), pp. 235–255. DOI: <https://doi.org/10.12700/APH.19.5.2022.5.12>. URL: http://acta.uni-obuda.hu/Tolner_Barta_Takacs_Eigner_123.pdf.
- [Tx2] Ferenc Tolner et al. “Investigation of High-Growth Firms in the SME sector via the Perspective of Owners and CEOs using Wordclouds”. In: *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. IEEE, May 2021. DOI: <https://doi.org/10.1109/SACI51354.2021.9465560>. URL: <https://ieeexplore.ieee.org/document/9465560>.
- [Tx3] Ferenc Tolner et al. “Clustering based on Preferences with K-modes using Categorical Variables”. In: *2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, Sept. 2021. DOI: <https://doi.org/10.1109/SISY52375.2021.9582485>. URL: <https://ieeexplore.ieee.org/document/9582485>.
- [Tx4] Ferenc Tolner et al. “Clustering of Business Organisations based on Textual Data - An LDA Topic Modeling Approach”. In: *2021 IEEE 21th International Symposium on Computational Intelligence and Informatics (CINTI)*. IEEE, Nov. 2021. DOI: <https://doi.org/10.1109/CINTI53070.2021.9668337>. URL: <https://ieeexplore.ieee.org/document/9668337>.