Indeks 371262
e-ISSN 2657-9545
ISSN 0033-2372

Przegląd Statystyczny
Statistical Review

Vol. 68 | No. 4 | 2021

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

INFORMATION FOR AUTHORS

Przegląd Statystyczny. Statistical Review publishes original research papers on theoretical and empirical topics in statistics, econometrics, mathematical economics, operational research, decision science and data analysis. The manuscripts considered for publication should significantly contribute to the theoretical aspects of the aforementioned fields or shed new light on the practical applications of these aspects. Manuscripts reporting important results of research projects are particularly welcome. Review papers, shorter papers reporting on major conferences in the field, and reviews of seminal monographs are eligible for submission, but only on the Editor's request.

Since 1 May 2019, the journal has been publishing articles in English.

Any spelling style is acceptable as long as it is consistent within the manuscript.

All work should be submitted to the journal through the ICI Publishers Panel (https://editors.publisherspanel.com/pl.ici.ppanel-app-war/ppanel/index).

For details of the submission process and editorial requirements please visit https://ps.stat.gov.pl/ForAuthors.

Przegląd Statystyczny
Statistical Review

Vol. 68 No. 4 2021

EDITORIAL OFFICE'S ADDRESS

University of Gdańsk, ul. Armii Krajowej 101, 81-824 Sopot, Poland

Language editing: Scientific Journals Division, Statistics Poland
Technical editing and typesetting: Statistical Publishing Establishment – team supervised by Wojciech Szuchta

Zakład Wydawnictw Statystycznych
Printed and bound: Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warsaw, Poland, zws.stat.gov.pl

Website: ps.stat.gov.pl

ISSN 0033-2372
e-ISSN 2657-9545
Index 371262

Information on the sales of the journal: Statistical Publishing Establishment
phone no. +48 22 608 32 10, +48 22 608 38 10

Order no. 76/2022 – 225 printed copies

CONTENTS

Przegląd Statystyczny. Statistical Review, 2021, vol. 68, 4, 1–20 DOI: 10.5604/01.3001.0015.7797
Received: 20.12.2021, revised: 9.02.2022, accepted: 15.02.2022

Comparison of methods used for filling partially unobserved contingency tables

Michał Kot,[a] Bogumił Kamiński[b]

Abstract. In this article, we investigate contingency tables where the entries containing small counts are unknown for data privacy reasons. We propose and test two competitive methods for estimating the unknown entries: our modification of the Iterative Proportional Fitting Procedure (IPFP), and one of the Monte Carlo Markov Chain methods called Shake-and-Bake. We use simulation experiments to test these methods in terms of time complexity and the accuracy of searching the space of feasible solutions. To simplify the estimation procedure, we propose to pre-process partially unknown contingency tables with simple heuristics and dimensionality-reduction techniques to find and fill all trivial entries. Our results demonstrate that if the number of missing cells is not very large, the pre-processing is often enough to find fillings for the unknown values in contingency tables. In the cases where simple heuristics are insufficient, the Shake-and-Bake technique outperforms the modified IPFP in terms of time complexity and the accuracy of searching the space of feasible solutions.

Keywords: contingency tables, Markov Chain Monte Carlo, Iterative Proportional Fitting Procedure

JEL: C15, C44

1. Introduction

The study of dependencies between variables is one of the key aspects of data analysis. The procedure of verifying associations between variables depends on the type of variables. In the case of ordinal or nominal variables, it is common to investigate this relationship by their joint frequency distribution provided by a multidimensional contingency table. The elements of such a table (referred to as entries or cells) contain the frequency of joint occurrences of the outcomes of all the underlying variables. Contingency tables, which are the focus of this article, are considered to be a simple and effective technique to analyse data (Payne & Payne, 2011), and are utilised in e.g. social sciences (Payne & Payne, 2011), medical research (Zelterman & Louis, 2019) or biology (Bailey, 1995).

In practical applications, due to confidentiality reasons, data aggregated in the form of a contingency table can be incomplete, and thus not useful. This problem occurs because certain combinations of features included in a contingency table may

[a] SGH Warsaw School of Economics, Institute of Econometrics, Decision Analysis and Support Unit, ul. Madalińskiego 6/8, 02-513 Warszawa, e-mail: mk46032@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-7885-4028.

[b] SGH Warsaw School of Economics, Institute of Econometrics, Decision Analysis and Support Unit, ul. Madalińskiego 6/8, 02-513 Warszawa, e-mail: bkamins@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-0678-282X.

be so rare that they can potentially disclose personal or sensitive data (Slavković, 2010; Slavković & Lee, 2010). In such a case, the end user of the data presented in the form of a contingency table is provided with information only on the marginal distributions of features along with cell sizes big enough to ensure data privacy. The problem with audiences being too small to be reported, common in marketing, is a good example of the above. For instance, Nielsen does not report TV ratings for audiences of sizes below the assumed level (Eastman & Ferguson, 2012), and neither Facebook nor Google show advertisements to custom audiences unless their size requirement is met (Facebook.com, n.d.; Google.com, n.d.). Such practice results from the fact that data providers striving to comply with the General Data Protection Regulation (GDPR) face a trade-off between securing data privacy and ensuring data utility (Slavković, 2010).

We encountered the same problem during the construction of an agent-based model for the calculation of the cross-media reach of advertising (Kot & Kamiński, 2021). When designing the model, we assumed that the synthetic population of agents reflects the existing population of Poland in terms of the agents' socio-demographic features. Since certain dimensions are correlated with others, sampling each feature independently would lead to a bias in the model. Therefore, the necessary condition for the agents' sampling procedure was to obtain a contingency table of features where the probabilities of all combinations would be known. Since we lacked a complete contingency table, we began to seek a solution to this problem.

The literature points to the Iterative Proportional Fitting Procedure (IPFP) as the solution to the problem of estimating unknown entries of contingency tables based on known marginal distributions (Bacharach, 1965; Deming & Stephan, 1940). The IPFP allows the estimation of the sizes of the empty cells that match marginal restrictions by the iterative adjustment of the sizes of missing entries across all the dimensions of the problem, until the marginal totals converge to the target ones. The starting point for the estimation is the construction of a matrix of the initial sizes of all entries, which reflect the prior information available to the researcher. A detailed discussion on the mechanics of the IPFP can be found in Lovelace et al. (2015).

Recent state-of-the-art methods of estimating contingency tables are presented in the literature on issues related to data disclosure. Research in this field was initiated by Diaconis and Sturmfels (1998). In their article, the authors proposed generating contingency tables using the Markov moves. They start with the initial contingency table meeting the known marginal requirements. Then, the Markov moves applied to this table introduce pairwise integer changes to the cells' sizes, resulting in the emergence of a new contingency table, where the marginal requirements remain fulfilled. The new table is accepted if all its entries are non-negative. A finite collection of Markov moves linking contingency tables with the same marginals is called a Markov basis (Dobra, 2003). However, the Markov basis in an explicit form is usually applied to only a few problems because of high computational costs related to its generation (Aoki et al., 2012). Instead of generating the entire Markov basis, Dobra (2012) proposes constructing a Markov chain of locally connected contingency tables. The resulting Markov chain is a subset of all the feasible tables. In this approach, it is possible to introduce additional constraints regarding the lower and upper boundaries of the cells.

Despite the indisputable advantages of the above-mentioned methods, we want to emphasize that three aspects of our problem make it different from the issues usually solved by them.

Firstly, the primary motivation of our research was to use completed contingency tables in the process of sampling the population of artificial agents. To meet this end, we were looking for a method to efficiently generate tables that were uniformly distributed on a set of feasible contingency tables. Sampling a population of agents from the set of contingency tables which differed from each other in terms of the estimated values of the unknown cells improved the results of the simulation study. It is because the conclusions drawn from the simulation results are valid for any feasible combination of features in the population, not only for a series of local, similar contingency tables.

The second unique feature of our problem was that only some elements of the contingency table were unknown. Other elements with a sufficiently high count were observed and had to remain fixed.

The final aspect of our problem was that we had no prior knowledge regarding the missing cells' entry size, except that they were non-negative and of a smaller or equal size to the known value (referred to as threshold). Thus, in our case, the set of feasible solutions was limited by the knowledge of the following features of the problem under study: (1) the marginal distributions of variables captured by the contingency table, (2) the sizes of entries with a sufficiently high count, and (3) the minimal required size of cells to become observable. In the next two paragraphs we will discuss how the unique structure of our problem affected the utility of the discussed algorithms.

In the case of the classic version of the IPFP, all entries in the contingency table are subject to change in each iteration of the algorithm. Therefore, no fixed cells are allowed, except for entries of the value of 0. The algorithm of Diaconis and Sturmfels, which could potentially be used in our case, has two shortcomings. Firstly, all the steps of the procedure are complicated and time-consuming, as they require solving several instances of optimisation problems. Secondly, as the method uses local Markov moves, it does not provide information whether the whole Markov basis was visited after a given number of steps.

At the same time, we are aware that our specific problem is not the only area that these methods may be applied to, i.e. they can be implemented to efficiently solve a wider spectrum of problems.

Since the use of the classic IPFP or the Markov moves would not allow the achievement of the goals listed in the previous paragraph, we decided to test other solutions which allowed the coverage of the searched area of the solution space and the time complexity. In the process of examining the competitive methods, we will use a simulation approach described in Section 3. Firstly, we will generate contingency tables with unobserved cells. Secondly, we will pre-process the contingency table by searching for all trivial cells' sizes (which in a few cases will allow the whole contingency table to be solved). Finally, if a contingency table still contains unobserved cells, we will try to address the situation with the methods listed in the next two paragraphs.

The first method we will test involves our modification of the IPFP algorithm, which has the capacity to find feasible solutions in a contingency table with fixed entries (a description of this modification is presented in detail in Section 2.3).

The second tested method is the Markov Chain Monte Carlo technique known as Shake-and-Bake (SB) (Boender et al., 1991), capable of approximating the distribution of the solution set by an asymptotic sampling of uniform points on a boundary of a convex polytope.

Despite being similar in terms of generating floating-point solutions, there are significant differences between the two competitive methods. In the modified IPFP, we are able to control the matrix of the initial weights only, which enables us to force specific entries' sizes to 0, but we cannot ensure that all the entries will meet the upper boundary requirement. On the other hand, in the SB, we do not assume the existence of any fixed relation between variables. Moreover, to reveal the size of a cell, we can apply prior knowledge regarding the minimum and maximum feasible value (or more generally, some more complex prior knowledge).

In order to test the coverage of the searched area of the solution space, we need a reference solution with all possible fillings of a given contingency table. For objective reasons, we can provide such a solution only for smaller-scale problems through the introduction of a pre-processing stage based on dimensionality reduction methods. We propose to consider exhaustive enumeration of the missing entries of a contingency table as the solution to a set of linear equations. To find all the possible non-negative integer solutions, we will use constraint programming. It is worth emphasizing that exhaustive enumeration and sampling are, along with the computation of sharp integer bounds and counting, the directions in which the research on contingency tables is currently heading (Dobra & Fienberg, 2010).

Our results show how to effectively sample solutions for partially unknown multidimensional contingency tables. This procedure is efficient mostly thanks to two elements of our analysis: the array pre-processing, based on heuristics and inspirations from the hypergraph theory, which results in a significant reduction of the problem's size and time complexity, and a detailed comparison of our modification of the IPFP and SB algorithms in terms of their ability to cover the space of feasible solutions and time required to return a given volume of samples.

The further part of the article consists of: Section 2, where we formulate the problem and describe the details of the tested algorithms, Section 3, where the planning of the simulation experiments and the results of the simulations' run in relation to benchmark methods are shown, and Section 4, in which we discuss the potential limitations of the used methods and propose the directions for their future development.

2. Formulation of the problem and the methods used

In the first part of this section, we formulate the research problem in detail. In the second part, we discuss the pre-processing of the problem. In the third part, we present the method of searching for all the solutions which served as a reference point for the two competitive methods applied in small-scale problems. Finally, we describe the methods that can be used to solve the research problem: in Section 2.4, we present our modification of the standard IPFP approach, and in Section 2.5 the SB algorithm.

