University of Sopron
Faculty of Wood Engineering and Creative Industries
József Cziráki Wood Science and Technologies
Doctoral School

THE THESES OF THE DOCTORAL DISSERTATION

ANALYSIS OF THE STRUCTURE, FINANCING, NETWORKS AND RESULTS OF THE 7TH AND 8TH FRAMEWORK PROGRAMMES' WOOD PROJECTS

Created by: Adrienn Novotni
Supervisor: Dr. habil Zoltán Pásztory

Sopron
2023

1. Introduction and Objectives

In the course of the research, in addition to literature studies, text mining, descriptive and descriptive statistics and network research tools were used to analyse mainly the FP7 and H2020 projects in the wood sector, the project participants, the relationships of the project participants and the individual project results (by comparing the content with the wood patents). While efficiency was considered important in the analysis, efficiency was also examined not only in terms of the economic result that can be obtained from the direct output, but also in terms of an efficient resource mobilisation strategy.

In accordance with the general objectives of the literature and research projects, it was also felt important to emphasise the compatibility of the wood industry and sustainability, and to examine the sustainable wood industry.

2. Data and Methods

Several data sources were used in the empirical research for the thesis.

The analysis of articles on the wood industry projects using text mining and science metrics tools was based on the BibTeX file filtered and downloaded from Scopus.

The source of the scientific weight of the sustainable timber industry was also a dataset filtered and downloaded from Scopus. To analyse the FP7 and H2020 wood industry projects, the data tables available in CORDIS were downloaded, converted into a relational database and finally filtered to extract the necessary rows.

The network analysis and cluster analysis of FP7 and H2020 wood projects were carried out after content analysis and screening, supplementing the previously used data tables.

The outputs of the wood industry projects were extracted from the database built on the CORDIS dataset. For the content analysis, data downloaded from the European Patent Office's Espacenet and the World Intellectual Property Organisation's Patentscope service were used.

The analyses used a number of univariate and multivariate methods. Before using multivariate (in our case bivariate) methods such as correlation, linear and non-linear re-

gression, χ^2 test, we sought to test or ensure their assumptions. We searched for univariate and bivariate outliers using various methods, and in addition to the normality of the data, we also examined the normality of the residuals and homoscedasticity.

In the SNA analysis, a number of central indicators were counted, a network of links and clusters was drawn, and notable distributions were examined. The analyses responded to the debate in the literature on inferential statistics. Where it made the least sense/meaning, stochastic statistical methods were used, but it was indicated that the result might be questionable depending on the professional approach. In addition to the commonly used tests (e.g. Kolmogorov-Smirnov, Shapiro-Wilk), less commonly used tests (e.g. Anderson-Darling, Breusch-Pagan, White, Goldfeld-Quandt) were used following examples from the literature.

What may seem like an overabundance of trial and error has sometimes been offset by less fashionable methods (e.g. breaking the scattering square into components). In text mining analyses, RAKE, POS, PMI algorithms were mostly used, but examples of dimensionality reduction and clustering were also given.

The analysis was carried out using Excel, SPSS and JASP software, and some packages of the statistical programming language R.

3. Theses

The 19 professional hypotheses put forward in the paper were tested using a number of statistical hypotheses, related tests or other methods. Among the professional hypotheses, the results of the analysis point clearly towards non-rejection in 14 cases and towards rejection in five cases. Based on our 19 professional hypotheses, we formulated 14 theses.

The professional hypotheses and theses are structured around the following themes:

1. Text mining and scientific metrics analysis of the literature
2. Results of statistical analysis of FP7 and H2020 wood projects
3. Results of the analysis of the wood project network
4. Efficiency and content analysis of project results

Thesis 1

The number of citations in the literature on wood industry projects follows a power function distribution after a certain x_{min} value. So most articles get very few citations, while some articles get very many citations.

Because of the extreme values (Figure 1), a bootstrapping procedure based on 1000 iterations to estimate p and test the distribution was used.

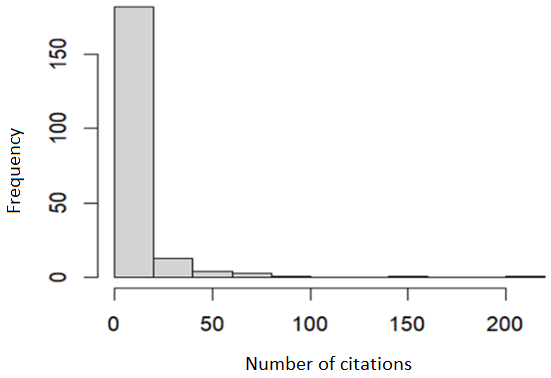

Figure 1. Number of citations of publications related to wood projects

Related publications: [3][4][8][12][18]

Thesis 2

There is a correlation between the number of articles and citations per year dealing with woodworking projects, not only for citations in the given year, but also for citations one or two years later than the publications.