2.1. Formulation of the research problem

We assume that a population of an N size is split by K features, with k_i states each. The split generates a Cartesian product (contingency table) with $\prod_{i=1}^{K} k_i$ entries. For simplicity, we assume that feature i takes values from set $V_i = \{1,2,\ldots,k_i\}$. By p_m we denote the size of cell $m \in V_1 \times \ldots \times V_K$. We assume that all p_m are non-negative. We also assume that we know all the marginal totals, which for dimension i are denoted by ${p_v}^i$ for $v \in V_i$. We assume that cells are observable if and only if their size is greater than threshold level T that is known to the researcher. Therefore, in terms of dimensions, the table of observable cells is identical to the complete table, with entries q_m structured as in the following equation:

$$q_m = \begin{cases} p_m \; if \; p_m > T \\ 0, otherwise \end{cases}. \tag{1}$$

Similarly, we denote the marginal distributions of the known cells for value $v \in V_i$ in dimension i as ${q_v}^i$, and by subtracting them from ${p_v}^i$, we obtain marginal boundaries for unobserved cell sizes $r_v{}^i = p_v{}^i - q_v{}^i$. An example of the calculation of $r_v{}^i$, $p_v{}^i$ and $q_v{}^i$ can be found in Table 1. The goal of the algorithm we want to develop is to find all values of $\hat{p}_m$ that solve the system of linear equations:

$$\begin{aligned} \sum_{m \in R_V{}^i} \hat{p}_m &= r_v{}^i \\ \hat{p}_m &\geq 0 \\ \hat{p}_m &\leq T \\ \hat{p}_m &\in \mathbb{Z} \end{aligned} \quad . \tag{2}$$

In Equation (2), $R_V{}^i$ is a set of indices in $m \in V_1 \times \ldots \times V_K$ such that $\forall m \in R_V{}^i : q_m = 0$ and in the i-th dimension they are of v value. Based on the example shown in Table 1, the unknown elements are $\hat{p}_{2,2}$, $\hat{p}_{2,3}$, $\hat{p}_{3,2}$ and $\hat{p}_{3,3}$.

Table 1. Example of the $p_v{}^i$, $q_v{}^i$ and $r_v{}^i$ calculation for $T = 15$

				$p_v{}^1$	$q_v{}^1$	$r_v{}^1$
	30	20	20	70	70	0
	25	$\hat{p}_{2,2}$	$\hat{p}_{2,3}$	40	25	15
	20	$\hat{p}_{3,2}$	$\hat{p}_{3,3}$	35	20	15
$p_v{}^2$	75	30	40			
$q_v{}^2$	75	20	20			
$r_v{}^2$	0	10	20			

Source: authors' calculations.
Note. values $p_{2,2}$, $p_{2,3}$, $p_{3,2}$, and $p_{3,3}$ are not observed.

The marginal restrictions for this example are presented in Equation (3):

$$\begin{aligned} \hat{p}_{2,2} + \hat{p}_{2,3} &= 15 \\ \hat{p}_{3,2} + \hat{p}_{3,3} &= 15 \\ \hat{p}_{2,2} + \hat{p}_{3,2} &= 10 \\ \hat{p}_{2,3} + \hat{p}_{3,3} &= 20 \end{aligned} \quad . \tag{3}$$

2.2. Pre-processing of the problem

To reduce the search space of the calculations, we introduce a pre-processing stage to be run before attempting to identify $\hat{p}_m$. Its main purpose is to find all trivial solutions to a given problem and therefore to accelerate the calculations. During the pre-processing stage, we iteratively search through the table of observable cells for:

- single missing cells in some dimensions (i.e. the cases where $|R_V{}^i| = 1$), where we input r_v in these dimensions;
- dimensions in which unknown entries exist and the marginal sum is 0, where we input 0 in each missing entry;
- dimensions in which the number of missing cells multiplied by the threshold value is equal to the marginal sum of the unknown cells (i.e. $|R_V{}^i| \cdot T = r_v{}^i$), where we input the threshold value in each missing entry.

Moreover, we extend the pre-processing stage described above to further simplify the calculations. Inspired by the hypergraph theory (Bretto, 2013), we treat the set of the remaining missing cells $\{m: q_m = 0\}$ as vertices of a hypergraph. The hyperedges of the hypergraph are defined by sets $R_V{}^i$. We seek a set of vertices V that separates the underlying total hypergraph into independent subhypergraphs. We assume set V to be minimal, i.e. the removal of any element of V results in V no longer separating the hypergraph. Then, the potential values of the cells in the separated subhypergraphs are conditioned on the values of the separating vertices in V. This approach significantly limits the solution space. Below we provide examples of how this procedure simplifies the problem and accelerates the computation process.

The first example of such a situation is presented in Table 2, where the threshold level is equal to $T = 20$. Without the separation, it can be computed that an algorithm should find feasible solutions among 441 potential combinations. One can notice that the p_{32} vertex can be used to separate the problem into two distinct problems that are easier to solve. Furthermore, only $p_{32} = 10$ allows all marginal constraints to be met and thus the size of the feasible set is reduced to 42 combinations.

Table 2. A contingency matrix where p_{32} is a separating vertex

p_{11}	p_{12}	30	30	80
p_{21}	p_{22}	30	30	80
30	p_{32}	p_{33}	p_{34}	60
30	30	p_{43}	p_{44}	80
80	60	80	80	300

Source: authors' calculations.

The second example is presented in Table 3, where the threshold level is equal to $T = 30$. It can be computed that without the separation, it would require finding feasible solutions among 1,224,912 potential combinations. The assumption that $p_{4,3} = 10$ (which, as above, can be proven to be the only admissible entry) results in an 80%-reduction of a combination set, to 122,491×2 elements, which in practice is much easier to compute.

Please note that in certain situations the pre-processing stage can solve the array and fill all missing entries; if it is not able to solve the array, the missing cells form a subarray of at least 2 elements across each dimension (the smallest possible is a 2×2 subarray).

Table 3. A contingency matrix where p_{43} is the separating vertex

p_{11}	p_{12}	p_{13}	40	40	40	150
p_{21}	p_{22}	p_{23}	40	40	40	150
p_{31}	p_{32}	p_{33}	40	40	40	150
40	40	p_{43}	p_{44}	p_{45}	p_{46}	120
40	40	40	p_{54}	p_{55}	p_{56}	150
40	40	40	p_{64}	p_{65}	p_{66}	150
150	150	120	150	150	150	870

Source: authors' calculations.

2.3. A method searching for all solutions

In this subsection, we discuss the algorithm used as a reference for the tested methods, which relies on finding all feasible integer solutions to a problem expressed by Equation (2). Since the feasible set of solutions is in this case limited by marginal sums across all dimensions, one can think of solving the contingency table as of solving a set of linear equations. In practice, to solve the linear system, we use software that can be used to solve constraint programming problems called MiniZinc (Nethercote et al., 2007).

Since the exhaustive enumeration of all solutions is highly time-consuming even for specialised software, we will use it as a reference for small-scale problems only.

An example presented in Table 4 is used to show the impact of different threshold levels on the estimation process. As we have previously assumed, the threshold level is known, i.e. the researcher is aware that all the sizes of unknown entries are smaller than some fixed value T. If the threshold was unknown, the presented problem would have 9 feasible solutions presented in Table 5. If the threshold level was set at 9, only 7 results would be valid, because answers #8 and #9 would contain elements greater than the threshold value. By the same token, for threshold levels 10 and 11, there would be 8 and 9 solutions, respectively.

Table 4. A contingency matrix without known cells

p_{11}	p_{12}	8
p_{21}	p_{22}	12
9	11	20

Source: authors' calculations.

Table 5. Solutions to the problem presented in Table 4

ID	p_{11}	p_{12}	p_{21}	p_{22}
#1	0	8	9	3
#2	1	7	8	4
#3	2	6	7	5
#4	3	5	6	6
#5	4	4	5	7
#6	5	3	4	8
#7	6	2	3	9
#8	7	1	2	10
#9	8	0	1	11

Source: authors' calculations.

2.4. Modified IPFP method

To present how we have modified the classic IPFP, we should start with discussing the mechanics of the classic version of this algorithm based on an example for the 2×2 contingency table presented in Table 6. Consider an unknown contingency table and assume that the target row and the column totals are (25, 35) and (20, 40), respectively. We have no prior knowledge about the entry sizes, thus we initially assume them to be equal to 1. The first step of each IPFP iteration creates multipliers to adjust the sum of the row elements in an entry table to the marginal totals. Then, the IPFP repeats the procedure with the column totals. In an extended case of a multidimensional array, multipliers are created along each dimension and the procedure follows the two-dimensional example. In consecutive iterations, the same procedure is repeated until the algorithm converges, i.e. the values in the cells change from iteration to iteration by less than the assumed tolerance level. The convergence of the IPFP algorithm was proven by Fienberg (1970) in the case of strictly positive tables. As regards non-negative tables where certain cells equal 0, convergence was proven by Csiszár (1975). In the presented example, the IPFP converges after 3 iterations, and the initial and final steps are presented in Tables 6 and 7, respectively.

As mentioned before, the problem considered in this paper is different from those normally solved by the classic IPFP. Therefore, we had to modify its normal specification to be able to use this method. Initially, we know all the fixed cells' sizes and the marginal totals. In the first step, we subtract the sums of the known cells across each dimension from the known marginal totals in order to obtain the sums for the unknown cells. In the second step, we create an array of the initial values – if the cell is known, then the initial value will be positive or otherwise equal to 0, resulting in an array with non-negative values. Our objective is to sample the possible fillings. A single run of a modified IPFP algorithm with the same initial weights would produce similar results. Thus, in our case, we propose applying a multi-start of the algorithm with random initial weights from the standard uniform distribution.

The adjustment of the cell values across all dimensions is performed in a similar way to the standard version of the IPFP algorithm, yet the target values are the marginal sizes of the unknown entries. The algorithm is completed when the convergence requirement is met, i.e. the highest multiplier across all dimensions does not exceed the assumed value. Since our initial array is non-negative and contains 0, the proof of the convergence of the modified IPFP algorithm follows from Csiszár (1975).

Table 6. Example of classic IPFP mechanics – initial table

1	1	25
1	1	35
20	40	60

Source: authors' calculations.

Table 7. Example of classic IPFP mechanics – final table

$8\frac{1}{3}$	$16\frac{2}{3}$	25
$11\frac{2}{3}$	$23\frac{1}{3}$	35
20	40	60

Source: authors' calculations.

2.5. Shake-and-Bake simulated result

The second competitive method that we use for solving problems with a high number of missing values is Shake-and-Bake, which is based on the Monte Carlo Markov Chain technique (Boender et al., 1991).

SB generates uniform points from the boundary of convex polytope L, limited by a set of m linear inequalities $\boldsymbol{Ax} \leq \boldsymbol{b}$. The method also assumes lack of redundant equations, and the boundary of polytope ∂L is defined as a set of points, for which only one inequality constraint is active at a time as per Equation (4):

$$\partial L = \bigcup_{i=1}^{m} \{\boldsymbol{x}: \boldsymbol{a}_i{}^T \boldsymbol{x} = b_i, \boldsymbol{a}_j{}^T \boldsymbol{x} = b_j, \forall j \neq i\}. \quad (4)$$

The mechanics of the SB algorithm can be presented in a few steps. The beginning of the algorithm is an arbitrary point on the boundary of the polytope. Then, random feasible search direction vector d is sampled, and the algorithm jumps from the starting point in a direction defined by d to a target point. The target point is defined as the closest to the starting point intersection of the line defined by the starting point, search direction vector and one of the polytope boundaries. All iterations undergo the described process. The detailed presentation and discussion of the SB algorithm can be found in Kroese et al. (2011).

Note that the problem described by Equation (2) meets the conditions of the SB algorithm if we remove the constraint due to the fact that the solutions must be an integer.

In this section, we defined the research problem and described the pre-processing of an incomplete contingency table that helps to simplify the research problem. We introduced three methods that will be used in simulations: (1) the method which enumerates all feasible integer fillings used as a reference point, (2) our modification of the IPFP, and (3) the SB method. Despite the fact that two of the tested methods sample feasible float fillings, their mechanics differ significantly. The modified IPFP applies a multi-start approach, which in each iteration starts from different uniform weights and follows a procedure of iterative adjustment across all of the array's marginals (as described in Section 2.4) to return a feasible solution. The SB approach, on the other hand, creates a convex polytope that represents the problem's constraints and iteratively jumps on its boundaries to produce a sample of feasible fillings. What is more, the modified IPFP allows the setting of only the lower boundary for solutions, while in SB both the lower and upper boundaries can be set.

3. Simulation setup and results

In this section, we use simulations to test the ability of the modified IPFP, and the SB methods to cover the searched area of the solution space and time complexity. To compare the coverage of the feasible area, we need a reference solution – the result of the method described in Section 2.3 listing all the possible integer fillings. Since the number of solutions grows rapidly with the increasing size of the contingency table, we run this test on two-dimensional arrays only, with a limited number of missing cells. In the first part of this section, we describe the simulation procedure, and in the second we present the results.

3.1. Simulation procedure

The method used to compare the performance of the competitive techniques firstly involved the sampling of the multidimensional contingency table. We assumed that the table would contain information about an artificial population of N units. Each unit was described by K nominal variables, with k_i states in each variable. Firstly, we sampled a random realisation of a standard uniform distribution for each cell of the table, which served as the weight of a given combination of features. Then, each unit of the population was randomly assigned to a given cell proportionally to the weights, sampled in Step 1. Therefore, each unit of the population was put into a single cell of the contingency table. Finally, we chose threshold T as percentile c of the joint distribution stored in the contingency table, and all cells of sizes smaller or equal to T were marked as missing. As a result of the above procedure, we received the known marginal distributions of a full array, the incomplete array of observable cells, and the threshold.

We planned the simulation procedure in such a way as to capture the coverage of the search area and the time required by each method to generate 1,000 solutions according to various parameters. We tested different parameters, including: (1) the marginal sizes k_i of the contingency tables (the number of the distinct states of a nominal variable), (2) the number of dimensions (features) K, (3) the number of units in population N, and (4) the threshold level in the form of a percentile of joint distribution c. Having managed not to lose generality, we sampled only hypercube contingency tables with equal sizes on each margin ($k_i = k$). We split the simulation procedure into three independent experiments, each focusing on different aspects of the investigated problem.

The aim of the first experiment was to analyse the pre-processing procedure. The purpose of the second experiment was to investigate the time complexity of the two competitive methods. Finally, in the third experiment, we compared the coverage of the searched area of the feasible solutions. Table 8 presents the ranges of the parameters which were set for each experiment.

Table 8. Ranges of simulation parameters

Parameter	Experiment 1	Experiment 2	Experiment 3
k_i	4, 5, 6, 7, 8, 9	6, 7, 8, 9	8
K	2, 3, 4	2, 3	2
N	10* k_i *K	10* k_i *K	640
c	0.10, 0.15, 0.20, 0.25, 0.30	0.25, 0.30	0.20
Number of simulations	11,000	2,000	1,000

Source: authors' calculations.

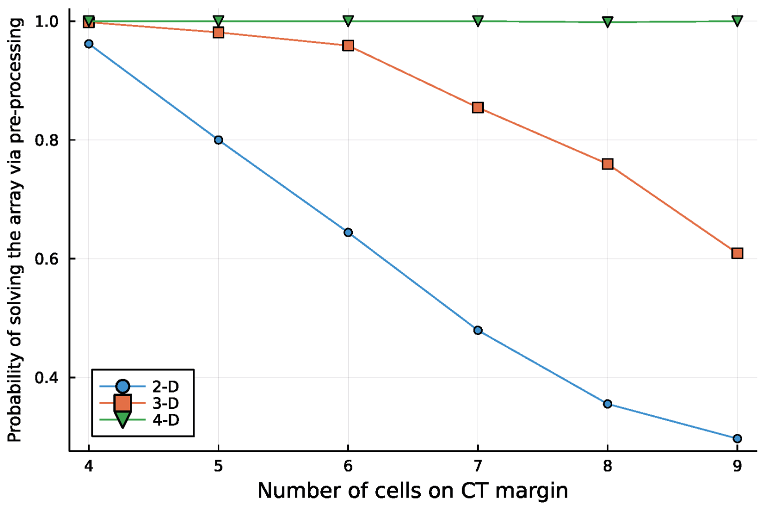
A single iteration run consisted, in the case of Experiment 1, of three steps, while for Experiments 2 and 3 it involved four steps. In the first step, we sampled k_i, K and c. In the second step, we sampled a contingency table for parameters k_i and K obtained in Step 1, and trimmed the entries below the threshold level, defined by c. In the third step, we pre-processed the incomplete contingency table to input all the trivial fillings. In the fourth step, we used the modified IPFP (with initial weights being randomly uniform) and SB algorithms to sample 1,000 possible fillings of the remaining unknown entries. In Experiment 3, we also searched for all the possible integer fillings to investigate the coverage of the searching of two sampling-based methods. In each experiment, we ran a simulation several times (please refer to row *Number of simulations* in Table 8) to obtain stable results. It is worth remembering that the modified IPFP and SB sample floating-point numbers. To compare their results with the integer solutions, we rounded them and ensured that the rounding does not bring solutions outside of the feasible set.

The simulations have been conducted in Julia language (Bezanson et al., 2017), where we developed all the methods used in this paper: the modified IPFP, the exact method incorporating MiniZinc (Nethercote et al., 2007) and SB using R language (R Core Team, 2018), and the hitandrun package (van Valkenhoef & Tervonen, 2019).

3.2. Simulation results

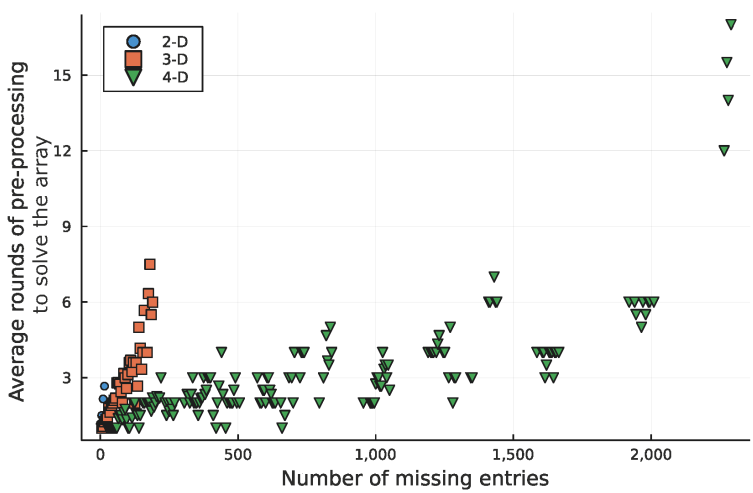
The results obtained in the first experiment showed a relation between the problem's size (defined by the number of array dimensions) and the probability that the pre-processing managed to solve it. We observed that a higher number of dimensions increases the probability that certain unknown entries could be revealed with simple heuristics of a pre-processing stage. For example, if the marginal size is equal to 7, the probability that pre-processing will solve a two-dimensional table is 0.48, while for three- and four-dimensional tables it is 0.85 and 1.00, respectively. At the same time, more complex problems with a higher number of unknown entries require a higher number of pre-processing rounds to be solved. On average, two-dimensional problems require 1.58 rounds to be solved, while three- and four-dimensional ones require 2.35 and 2.58 rounds, respectively. Figures 1 and 2 present the results of Experiment 1, which confirm the two above observations.

Figure 1. Probability of solving the array by pre-processing



Source: authors' calculations.

Figure 2. Number of pre-processing rounds necessary to solve the array

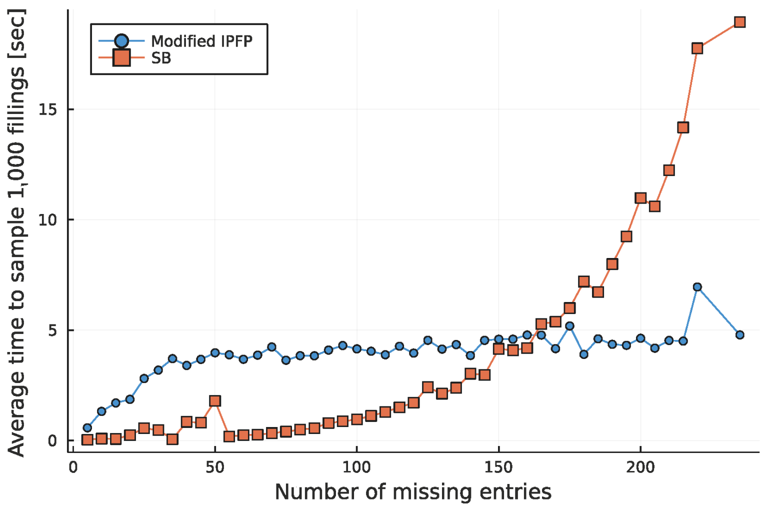


Source: authors' calculations.

In the second experiment, we tested two competitive methods, i.e. our modified version of the IPFP and SB, in terms of time complexity. The results, presented in Figure 3, show that SB requires less time to generate 1,000 samples than the modified

IPFP needs to solve problems of a small or moderate size (up to 150 unknown entries). As the problem's size increases, the amount of time required for the modified IPFP to solve the problem increases, but only up to 50 missing entries, and it remains on a similar level for more complex problems. In the case of SB, time complexity grows exponentially along the problem's complexity. Since complex problems with a high number of missing counts occur less often due to the pre-processing stage, the average time for the modified IPFP to sample 1,000 solutions was 2.76 seconds, while for SB it was 1.01 seconds.

Figure 3. Time complexity of the modified IPFP and SB



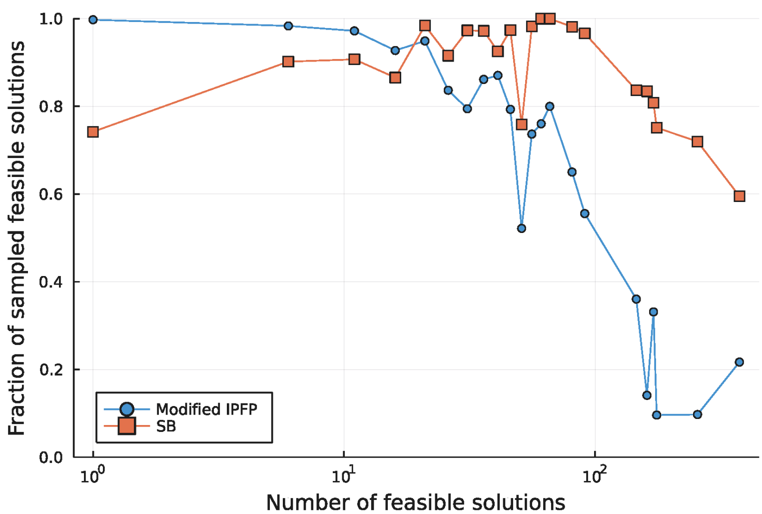
Source: authors' calculations.

The third experiment was designed to compare the ability of both methods to effectively search through the space of feasible solutions. An optimal method should be able to meet two requirements: to sample each possible filling and to generate a uniform sample where all fillings occur equally often.

We present the results of an experiment relating to the first requirement in Figure 4. The modified IPFP can sample all feasible solutions for problems to which there are fewer than 50 solutions. SB outperforms the modified IPFP, as it can sample all solutions for problems with 100 feasible solutions. In the case of problems with over 100 feasible solutions, however, none of the above methods is able to effectively sample all possible fillings.

As regards the second requirement, we propose testing the sampling uniformity with two metrics, the first of which is the ratio of the number of occurrences of the solutions sampled most often to the number of occurrences of solutions sampled least often, as displayed in Figure 5. For example, if the ratio is equal to 10, a given algorithm samples the most frequent solution ten times as often as the least frequent one. In the case of the modified IPFP and SB, the first uniformity metric increases as the number of feasible solutions grows up to 30 and 50, respectively, and then sharply decreases. The shapes of both curves result from the fact that for problems with a low number of feasible fillings, it is more probable that both will be sampled, and hence the first uniformity metric accounts for low values. On the other hand, when the number of feasible solutions is high, each solution is sampled only a few times, and therefore the ratio of the most frequent occurrences to the least frequent ones is naturally limited. On average, SB registers the first uniformity metric 36% more efficiently than the modified IPFP.