The perhaps unconventional delimitation of the years of publications and citations was justified, on the one hand, by the need to ensure a possible time lag and, on the other hand, by the equal number of elements required for correlation analysis of the two time series under study.

Related publications: [3][4][8][12][18]

Thesis 3

The publications describing wood industry projects available from Scopus do not follow a Bradford distribution and are therefore unlikely to be a true professional representation of publications on the subject.

The ratio of the number of journals in each zone - 24:65:67 - is not close to the value in Bradford's theorem.

Related publications: [3][4][8][12][18]

Thesis 4

The frequency of author productivity in publications describing woodworking projects available in Scopus follows a Lotka distribution.

The β value of the two-sample Kolmogorov-Smirnov test of the Lotka distribution (4.72) differs from the general value estimated by Alfred J. Lotka (2). However, this does not mean that the frequency distribution of author productivity does not follow a Lotka distribution, since the p value (0.52) is significantly higher than the significance level ($\alpha = 0.05$). (This significance level is always chosen in the later sections.)

Related publications: [3][4][8][12][18]

Thesis 5

There is a very strong correlation (strong Spearman rank correlation) between the total cost of each project and community support for FP7 and H2020 wood projects.

The value of the Spearman rank correlation coefficient ($\rho = 0.99$) indicates a very high correlation in FP7 (Figure 2). The value of the Spearman correlation coefficient between the outlier-cleaned data is also very high in H2020 ($\rho = 0.97$) (Figure 3).

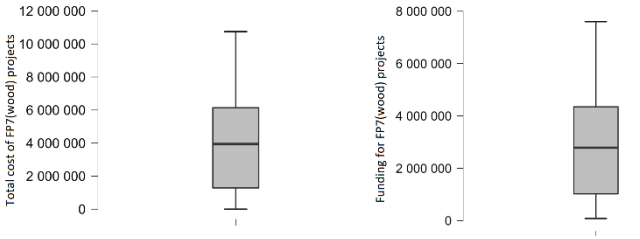

Figure 2. Total cost and funding for FP7 projects in the wood sector (in Euro)

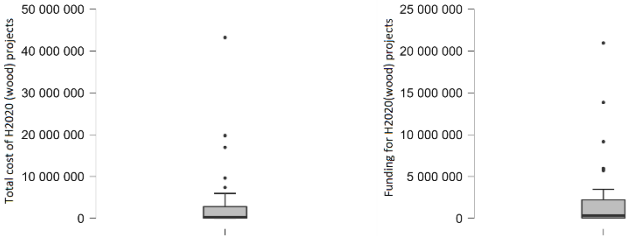

Figure 3. Total cost and funding for H2020 projects in the wood sector (in Euro)

Related publications: [1][5][13][14][15]

Thesis 6

In FP7 and H2020 wood projects, there is a correlation between the frequency of participation per country (Spearman rank correlation stronger than medium).

Using Spearman's rank correlation, a relatively strong (stronger than medium) positive correlation was measured between FP7 (wood industry) and H2020 (wood industry) project participation rates per country ($\rho = 0.719$).

Related publications: [2][9][17]

Thesis 7

The distribution of individual project shares in FP7 and H2020 wood projects in the two programmes and the distribution of project shares per country in the two programmes are derived from the same distribution, i.e. they can be considered to be identically distributed.

The result was supported by Cramér's V index (0.11), Kullback-Leibler divergence (0.03) and Jensen-Shannon divergence (0.008) in the first case, and by a two-sample Kolmogorov-Smirnov test (D = 0.12, p = 0.83) in the second case.

Related publications: [2][4][15][17]

Thesis 8

There is a correlation between the activity and the frequency of participation in FP7 and H2020 wood projects (medium strength according to Cramér's V).

The Cramér V-index is 0.21, reflecting a medium effect, and Pearson's χ^2 is 18.425 (p = 0.001).

Related publications: [2][3][4]

Thesis 9

There is a correlation between the level of community support per country in FP7 and H2020 wood projects. A Spearman rank correlation is stronger than medium and a linear regression is stronger than medium.