Figure 4. Fraction of sampled feasible solutions



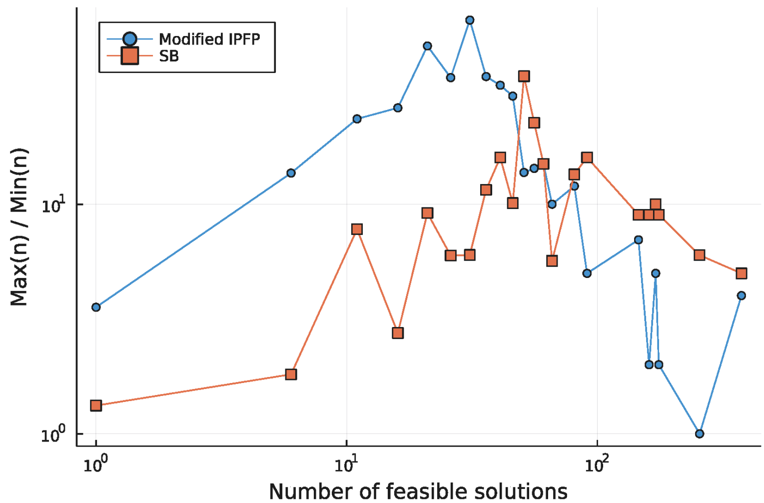
Source: authors' calculations.

Secondly, we test the uniformity with Total Variation Distance δ, defined as in Equation (5) (Levin et al., 2009), where a and b are probability distributions on D (shown in Figure 6). In our case, probability distribution a is an empirical distribution returned by the sampling algorithm, and b is a uniform distribution:

$$\delta(a, b) = \frac{1}{2} \sum_{x \in D} |a(x) - b(x)|. \tag{5}$$

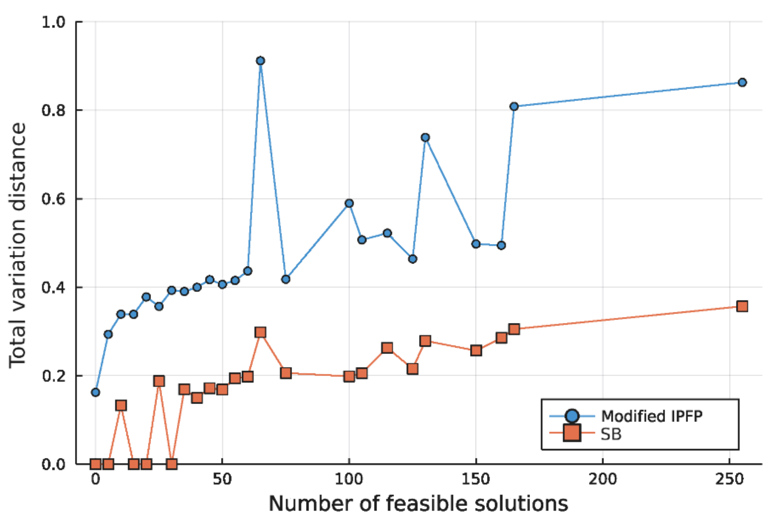
When the uniformity of the sampled solutions is measured with the total variation distance, the SB method outperforms the modified IPFP, regardless of the size of the problem measured with the number of feasible solutions. Although the total variation distance grows for both methods as problems become increasingly more complicated, for SB it never exceeds 0.40, while for the modified IPFP it reaches the level of 0.85.

Figure 5. Uniformity of the sampled solutions – the ratio of the most common solution to the least common solution



Source: authors' calculations.

Figure 6. Uniformity of the sampled solutions – total variation distance



Source: authors' calculations.

4. Conclusions

In this article, we investigated the problem of partially unknown contingency tables. In such arrays, the problem of unknown entries is caused by data privacy requirements, and thus cells with low counts are not reported. We presented our modification of the classic IPFP algorithm and proposed a simulation method incorporating the Shake-and-Bake algorithm. In addition to these methods, we developed a list of heuristics and dimensionality-reduction techniques which, if applied first, simplify the problem and search for all trivial fillings. We conducted a series of experiments to compare both methods in terms of their ability to effectively search through the space of feasible solutions and time complexity.

Our results show that with the increasing dimensionality of the contingency table, the probability that simple heuristics could solve all missing entries rises. In the case of moderately-sized problems, pre-processing required on average 2.3–2.6 rounds to find a solution. Wherever pre-processing was not able to solve the contingency table, we used two competitive methods. In terms of time complexity, our results show that SB outperforms the modified IPFP algorithm when solving smaller problems (of the number of missing entries lower or equal to 150). On average, the time required for SB to sample 1,000 solutions was lower by 64% than the time required for the modified IPFP to do the same. In terms of the ability to search through the space of solutions, SB was able to find 85% of the feasible solutions, while the modified IPFP was able to locate 78%. Moreover, SB samples are characterised by a greater uniformity, which was proven by two different metrics: the ratio of the most frequent solution to the least frequent solution and the total variation distance. We can therefore conclude that SB outperforms the modified IPFP algorithm, as it offers a lower time complexity and a more thorough search of the space of feasible solutions.

In terms of the future development of this study, we plan to improve the functionality of the simulated approach described herein, so as to make possible the sampling from the lattice (Xie et al., 2017). In addition, since the algorithms are designed to solve problems based on sample data, we would like to add an adjustment for population marginal totals (as an optional step).

Acknowledgements

The project has been financed by the Polish National Agency for Academic Exchange. The authors would like to thank Maciej Beręsewicz, PhD, for his substantial help in choosing literature relevant to the presented analysis.

References

Aoki, S., Hara, H., & Takemura, A. (2012). Running Markov Chain Without Markov Bases. In S. Aoki, H. Hara & A. Takemura (Eds.), *Markov Bases in Algebraic Statistics* (pp. 275–286). Springer. https://doi.org/10.1007/978-1-4614-3719-2_16.

Bacharach, M. (1965). Estimating Nonnegative Matrices from Marginal Data. *International Economic Review*, *6*(3), 294–310. https://doi.org/10.2307/2525582.

Bailey, N. T. J. (1995). *Statistical methods in biology* (3rd edition). Cambridge University Press. https://doi.org/10.1017/CBO9781139170840.

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671.

Boender, C. G. E., Caron, R. J., McDonald, J. F., Kan, A. H. G. R., Romeijn, H. E., Smith, R. L., Telgen, J., & Vorst, A. C. F. (1991). Shake-and-Bake Algorithms for Generating Uniform Points on the Boundary of Bounded Polyhedra. *Operations Research*, *39*(6), 945–954. https://doi.org/10.1287/opre.39.6.945.

Bretto, A. (2013). *Hypergraph Theory*. Springer. https://doi.org/10.1007/978-3-319-00080-0.

Csiszár, I. (1975). *I*-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, *3*(1), 146–158. https://doi.org/10.1214/aop/1176996454.

Deming, W. E., & Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics*, *11*(4), 427–444. https://doi.org/10.1214/aoms/1177731829.

Diaconis, P., & Sturmfels, B. (1998). Algebraic Algorithms for Sampling from Conditional Distributions. *Annals of Statistics*, *26*(1), 363–397. https://doi.org/10.1214/aos/1030563990.

Dobra, A. (2003). Markov bases for decomposable graphical models. *Bernoulli*, *9*(6), 1093–1108. https://doi.org/10.3150/bj/1072215202.

Dobra, A. (2012). Dynamic Markov Bases. *Journal of Computational and Graphical Statistics*, *21*(2), 496–517. https://doi.org/10.1080/10618600.2012.663285.

Dobra, A., & Fienberg, S. E. (2010). The Generalized Shuttle Algorithm. In P. Gibilisco, E. Riccomagno, M. P. Rogantin & H. P. Wynn (Eds.), *Algebraic and Geometric Methods in Statistics* (pp. 135–156). Cambridge University Press. https://doi.org/10.1017/CBO9780511642401.

Eastman, S. T., & Ferguson, D. A. (2012). *Media Programming: Strategies and Practice* (9th edition). Wadsworth, Cengage Learning.

Facebook.com. (n.d.). *Facebook lookalike audiences*. Retrieved January 23, 2021, from https://www.facebook.com/business/a/custom-to-lookalike-audiences.

Fienberg, S. E. (1970). An Iterative Procedure for Estimation in Contingency Tables. *The Annals of Mathematical Statistics*, *41*(3), 907–917. https://doi.org/10.1214/aoms/1177696968.

Google.com. (n.d.). *About similar segments on the Display Network*. Retrieved January 23, 2021, from https://support.google.com/google-ads/answer/2676774?hl=en.

Kot, M., & Kamiński, B. (2021). Agent Based Model of Cross Media Reach of Advertising. In P. Ahrweiler & M. Neumann (Eds.), *Advances in Social Simulation: Proceedings of the 15th Social Simulation Conference: 23–27 September 2019* (pp. 45–57). Springer. https://doi.org/10.1007/978-3-030-61503-1_5.

Kroese, D. P., Taimre, T., & Botev, Z. I. (2011). *Handbook of Monte Carlo Methods*. John Wiley & Sons. https://doi.org/10.1002/9781118014967.

Levin, D. A., Peres, Y., & Wilmer, E. L. (2009). *Markov Chains and Mixing Times* (2nd edition). American Mathematical Society. http://dx.doi.org/10.1090/mbk/058.

Lovelace, R., Birkin, M., Ballas, D., & van Leeuwen, E. (2015). Evaluating the performance of Iterative Proportional Fitting for spatial microsimulation: new tests for an established technique. *Journal of Artificial Societies and Social Simulation*, *18*(2), 1–25. http://dx.doi.org/%2010.18564/jasss.2768.

Nethercote, N., Stuckey, P. J., Becket, R., Brand, S., Duck, G. J., & Tack, G. (2007). MiniZinc: Towards a Standard CP Modelling Language. In C. Bessière (Ed.), *Principles and Practice of Constraint Programming – CP 2007* (pp. 529–543). Springer. https://doi.org/10.1007/978-3-540-74970-7_38.

Payne, G., & Payne, J. (2011). *Key Concepts in Social Research*. SAGE Publications. https://dx.doi.org/10.4135/9781849209397.

R Core Team. (2018). *R: A language and environment for statistical computing*. Retrieved January 23, 2021, from https://www.R-project.org.

Slavković, A. B. (2010). Partial Information Releases for Confidential Contingency Table Entries: Present and Future Research Efforts. *Journal of Privacy and Confidentiality*, *1*(2), 253–264. https://doi.org/10.29012/jpc.v1i2.577.

Slavković, A. B., & Lee, J. (2010). Synthetic two-way contingency tables that preserve conditional frequencies. *Statistical Methodology*, *7*(3), 225–239. https://doi.org/10.1016/j.stamet.2009.11.002.

van Valkenhoef, G., & Tervonen, T. (2019, January 8). *'Hit and Run' and 'Shake and Bake' for Sampling Uniformly from Convex Shapes*. https://cran.r-project.org/web/packages/hitandrun/hitandrun.pdf.

Xie, C., Zhong, W., & Mueller, K. (2017). A Visual Analytics Approach for Categorical Joint Distribution Reconstruction from Marginal Projections. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 51–60. https://doi.org/10.1109/TVCG.2016.2598479.

Zelterman, D., & Louis, T. A. (2019). Contingency Tables in Medical Studies. In J. C. Bailar & F. Mosteller (Eds.), *Medical Uses of Statistics* (2nd edition; pp. 293–310). CRC Press. https://doi.org/10.1201/9780429187445.

Przegląd Statystyczny. Statistical Review, 2021, vol. 68, 4, 21–44 DOI: 10.5604/01.3001.0015.7798
Received: 30.12.2021, revised: 23.02.2022, accepted: 25.02.2022

Influence of previous experience and socioeconomic characteristics on willingness to pay for physiotherapy in Poland

Justyna Proniewicz[a]

Abstract. The healthcare system in many countries is characterised by the co-existence of public and private medical services. Patients' decisions regarding the purchase of private health services are made taking into account the trade-off between the price of a treatment and its quality and the waiting time. The aim of this study is to find out which factors impact the willingness to pay for health insurance or the willingness to pay for medical treatment. The study demonstrates that besides socio-demographic characteristics, some negative experiences (e.g. unavailable treatments, long waiting times, long journeys involved) and the experience of already having paid for treatments impact the willingness to pay. The results suggest that negative experiences are likely to cause a change in patients' habits.

Keywords: willingness to pay, preferences, rehabilitation, health care financing, public health insurance
JEL: D12, D91, I11, I13

1. Introduction

When stating what their willingness to pay for healthcare services is, people often have to trade off the quality of medical treatment and the time they need to wait to receive it against the price of the services. Another issue is that the quality of services is not fully known before the purchase. Also, not all health services are the same, nor can all of them be provided at the same time by the same provider. Therefore, patient's willingness to pay for the general healthcare services and for the access to a more specific type of treatment might not be the same. Physiotherapy is one of those health services that are usually needed urgently, due to pain or discomfort. Moreover, a single series of treatment is often not enough to fully recover. Lack of relatively quick access to physiotherapy may result in chronic ailments. At the same time, there is a general sentiment among physiotherapists in Poland that they are paid too little. They often prefer to work at private medical facilities where they are offered better salaries. Such decisions affect the availability of physiotherapy in public facilities. The whole situation poses a question about how much patients value the access to physiotherapy; in other words, are they willing to pay for physiotherapy treatment, and if yes, how much? These questions might not be easy to answer. Physiotherapy services are specific in many ways, and possibly some changes in the

[a] SGH Warsaw School of Economics, Collegium of Economic Analysis, Institute of Econometrics, Department of Support and Decision Analysis, ul. Madalińskiego 6/8, 02-513 Warszawa,
e-mail: jp53741@doktorant.sgh.waw.pl, ORCID: https://orcid.org/0000-0002-6241-7341.

scheme for financing them should be introduced. This study focuses only on physiotherapy (not any other type of healthcare services).

Recently the average waiting time for an appointment within the public healthcare system in Poland has significantly increased, which fuelled the demand for commercial healthcare services (Najwyższa Izba Kontroli [NIK], 2014). As it is only possible to assess the quality of such services after they have been bought and used, patients' opinions about previous medical treatments might influence their future decisions in this respect. My research shows the elements of the past experience which are likely to affect future health-related purchases. This kind of analysis provides a broad view on the willingness to pay for physiotherapy, influenced not only by patients' demographic characteristics, but also by their previous decisions and their outcomes. The analysis focuses mainly on two effects influencing or the willingness to pay for healthcare treatment the willingness to pay for health insurance: the income effect and the impact of past experience.

The effect of experience results from the overall knowledge an experienced patient has. As mentioned before, a person who has already received several physiotherapy treatments is likely to use his or her knowledge about their quality and effectiveness while making similar decisions in the future. Each treatment received gives some additional information, i.e. arouses either positive or negative emotions, involves spending either substantial sums of money or no costs, is smoothly delivered or there are some issues. Following the Bayesian Updating theorem (Viscusi, 1985), this additional information is likely to influence one's valuation of a medical service through the freshly-acquired knowledge about its quality.

A person who remembers the procedures necessary to receive physiotherapy treatment within the public healthcare system as problematic might be reluctant to try them again, and might decide to try commercial physiotherapy services instead. There might be a few types of problems on the way to getting a physiotherapy treatment. Some treatments might be unavailable in the closest healthcare facility (lack of equipment or trained personnel), long waiting time might be involved, or a patient would need to travel far to receive services of a sufficiently high quality. Another factor here is that patients' habits might be difficult to change. For example, if somebody got accustomed to paying for physiotherapy, there is a chance he or she will continue paying in the future. A patient who knows the current treatment prices and is aware what quality and outcomes of this kind of treatment he or she might expect is likely to value future treatments higher than an unexperienced patient. A reverse situation might be expected in the case of patients who are used to receiving physiotherapy treatments through the public healthcare system, namely they are likely not to be willing to pay for such treatments (having been used to receving them for free).

To measure the effect of experience, the analysis of the influence of the physiotherapy type received most recently on the willingness to pay was performed. This kind of analysis was necessitated by the fact that there are various types of physiotherapy treatments delivered and priced differently. Some kinds of physiotherapy are meant to reduce pain or discomfort. Some are offered rarely because they require highly specialised personnel or special equipment. Some are only partly financed by the National Health Fund (Pol. Narodowy Fundusz Zdrowia – NFZ) due to their high price. All the above shows that the patient's overall experience related to physiotherapy might be influenced by the type of treatment. The study deals with three main kinds of physiotherapy: kinesitherapy, physical therapy and massage.

The effect of past experience is definitely worth considering and testing, especially in the context of the Polish healthcare system. Scarce publications focus thoroughly on its aspects. But it is equally worthwhile to measure the impact of the number of treatments received (separately within the public and the private healthcare systems) on the patients' willingness to pay for physiotherapy treatment. In this study, I tried to address all the above-mentioned areas related to one's experience with healthcare services, namely: previous habits and preferences (e.g. private or public healthcare), needs in the past (number of treatments already received), kinds of physiotherapy treatments received (price range of treatments), problems encountered (e.g. long waiting time, unavailability of necessary treatments), and the effects of the treatment received (any improvements in the health state after the treatment).

The innovativeness of this study lies also in the fact that it analyses the degree of satisfaction with treatments either purchased in separate treatment series (usually around ten physiotherapy sessions in a series) or through a subscription allowing the use of physiotherapy (without any additional costs) for a year. In this second option, respondents were informed that the yearly subscription would enable them to use physiotherapy treatment as needed, anywhere in Poland, in quantity and form as prescribed to them. So another objective of this work is to check whether it is the same or a different set of determinants that impacts patients' willingness to pay for a series of treatments and their willingness to pay for a subscription. Following Exworthy and Peckham (2006), I expected that respondents would be willing to pay for the treatment and there would be relatively few answers expressing unwillingness to pay, i.e. 'protest answers'.

As experience is the key area studied in this work, the research sample consisted only of those individuals who received physiotherapy treatment at least once. Due to frequent injuries in sports, sportspeople often resort to physiotherapy, so a part of the respondents were selected from among the students and employees of a Warsaw-based sports college. Other respondents were recruited from among the students and employees of one of medical colleges, as the author also needed answers from people familiar with patient care. The academic and professional knowledge of students and employees of these schools enabled the author to receive answers unbiased by media or trends.

Two methods were used to capture the relationships between the studied attributes and the willingness to pay: Welch's ANOVA test together with Games-Howell post-hoc test for the assessment of correlation, and the Bayesian network for the evaluation of causation between variables. The use of these two methods allowed a broad problem analysis and a better understanding of the influence of the respondents' characteristics and experience on their valuation. The author firstly checked whether the mean willingness to pay was different among groups divided by the attribute level, and secondly, assessed the force and direction of the influence.

2. Willingness to pay for healthcare

Most Polish citizens are entitled to the public healthcare financed by the government from taxes and contributions paid by each employed individual. However, as mentioned before, the public healthcare system has several shortcomings. One of them is long waiting time for appointments. Long queues for treatments are commonplace, which effectively forces some patients to use commercial healthcare services (especially when their health state makes it impossible for them to wait long, or when they are dissatisfied with the quality of public medical services). In Poland, the public healthcare system is coordinated by the already-mentioned National Health Fund (NFZ). Each month employed citizens pay the health insurance premium, thanks to which they can use health services in all the NFZ's medical facilities, but the waiting times, as indicated before, are long and still increasing. This applies especially to physiotherapy (please refer to the analyses of Agencja Oceny Technologii Medycznych i Taryfikacji, 2018). In 2015, the average waiting time for admission to a rehabilitation ward was 37 days for urgent cases and 347 days for not urgent ones. In 2018, the waiting periods increased to 51 and 464 days, respectively. For this reason, some people choose services offered by private medical centres, where they can receive medical help faster. Exworthy and Peckham (2006) demonstrated that patients are willing to pay more and travel further in order to reduce to the largest possible extent the waiting time for medical treatment. Needless to say, the fact that some patients are effectively forced to either pay for treatments or to wait for them for a long time negatively affects their assessment of the public healthcare system (Łosiewicz and Ryłko-Kurpiewska, 2015).

Patients' valuation of the access to the healthcare system provides information that might be useful in planning an extension of or changes in the range of medical services or prices of the services offered by a medical services provider. It also helps medical services providers meet customers' needs and expectations more accurately.

Healthcare products and services differ from other products and services analysed by economists (Arrow, 1963). The demand for healthcare is not constant and is therefore difficult to predict. What is also difficult to foresee is the quality of healthcare services, so the decision to use them requires some degree of trust between the patient and the provider of the service. Additionally, the recovery process is as unpredictable as the illness itself, and moreover one cannot test healthcare services before purchasing them. The above-mentioned aspects matter when the profitability of specialised healthcare services and the allocation of resources (often scarce) within a particular (public or private) system are considered.

One of the indicators of a patient's assessment of medical services is the willingness to pay (WTP) value. This value represents the maximum price a person is ready to pay for a good that currently is not in his or her possession (Horowitz & McConnell, 2003). The willingness to pay is related to another value called the willingness to accept (WTA), which represents a minimum price a person is ready to accept to sell or give up a good which is currently in his or her possession. The willingness to pay and the willingness to accept usually differ from each other in such a way that the WTP is often smaller than the WTA. Horowitz and McConnell (2002) and O'Brien et al. (2002) showed in their research related to health that such difference might be even sevenfold. Horowitz and McConnell (2002) moreover noticed that a bigger difference is related to non-market and public goods (e.g. healthcare) than to typical market goods or money.

The concept of the willingness to accept might be difficult to understand for patients. For example, if a person has a medical treatment scheduled that would significantly improve his or her health and quality of life, it would be strange to ask this person about a price he or she is willing to accept to give it up to somebody else. Even if the patient is ready to provide such a valuation, it might not be the value we expected. Due to such issues, the value of the willingness to pay is used more often in studies related to health.

The WTP is usually calculated by the contingent valuation method or discrete choice experiments. The first of these methods uses a set of questions about the maximum amount the respondent is willing to pay for a specific good under defined conditions – for example, to start treatment for a given illness (Bayoumi, 2004). This yields a monetary value directly and allows the generalisation regarding different health states and levels of risk. However, the method is prone to many bias--inducing effects. In an ideal world, the answers should correspond very closely to the values of the willingness to pay and the willingness to accept from a real-life situation, but usually they do not. The most effective test would be to compare values received in research with amounts paid or accepted by respondents on the market. Such tests are generally rare, but their results show differences between those values. These differences are often moderate in scale (Johannesson et al., 1999). It is also possible to compare respondents' valuation of healthcare services with their characteristics; for example, whether somebody's willingness to pay does not exceed his or her budget.

The second method, the discrete choice experiment, is based on a set of choices between defined and statistically independent pairs of scenarios (Ryan & Gerard, 2003). Each respondent's choice represents his or her utility from a given choice, based on the presented levels of the used variables. Such methodology allows the assessment of compromises between the levels of variables, but requires defining assumptions regarding the shape of the utility function. It is possible to calculate the monetary value using a cost function. The valuation of additional costs and the health insurance pricing usually uses this method (Ryan, 2004).

Many biases and effects influence consumers' willingness to pay (Brown & Gregory, 1999), e.g. the endowment effect, the income effect, the lack of substitutes and the lack of experience. We talk about the endowment effect when an individual values a good higher if he or she owns it or owned it in the past. The income effect, on the other hand, puts a limit (related to the respondent's salary) on the price one can pay for a given good. Lack of easy access or highly priced substitutes increases the valuation. Negative experiences might affect the valuation to such an extent that in some cases they even override the endowment effect.

Respondents' demographic characteristics also influence their willingness to pay for healthcare. Aizuddin et al. (2012) showed a significant relationship between the willingness to pay for healthcare services and the respondents' age, level of education, income, rural/urban place of residence, household size, and the quality of available healthcare services. Statistics from Poland (NIK, 2014) showed that the waiting time for starting physiotherapy in this country varied across regions, which indicates that respondents' place of residence influences the waiting times for this kind of services. Gonen and Bokek-Cohen (2018) demonstrated that emotions related to medical treatment influence patients' valuation of similar procedures. Also, socio-economic status and the level of satisfaction with previous physiotherapy treatments proved to be correlated with patients' willingness to pay for future treatments (Fatoye et al., 2020).

3. Methods

3.1. Design

The research was based on the author's custom survey form. Respondents were not paid for their participation in the study. The survey had a theoretical character, as the evaluated public good was not provided to participants following the end of the experiment. The responses were collected individually and were not shared.