Using Spearman rank correlation, a relatively strong (stronger than medium) positive correlation was measured between FP7(wood) and H2020(wood) community support per country (ρ = 0.65). In linear regression, the coefficient of determination R^2 is 0.495. The result indicates a medium-strong relationship (p = 0)

Related publications: [1][2][3][4][5][17]

Thesis 10

The distribution of project funding per country in FP7 and H2020 wood projects comes from the same distribution.

Support was mainly based on a two-sample Kolmogorov-Smirnov test (D = 0.16, p = 0.77).

Related publications: [1][2][3][4][17]

Thesis 11

The link between this activity and the support received under FP7 and H2020 is significant. The difference in support between categories is not due to heterogeneity by criterion but to heterogeneity within categories.

In the analysis, we calculated main averages and sub-averages based on the mixed relationship. The total variance squared was then decomposed into components (external and internal variance squared) and variance squared ratio type indicators were calculated. To show the change between the two aid schemes, the main averages were compared using standardisation.

Related publications: [1][2][3][4][17]

Thesis 12

There is a correlation (Spearman rank correlation stronger than medium) between the funding received per country from FP7 and H2020 and the output of the wood sector.

A production quantity (A and B) is defined for each period by a composite calculation. Spearman's rank correlation coefficient (ρ) for FP7 support to wood projects - Production (A) is 0.57 and for H2020 support to wood projects - Production (B) is 0.70 (Figure 4).

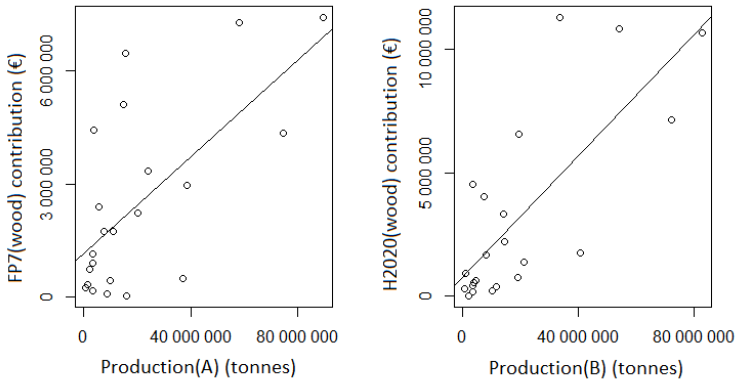

Figure 4. Link between production and community support

Related publications: [1][2][3][4][17]

Thesis 13

The major network clusters of wood project networks have a hierarchical structure.

The largest clusters (Figure 5) almost always have peripheral or insignificant participants in addition to those that are central or at least significant in the network. This highlights the strategy of effective partner search.

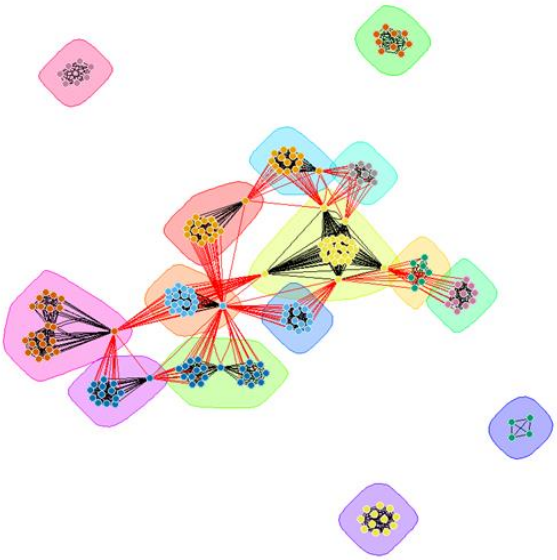

Figure 5. Clusters of FP7 projects in the wood sector

Related publications: [2][3][10][11][15][17]

Thesis 14

The results of FP7 and H2020 projects in the wood sector have produced limited open and closed innovation results compared to the Community resources used, and the overall correlation between funding and results is less than medium in numerical terms.

FP7 projects in the wood sector have produced 8 patents and 111 publications. The H2020 wood projects produced 137 publications, 61 documentations and reports, 19

websites, patents and videos submitted (the latter three in one category), five presentations, experimental models and prototypes (these three also in one category) and three other outputs.