The survey form consisted of three open-ended and 17 close-ended questions, of which four were related to the experience with physiotherapy treatment, seven to the willingness to pay for physiotherapy, and nine to patient's demographic characteristics. Open-ended questions, regarding the willingness to pay for a series of physiotherapy treatments and the willingness to pay for a yearly subscription (that would allow the use of physiotherapy when needed) were presented as short scenarios in which respondents were asked to imagine that they did not have access to physiotherapy within the public healthcare. They were asked about the amount they would pay for (a) one series of treatments, or (b) a yearly subscription enabling them to use physiotherapy treatment without limits (a and b options as separate valuation questions). They could use the evaluated good in any healthcare facility in Poland. The price for the evaluated service would have to be paid in advance, before its consumption. The 'I do not want to pay' option was also provided, in order to avoid the protest effect.

The respondents were told that their answers regarding the evaluation of healthcare services would help recognise the necessary changes in the Polish healthcare system as well as improve the quality of services. Respondents were also informed that the aim of the study was to find both the areas where the Polish healthcare system needed improvements, and those which were worth preserving (as functioning well). Figure 1 presents variables from the survey and their definitions. Most of the outcomes were nominal.

Figure 1. Three sets of variables used in the study

Name	Meaning
Variables related to experience with physiotherapy	
Treatment Type	Dominating type of the respondent's last physiotherapy (physical therapy, kinesitherapy or massage)
Improved Health	Any health improvement as a result of the last physiotherapy treatment/series
Difficulties	Any problems with receiving physiotherapy treatment (e.g. long waiting time, faraway travels)
Treatment Range	Respondent's opinion about the range of treatments available at his/her place of residence (very broad, sufficient or poor)

Figure 1. Three sets of variables used in the study (cont.)

Name	Meaning
Variables related to the willingness to pay for physiotherapy	
WTP Treatment / WTP Subscription	Respondent's WTP for a series of physiotherapy treatments / for a subscription for physiotherapy treatments
Tax Deduction	Respondent's deduction of physiotherapy costs from taxes
Max Distance	Maximum distance that the respondent is willing to travel to receive physiotherapy
Freq NFZ / Private	Number of treatments financed by the NFZ/by the respondent receivied within the last 10 years
Lst Pymt Scheme	Way of financing the respondent's last physiotherapy (by NFZ or privately)
Variables related to demographic characteristics	
Population	Population of the respondent's place of residence
Salary	Respondent's salary
Phys Family	Respondent's family member/s who also used physiotherapy
Age	Respondent's age
Sex	Respondent's sex
Residence	Respondent's place of residence (urban or rural area)
Education	Respondent's education level (secondary or higher)
Type Work	Type of work performed by the respondent (mental, physical or mental-physical)
Work Exp	Respondent's work experience

Source: author's work.

3.2. Procedure

Data received from the survey was analysed by means of two methods: Welch's ANOVA test and the Games-Howell post-hoc test for the assessment of correlation, and the Bayesian network for the evaluation of causation between variables.

Fisher's analysis of variance (ANOVA) is based on the F-test. It determines whether there are statistically significant differences between the means of the analysed groups. If the equal variance assumption within groups is violated, it is possible to use Welch's ANOVA test (Delacre et al., 2019), which is insensitive to unequal variances. Combinations of groups created on the basis of the socio-economic and demographic characteristics can be compared with each other by means of the Games-Howell post-hoc test. Although having a similar form to Tukey's test, the Games-Howell test does not assume equal variances and sample sizes (Lee & Lee, 2018; Shingala & Rajyaguru, 2015). It was designed on the basis of Welch's degrees of freedom correction using Tukey's studentized range distribution. The test employs different pooled variances for each pair instead of the same pooled variance. As variances in this study are heterogeneous, this test was used to compare the average values of the willingness to pay between groups.

The impact of the experience on the willingness to pay was assessed by means of the Bayesian network. This kind of a model consists of three main elements (Stephenson, 2000):

- V – set of variables;
- A – set of directed arcs between variables; together with V creates a graphical structure $G = (V, A)$;
- P – set of conditional probabilities of all variables given their respective parents, where $P = (v|\pi_y): v \in V$, and π_y is a set of parents of v.

Variables and arcs together form a directed acyclic graph (DAG), where all edges need to be directed, and no cycles are allowed. Directed edges indicate which variables influence the given variable. Firstly, the network structure needs to be learnt on the basis of constraint-based or score-based algorithms. After having learnt the structure of the network, the parameters of the local distribution functions are estimated. Each variable has its conditional probability table calculated on the basis of all the configurations of the values of the parents of the variable. Bayesian networks provide a powerful tool to visualise probabilities of given scenarios and review relationships between variables found in the data. For the above reasons, I used a similar model to assess the influence of experience on the willingness to pay.

3.3. Subjects

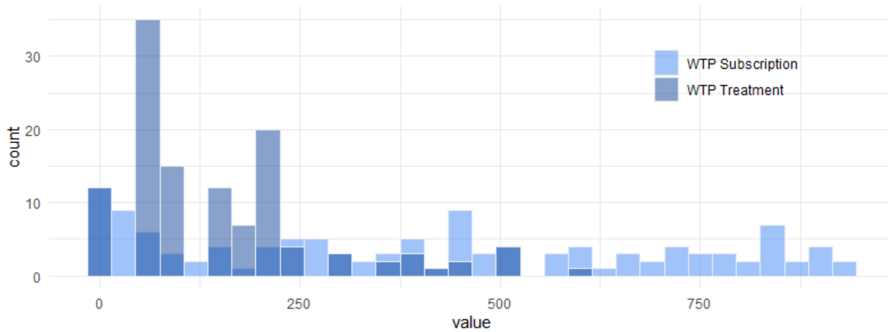
The survey was conducted on a group of 121 respondents who received physiotherapy treatment at least once. They were selected from amongst the employees and students of the Radom University and the Education in Sport University, both from their undergraduate and graduate programs and the 'Third Age University' (where academic classes are offered to older people). The research was carried out in February and March 2020, and the responses collected upon the respondents' oral consent for the participation in the study. All the information obtained was processed and stored anonymously, meeting the data confidentiality requirements as foreseen by Polish law.

The dominating demographic characteristics of the questioned individuals were:

- sex – female (64%);
- age – 18-25 year-olds (43%);
- place of residence – urban (90%);
- population of the place of residence – more than 500 thousand (33%);
- type of job – white collar (71%);
- years spent in education – 12 (68%);
- salary – over PLN 4,000 (55%);
- years of professional experience – 0–10 (62%).

There was a small percentage of 'protest answers' – only 10% for both the willingness to pay for treatment series and the willingness to pay for physiotherapy subscription. The figures provided by the respondents as the valuation of treatment series ranged between PLN 50-250 with a maximum of PLN 600 and the median at PLN 80. Valuations for a yearly subscription were spread wider through the scale, with a maximum of PLN 930 and a median at PLN 400 (Figure 2 and Table 1).

Figure 2. Respondents' WTP for treatment series and their WTP for subscription



Source: author's calculations based on data collected in Feb-March 2020.

Table 1. Minimum, median, mean and maximum values of the WTP for a treatment series and the WTP for subscription

	Min	Median	Mean	Max
WTP Treatment	0	80	149.8	600
WTP Subscription	0	400	397.2	930

Source: author's calculations based on data collected in Feb-March 2020.

4. Results

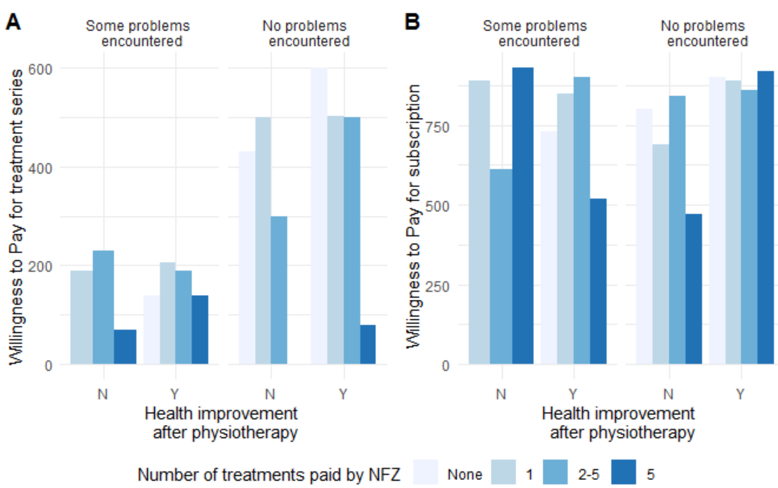
Experiences related to physiotherapy were proven impactful on the willingness to pay for a treatment series or subscription. More specifically, such effect was observed for variables representing difficulties encountered in arranging treatment and the frequency of the past use of either public or private healthcare services. Opinions about the range of treatments offered in local facilities also impacted the maximum distance an individual was ready to travel to receive treatment. The income effect was observed as well.

4.1. Correlations

As expected, experiences such as having already paid for treatments or having encountered problems with receiving physiotherapy were correlated with the treatment valuation. What came as a surprise, though, was that the salary factor was correlated only with the valuation of the treatment series, not the valuation of the physiotherapy subscription. Significant correlation of the dominating type of the last physiotherapy treatment with the WTP for physiotherapy subscription was observed, but the same did not hold for the WTP for treatment series. However, the above was expected, and could be explained by the fact that different kinds of physiotherapy are differently priced. Besides the respondent's salary, the WTP for treatment series was correlated with his or her age and place of residence. The WTP for a yearly subscription, on the other hand, was correlated only with having a family member who used physiotherapy, and with the population of the respondent's place of residence.

Figure 3(A) shows that respondents valued physiotherapy more when they did not have much experience with treatments financed by the NFZ, they encountered some problems while trying to receive it through the public healthcare system, or their health improved after the last physiotherapy treatment. As regards the WTP for the subscription for physiotherapy, the correlation is not visible at first sight (Figure 3B). Slightly higher valuations were received from respondents who both experienced problems while arranging/using physiotherapy within the public healthcare system, and whose health improved after the last treatment. The number of treatments financed by the NFZ turned out to be of no significance.

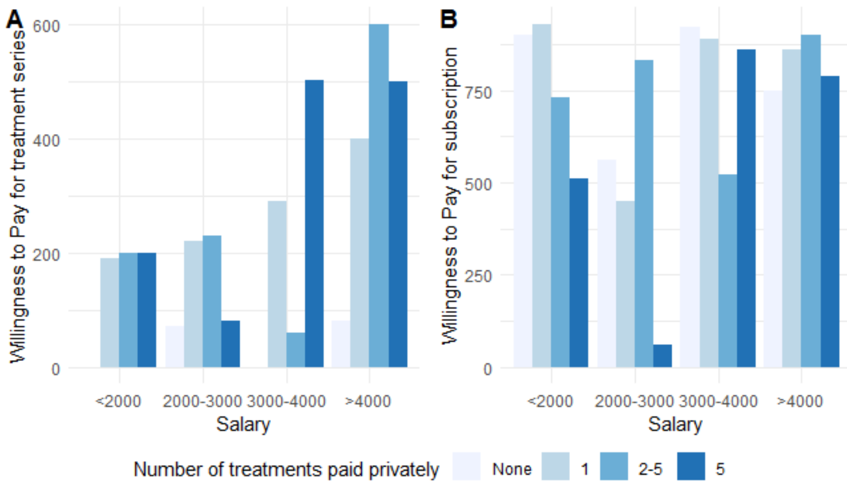
Figure 3. WTP for treatment series (A) and WTP for subscription (B) with and without problems encountered in the past



Source: author's calculations based on data collected in Feb-March 2020.

What could be observed and was expected was the fact that the WTP increased along with growing salaries. Figure 4(A) moreover demonstrates that another treatment series was valued higher by respondents who already paid for physiotherapy in the past. However, when it comes to the WTP for physiotherapy subscription, salary level and previous experiences with private physiotherapy made only a slight difference, as demonstrated in Figure 4(B).

Figure 4. WTP for treatment series (A) and WTP for subscription (B) as dependent on salary and experience in paying for physiotherapy



Source: author's calculations based on data collected in Feb-March 2020.

Table 2 presents the results of Welch's one-way ANOVA tests. Due to its large size, Table 3 with results of the Games-Howell test was placed in the Appendix. Table 3 features statistically significant results only. The average WTP for treatment series was significantly higher in the case of respondents who had never used physiotherapy within the public healthcare system than in the case of both the regular and occasional users of the NFZ-provided physiotherapy. The average valuation was also higher when the number of physiotherapy treatments paid by the patient was larger than 0. This suggests that broader experience with commercial physiotherapy makes patients more prepared to pay for such services. By the same token, a lack of or relatively modest experience with the NFZ-financed physiotherapy proved related to a higher mean valuation. Average valuations were also higher when individuals encountered problems with receiving physiotherapy through the public system, which probably results from the belief that it is easier to receive physiotherapy treatment within the private healthcare system, and that medical personnel are generally more patient-friendly there.

Only a few age and salary groups differed significantly from each other as regards the WTP for treatment series. This is in line with the common expectation that people earning more will be prepared to pay more for medical treatment. Also patients in older age groups, close to the pension age, are believed more likely to spend larger amounts on their health than individuals from younger age groups.

When it comes to the WTP for a yearly subscription, significantly different averages were observed only between a few groups. Patients using mainly massage during their most recent physiotherapy treatment granted lower valuation to the subscription than those using kinesitherapy. As each physiotherapy type is priced differently and is prescribed for specific needs, it is expected that people's valuation will depend on the type of treatment. What came as a surprise, though, was that the average valuation of a subscription was lower in the group of respondents who used commercially-provided physiotherapy relatively often, i.e. more than five times in the last 10 years, than in the group of people who used it moderately often (two to five times in the last 10 years).

Respondents whose parents or spouses used physiotherapy services granted higher average valuations to a yearly subscription than those whose other family members used it. This difference suggests that people were ready to pay more when someone relatively close to them was using physiotherapy, so they were taking into account a close persons' experience. There was also a significant difference between the averages of groups of respondents coming from areas with the relatively smallest and the relatively largest populations, which indicates that residents of large urban areas are accustomed to higher prices and greater spending.

Table 2. Welch's ANOVA test results (only variables with significant differences in groups)

		F	*p*-value
WTP Treatment	Freq NFZ	25.145	5.496e-10***
	Freq Private	27.366	1.006e-10***
	Difficult Receive	43.037	2.24e-09***
	Salary	9.583	9.01e-05***
	Age	5.119	0.003**
	Residence	4.999	0.028*
WTP Subscription	Treatment Type	5.030	0.009**
	Freq Private	3.573	0.022*
	Difficult Receive	4.616	0.034*
	Phys Family	11.362	5.064e-06**
	Population	8.689	7.575e-05**

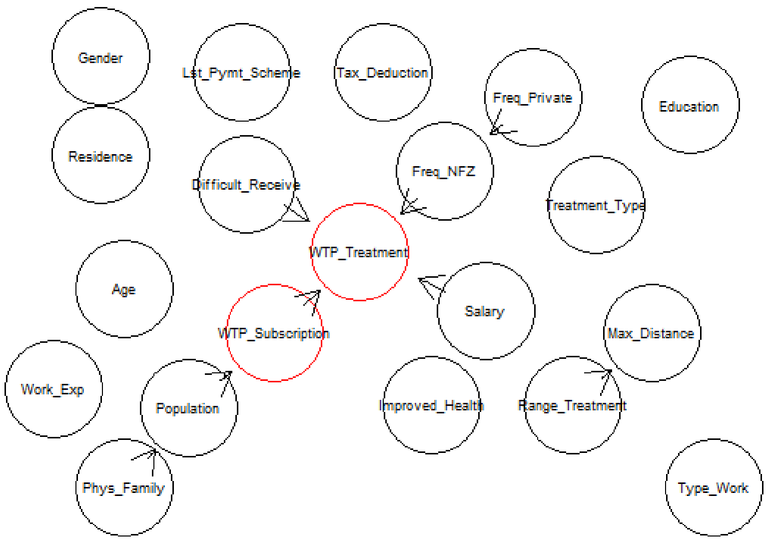
Note. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Source: author's calculations based on data collected in Feb-March 2020.

4.2. Causation

The Bayesian network was used to assess the influence of respondents' demographic characteristics and past experiences with physiotherapy on the WTP for a treatment series. In the diagram, the nodes representing the two WTP variables were marked in red (Figure 5).

Figure 5. The Bayesian network and relationships between variables used



Source: author's work based on data collected in Feb-March 2020.

As expected, the salary level, previous experience of difficulties in arranging or using physiotherapy, and experiences with physiotherapy within the public healthcare system influenced the WTP for a treatment series. Experience related to paying for treatment also influenced this indicator, but indirectly, i.e. it impacted the number of the NFZ-financed treatments that respondents used, which, in turn, affected the WTP. An interesting relationship was also observed between the two analysed variables (the WTP for a yearly subscription and the WTP for treatment series), where the former affected the latter. The WTP for a yearly subscription was directly affected by the size of the population of the respondent's place of residence, and indirectly by the prior experience with physiotherapy treatment of the respondent's close family member. Another relationship was found between the respondent's opinion about the range of treatments offered in local medical facilities and the maximum distance he or she was willing to travel to receive treatment.

Tables 4 and 5 present conditional probabilities for nodes representing the WTP for a physiotherapy subscription and the WTP for treatment series, respectively. Due to the tables' large sizes, the author placed both of them in the Appendix.

Conditional probabilities suggest that people living in highly-populated areas (with more than 500,000 inhabitants) are likely to value the yearly access (subscription) to physiotherapy treatment higher (are willing to spend PLN 660–930 on such a subscription) than people from places with smaller populations. The latter, i.e. people coming from places with populations up to 10,000 inhabitants and between 10,000 and 50,000 inhabitants, are willing to pay considerably smaller amounts, i.e. the maximum of PLN 120 and PLN 120–400, respectively.

In the case of some respondents, the quoted value of the WTP for physiotherapy subscription increased the probability of indicating concrete values for the WTP for physiotherapy series. For example, the WTP for a physiotherapy subscription in the range of (660,930] was related to a high probability (~0.42) of the WTP for treatment series valuation between PLN 60–80. Low values of the WTP for a physiotherapy subscription [0,120], on the other hand, were associated with a high probability (~0.32) of a relatively high valuation of treatment series (80,200].

The income effect was visible in the valuation of the physiotherapy treatment series. There was a noticeable probability (~0.34) that people with high earnings (between PLN 3,000 and PLN 4,000 per month) will quote the highest prices (between PLN 200 and PLN 600). Patients with lower salaries (less than PLN 2,000) were more likely (~0.51) to value the a physiotherapy treatment series at the maximum of PLN 60.

The effect of previous experience was also noticeable. Patients who encountered problems with getting physiotherapy treatment in the past (as mentioned before, e.g. certain types of treatment unavailable, long waiting times, long journeys involved) were more likely to value the access to treatment series higher. Approximately 33% of them were prepared for a price within the range of PLN 80–200, and about 32% for a price within the range of PLN 200–600.

Patients who often used physiotherapy financed by the NFZ were on the other hand more likely (~0.3) to declare lower values of the WTP for a treatment series (a maximum of PLN 60). Those who never used physiotherapy provided within the public healthcare system would probably (~0.31) declare higher values (PLN 80–200). Similarly, there was a high probability (~0.29) that patients who often paid for physiotherapy in the past would be willing to pay between PLN 80 and PLN 200 for the next treatment series. Those who did not have any experience with paid treatment would probably (~0.3%) state a minimal price of PLN 60.

Those respondents who have never had any problems with receiving physiotherapy, who often used treatment paid by the NFZ, or whose salaries are relatively low, would probably declare the WTP for treatment series between 0 and PLN 60. Those who had difficulties in getting access to physiotherapy in the past, who have never used treatment financed by NFZ, or whose salary is in the highest income range, would probably value physiotherapy treatment within the price range of PLN 200–600.

A few variables correlated with the WTP were not affecting the valuation. The WTP for treatment series was correlated with the respondents' age and place of residence, but those variables were not found impactful. Also, the dominant treatment type of the most recent physiotherapy, the frequency of receiving private treatments, and difficulties with arranging or using physiotherapy were correlated with the WTP for subscription, but did not impact its valuation.

5. Conclusions

While planning changes in the prices of treatments, it would be useful for owners and managers of medical facilities to accurately predict patients' behaviour and decisions. Due to the fact that healthcare services are not typical services, it is not easy to understand how clients value particular medical treatments. A 'willingness to pay' indicator proves helpful here, as it allows a better understanding of customers' needs and expectations. Depending on the chosen method, the values of this indicator might be calculated directly (by asking respondents about the maximum price they would be prepared to pay for a good they do not have) or indirectly (by inferring these values from respondents' decisions between a set of scenarios). In this research, only two determinants were chosen for the analysis of the willingness to pay, namely the income effect, related to respondents' salaries, and the effect of experience, shaped by respondents' memories of past treatments and the emotions connected with them which might affect the willingness to pay for similar goods in the future.

This research has shown that sociodemographic characteristics of patients and their previous experiences impact the willingness to pay for physiotherapy. As expected, the willingness to pay for a treatment series depended on the salary level, the experiences with using physiotherapy (especially within the public healthcare system, but not only), and the potential problems encountered while receiving physiotherapy. What turned out against expectations, though, was the dependence of the willingness to pay on the size of the population of respondents' place of residence and, indirectly, on whether respondents' close family members also used physiotherapy. It was expected that the salary level and treatment type would impact the value a person was prepared to pay for the subscription to physiotherapy, often more expensive than a single treatment series, but overall likely to prove cheaper (when costly treatments are necessary). On the other hand, the lack of dependence between the type of treatment received in the most recent physiotherapy series and the willingness to pay for another treatment series might result from patients' uncertainty about the potential future health issues and treatments needed. Another unexpected outcome of this study was the lack of correlation between the observed improvement in the patient's health state after the most recent physiotherapy treatment and his or her willingness to pay. However, this inconsistency could be explained by patients' uncertainty regarding the effects and the quality of potential future treatments (the already-mentioned unpredicatability related to health problems and recovery).

The study also brought to light an interesting relationship between the valuations of the two analysed goods. The willingness to pay for a treatment series turned out to be dependent on the willingness to pay for a yearly subscription. Although this relationship was not completely linear at some levels of the rest of the parent nodes, it nevertheless strengthened the effect.

The healthcare system in Poland and the patients' attitudes towards paying for medical services are specific, as demonstrated by other researchers as well. Pajewska-Kwaśny (2016) showed that even though the public healthcare system is not working optimally, and the range of services offered is relatively narrow and not of the best quality, patients in Poland are reluctant to buy additional healthcare insurance. Aspects influencing the exact valuation of healthcare services were studied in a similar way as in this work.

Bielawska and Lyskawa (2021) demonstrated that age, years in education and the size of the population of the place of residence all influence the willingness to pay for medical services (in my research, only the size of the population of the place of residence proved significant). Dror et al. (2007) showed the correlation between the willingness to pay for health insurance and the respondents' salaries and the education level. Salary was also found the key determinant of the willingness to pay for internal preferences concerning multiple health statuses presented (Javan-Noughabi et al., 2017). Nielsen et al. (2003) demonstrated that socio-demographic characteristics like respondents' sex, education, place of residence and age significantly influence their willingness to pay for health services, so in other words, their willingness to pay for the reduction of the future health risk. Research conducted on the basis of data from the Social Diagnosis' databases moreover showed that income and previous medical expenses might influence patients' willingness to pay for healthcare services (Jewczak, 2014). Żółtaszek (2012) presented similar observations to the results of my study, namely that the willingness to pay grows along with increasing income and experience with private medical services or insurance. Similarly Dudziński (2019) – proved that the effectiveness of the public healthcare system might influence the willingness to pay for healthcare services.

Like my research, other studies as well show that emotions – either by themselves or in interaction with other variables – cause changes in the willingness to pay, not only for health services but also for other goods (Silva et al., 2019). Bigné et al. (2008) showed that satisfaction both impacts customers' loyalty and increases their willingness to pay for a service. Most of my respondents were willing to pay at least a small amount for an improved access to physiotherapy, even though the valuation questions were open-ended. Some other analyses showed that patients in Poland declare unwillingness to pay for healthcare services if the question regards solely their general inclination (Magda & Szczygielski, 2012). The willingness to pay grows significantly when a specific price per visit is mentioned. Unlike this study, Markiewicz (2021) showed that sociodemographic characteristics are not always influencing patients' willingness to pay for healthcare-related treatments. It might turn out in the course of further research that only health improvements plausible 'here and now' influence the valuation.