Related publications: [2][3][6][7][10][11][13][14][15][16]

4. Bibliographic Data of Related Publications

4.1. Publications in foreign languages

[1] Báder, M., Németh, R., Vörös, Á., Tóth, Z. & Novotni, A., 2023. The effect of agroforestry farming on wood quality and timber industry and its supportation by Horizon 2020. *Agroforestry Systems*, 2023 (Published: 2023. January 24th), pp. 1–17. https://doi.org/10.1007/s10457-023-00812-8

[2] Novotni, A., Pásztory, Z. & Tóth, Z., 2022. Social Network Analysis in Wood Industry Projects. *Acta Silvatica et Lignaria Hungarica: An International Journal in Forest, Wood and Environmental Sciences*, 18(2), pp. 89–101. https://doi.org/10.37045/aslh-2022-0006

[3] Novotni, A. & Tóth, Z., 2021. Analyzing projects related to sustainable wood. In *International Conference of Economics PhD Students and Researchers in Komarno*. pp. 117–123.

[4] Tóth, Z. & Novotni, A., 2021. Analysis of Wood Industry Project Networks with R. In *III. International Conference of Economics PhD Students and Researchers in Komarno*. pp. 183–189.

[5] Tóth, Z. & Novotni, A., 2021. Demographic Reasons for High Birth Rates in Vienna in March 1946. *RussianStudies.hu*, 3(2), pp. 1–22. https://doi.org/10.38210/RUSTUDH.2021.3.g.1

[6] Novotni, A. & Tóth, Z., 2020. Retro on the furniture market. In *VI. International Winter Conference of Economics PhD Students and Researchers: Conference Proceedings*. pp. 181–187.

[7] Novotni, A. & Tóth, Z., 2020. IKEA and global marketing trends. In *VI. International Winter Conference of Economics PhD Students and Researchers: Conference Proceedings*. pp. 172–180.

[8] Novotni, A. & Tóth, Z., 2020. Innovation models as a basis of efficiency analysis. In *VI. International Winter Conference of Economics PhD Students and Researchers: Conference Proceedings*. pp. 165–171.

[9] Novotni, A., 2020. The role of the research of wood industry waste and by-products in H2020 projects. In *Proceedings of the Miskolc IPW- IV*. Sustainable raw materials international project week. pp. 1–12.

[10] Pásztory, Z. & Novotni, A., 2020. The Utilization of Tree Bark as Thermal Insulation Panels and Formaldehyde Absorber. *Geosciences and Engineering: A Publication of the University of Miskolc*, 8(12), pp. 205–216.

[11] Novotni, A., 2019. Info-communication Trends in Marketing Research. In *Modern gazdaság, okos fejlődés Nemzetközi Tudományos Konferencia*. Sopron, 2019. november 7. – Konferenciakötet / Modern Economy, Smart Development International Scientific Conference. Sopron, 7 November 2019. – Publications. pp. 641–653.

[12] Novotni, A., 2019. A science metric analysis of publications about sustainable wood. In *III. Ring – Fenntartható nyersanyag-gazdálkodás - III. Sustainable Raw Materials Konferenciakötet - Proceedings*. pp. 262–271.

4.2. Publications in Hungarian

[13] Novotni, A., 2021. A Horizont 2020 által támogatott agrárerdészeti projektek II. In „*Termeljünk együtt a természettel! – Az agrárerdészet, mint új kitörési lehetőség*". pp. 213–219.

[14] Novotni, A., 2021. A Horizont 2020 által támogatott agrárerdészeti projektek I. In „*Termeljünk együtt a természettel! – Az agrárerdészet, mint új kitörési lehetőség*". pp. 208–212.

[15] Novotni, A. & Tóth, Z., 2021. Két projekttámogatási eloszlás következményeinek vizsgálata faipari projektekben. In *Közgazdász Doktoranduszok és Kutatók VII. Nemzetközi Téli Konferenciája Konferenciakötet*. pp. 240–248.

[16] Novotni, A., 2020. A fenntarthatóság és az IKEA. In *IX. Interdiszciplináris Doktorandusz Konferencia 2020 [9th Interdisciplinary Doctoral Conference 2020].* pp. 375–386.

[17] Tóth, Z. & Novotni, A., 2020. A Horizont 2020 által támogatott fenntartható faipari projektek. In *Válság és kilábalás: Innovatív megoldások Nemzetközi Tudományos Konferencia Sopron, 2020. november 5. – Konferenciakötet* / Crisis and Recovery: Innovative Solutions International Scientific Conference 5 November 2020, Sopron – Conference Proceedings. pp. 545–551.

[18] Tóth, Z. & Novotni, A., 2019. A fenntarthatósághoz kapcsolódó faipari kutatások tudománymetriai vizsgálata. In *Logisztika-Informatika-Menedzsment Nemzetközi Konferencia 2019*. pp. 67–73.