The aim of this research has been fulfilled. However, some limitations occurred that might be addressed in the future. What is worth consideration in this context is the use of a more precise tool to define respondent's previous, current, and possible future health states, for example the EuroQol-5D (EQ-5D) scale (Brooks et al., 2003) or the Quality-Adjusted Life Years (QALY) (Haninger & Hammitt, 2006). The emotion-related aspects of patients' choices might also be studied in more detail by enabling respondents to comprehensively describe their experiences and feelings related to the analysed good. Another possible development of this study could involve the comparison of two groups of respondents: those who used physiotherapy treatment in the past and those who did not.

References

Agencja Oceny Technologii Medycznych i Taryfikacji. (2018). *Koncepcja zmian organizacji i funkcjonowania rehabilitacji leczniczej w systemie ochrony zdrowia w Polsce*.

Aizuddin, A. N., Sulong, S., & Aljunid, S. M. (2012). Factors influencing willingness to pay for healthcare. *BMC Public Health, 12*(Suppl 2), 1. https://doi.org/10.1186/1471-2458-12-S2-A37.

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review*, *53*(5), 941–973. https://assets.aeaweb.org/asset-server/files/9442.pdf.

Bayoumi, A. M. (2004). The measurement of contingent valuation for health economics. *PharmacoEconomics, 22*(11), 691–700. https://doi.org/10.2165/00019053-200422110-00001.

Bielawska, K., & Lyskawa, K. (2021). Retirees' Healthcare Needs and Satisfaction with their Coverage. *European Research Studies Journal*, *24*(2), 1007–1018. https://doi.org/10.35808/ersj/2169.

Bigné, J. E., Mattila, A. S., & Andreu, L. (2008). The impact of experiential consumption cognitions and emotions on behavioral intentions. *Journal of Services Marketing*, *22*(4), 303–315. https://doi.org/10.1108/08876040810881704.

Brooks, R., Rabin, R., & de Charro, F. (Eds.). (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. Springer Science & Business Media. https://doi.org/10.1007/978-94-017-0233-1.

Brown, T. C., & Gregory, R. (1999). Why the WTA–WTP disparity matters. *Ecological Economics, 28*(3), 323–335. https://doi.org/10.1016/S0921-8009(98)00050-0.

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking Parametric Assumptions Seriously: Arguments for the Use of Welch's *F*-test instead of the Classical *F*-test in One-Way ANOVA. *International Review of Social Psychology*, *32*(1), 1–12. http://doi.org/10.5334/irsp.198.

Dror, D. M., Radermacher, R., & Koren, R. (2007). Willingness to pay for health insurance among rural and poor persons: Field evidence from seven micro health insurance units in India. *Health Policy*, *82*(1), 12–27. https://doi.org/10.1016/j.healthpol.2006.07.011.

Dudziński, P. (2019). The effect of financial risk on the demand for supplementary health insurance. *Współczesna Gospodarka, 10*(2), 1–8. https://doi.org/10.26881/wg.2019.2.01.

Exworthy, M., & Peckham, S. (2006). Access, Choice and Travel: Implications for Health Policy. *Social Policy & Administration*, *40*(3), 267–287. https://doi.org/10.1111/j.1467-9515.2006.00489.x.

Fatoye, F., Mbada, C., Oluwatobi, S., Odole, A., Oyewole, O., Ogundele, A., & Ibiyemi, O. (2020). Pattern and determinants of willingness-to-pay for physiotherapy services. *European Journal of Physiotherapy*, *22*(4), 221–227. https://doi.org/10.1080/21679169.2019.1591502.

Gonen, L. D., & Bokek-Cohen, Y. (2018). Valuing the invaluable: Do emotional experiences during fertility treatments affects the willingness to pay for them?. *European Review of Applied Psychology*, *68*(2), 45–60. https://doi.org/10.1016/j.erap.2018.01.002.

Haninger, K., & Hammitt, J. K. (2006, February). *Willingness to Pay for Quality-Adjusted Life Years: Empirical Inconsistency Between Cost-Effectiveness Analysis and Economic Welfare Theory*. http://idei.fr/doc/conf/fpi/papers_2006/hammitt.pdf.

Horowitz, J. K., & McConnell, K. E. (2002). A review of WTA/WTP studies. *Journal of Environmental Economics and Management, 44*(3), 426–447. https://doi.org/10.1006/jeem.2001.1215.

Horowitz, J. K., & McConnell, K. E. (2003). Willingness to accept, willingness to pay and the income effect. *Journal of Economic Behavior & Organization*, *51*(4), 537–545. https://doi.org/10.1016/S0167-2681(02)00216-0.

Javan-Noughabi, J., Kavosi, Z., Faramarzi, A., & Khammarnia, M. (2017). Identification determinant factors on willingness to pay for health services in Iran. *Health Economics Review*, *7*, 1–6. https://doi.org/10.1186/s13561-017-0179-x.

Jewczak, M. (2014). Spatial distribution of Willingness to Pay for health services of households in Poland. *Acta Universitatis Lodziensis. Folia Oeconomica, 3*(302), 241–251. https://czasopisma.uni.lodz.pl/foe/article/view/57.

Johannesson, M., Blomquist, G. C., Blumenschein, K., Johansson, P. O., Liljas, B., & O'Conor, R. M. (1999). Calibrating hypothetical willingness to pay responses. *Journal of Risk and Uncertainty*, *18*(1), 21–32. https://doi.org/10.1023/A:1007708326467.

Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test?. *Korean Journal of Anesthesiology*, *71*(5), 353–360. https://doi.org/10.4097/kja.d.18.00242.

Łosiewicz, M., & Ryłko-Kurpiewska, A. (2015). Public perception of healthcare personnel in Poland and some other European countries in view of selected studies. *International Business and Global Economy*, *34*, 198–209. https://doi.org/10.4467/23539496IB.13.016.3989.

Magda, I., & Szczygielski, K. (2012). Współpłacenie za usługi medyczne w Polsce: możliwości wprowadzenia i potencjalne skutki. In E. Nojszewska (Ed.), *Racjonalizacja kosztów w ochronie zdrowia* (pp. 109–119). Wolters Kluwer Polska.

Markiewicz, O. (2021). Value of Life Year and Cost-Effectiveness Thresholds: The Case of Poland. *Central European Economic Journal*, *8*(55), 256–268. https://doi.org/10.2478/ceej-2021-0019.

Najwyższa Izba Kontroli. (2014). *Dostępność i finansowanie rehabilitacji leczniczej. Informacja o wynikach kontroli.*

Nielsen, J. B., Gyrd-Hanses, D., Kristiansen, I. S., & Nexøe, J. (2003). Impact of Socio-demografic Factors on Willingness to Pay for the Reduction of a Future Health Risk. *Journal of Environmental Planning and Management*, *46*(1), 39–47. https://doi.org/10.1080/713676699.

O'Brien, B. J., Gertsen, K., Willan, A. R., & Faulkner, L. A. (2002). Is there a kink in consumers' threshold value for cost-effectiveness in health care?. *Health Economics*, *11*(2), 175–180. https://doi.org/10.1002/hec.655.

Pajewska-Kwaśny, R. (2016). Możliwości implementacji w Polsce wybranych aspektów amerykańskiego systemu ubezpieczeń zdrowotnych. *Ekonomia i Prawo w Ochronie Zdrowia*, (1), 95–111.

Ryan, M. (2004). Discrete choice experiments in health care. *The BMJ*, *328*(7436), 360–361. https://doi.org/10.1136/bmj.328.7436.360.

Ryan, M., & Gerard, K. (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied Health Economics and Health Policy*, *2*(1), 55–64.

Shingala, M. C., & Rajyaguru, A. (2015). Comparison of Post Hoc Tests for Unequal Variance. *International Journal of New Technologies in Science and Engineering*, *2*(5), 22–33. https://www.ijntse.com/upload/1447070311130.pdf.

Silva, A., Von Hausen, N., & Higgins, L. M. (2019). The interaction effect of socioeconomic status and emotions on willingness to pay. *The International Review of Retail, Distribution and Consumer Research*, *29*(1), 118–124. https://doi.org/10.1080/09593969.2018.1556180.

Stephenson, T. A. (2000). *An Introduction to Bayesian Network Theory and Usage*. Dalle Molle Institute for Perceptual Artificial Intelligence.

Viscusi, W. K. (1985). A Bayesian perspective on biases in risk perception. *Economics Letters*, *17*(1–2), 59–62. https://doi.org/10.1016/0165-1765(85)90127-2.

Żółtaszek, A. (2012). Private financing of health care in Poland – microsimulation model. *Przegląd Statystyczny*, *59*(3), 290–301. https://ps.stat.gov.pl/PS/2012/3/2012_59_3_290-301.pdf.

Appendix

Table 3. Games-Howell post hoc test results

Variable	Group 1	Group 2	Estimate	Conf.low	Conf.high	p. adj.
			WTP Treatment			
Freq NFZ	1	>5	−144.	−213.0	−74.3	1.2e-05***
	2–5	>5	−98.7	−144.0	−53.1	2.29e-06***
	2–5	None	110.0	1.73	218.0	0.046*
	>5	None	209.0	104.0	313.0	11.3e-05***
Freq Private	1	>5	102	6.78	197.0	0.032*
	1	None	−111.0	−155.0	−66.5	1.95e-07***
	2–5	None	−147.0	−217.0	−76.9	8.07e-06***
	>5	None	−213.0	−306.0	−119.0	7.27e-06***
Diff. Receive	N	Y	127.0	88.6	165.	2.24e-9***
Salary	<2000	>4000	117.0	54.4	180.0	3.73e-5***
	<2000	3000–4000	160.0	13.4	306.0	0.031*
	>4000	2000–3000	−73.0	−133.0	−12.7	0.011*
Age	26–35	36–45	61.9	−40.8	165.0	0.415*
Residence	City	Village	−49.7	−94.0	−5.44	0.028*
			WTP Subscription			
Treatment Type	Kinesitherapy	Massage	−237.0	−418.0	−57.0	0.007**
Freq Private	2–5	>5	−215.0	−403.0	−27.6	0.019*
Diff. Receive	N	Y	122.	9.28	234.	0.034*
Phys Family	None	Other	−321.0	−551.0	−91.0	0.003**
	Other	Parent	428.0	206.0	650.0	7.81e-5***
	Other	Spouse	344.0	157.0	530.0	2.05e-4***
Population	10	10–50	74.0	−123.0	271.0	0.820**
	10	101–500	262.0	50.2	474.0	0.009***
	10–50	>500	283.0	82.5	484.0	0.002**

Note. ∗ $p<0.05$; ∗∗ $p<0.01$; ∗∗∗ $p<0.001$.

Source: author's work based on data collected in Feb-March 2020.

Table 4. Conditional probability table for the willingness to pay for a yearly subscription node

WTP Subscription	Population (in thousands)				
	10	10–50	51–100	101–500	500
[0,120]	0.52	0.35	0.33	0.13	0.10
(120,400]	0.27	0.38	0.11	0.35	0.20
(400,660]	0.22	0.15	0.33	0.26	0.25
(660,930]	0.00	0.12	0.22	0.26	0.45

Source: author's work based on data collected in Feb-March 2020.

Table 5. Conditional probability table for the willingness to pay for a treatment series node

	Salary (in PLN)	WTP Treatment	WTP Subscription			
			[0,120]	(120,400]	(400,660]	(660,930]
Difficult Receive = N Freq NFZ = <=1	<2000	[0,60]	1.00		1.00	
		(60,80]	0.00		0.00	
		(80,200]	0.00		0.00	
		(200,600]	0.00		0.00	
	2000–3000	[0,60]	0.00		0.00	
		(60,80]	0.50		1.00	
		(80,200]	0.50		0.00	
		(200,600]	0.00		0.00	
	>4000	[0,60]	0.00	1.00	0.00	0.00
		(60,80]	0.00	0.00	0.00	0.00
		(80,200]	0.50	0.00	1.00	1.00
		(200,600]	0.50	0.00	0.00	0.00
Difficult Receive = N Freq NFZ = 2-5	<2000	[0,60]	1.00	1.00		1.00
		(60,80]	0.00	0.00		0.00
		(80,200]	0.00	0.00		0.00
		(200,600]	0.00	0.00		0.00
	2000–3000	[0,60]		1.00	1.00	
		(60,80]		0.00	0.00	
		(80,200]		0.00	0.00	
		(200,600]		0.00	0.00	
	>4000	[0,60]	0.00	0.80	0.00	
		(60,80]	0.50	0.00	0.50	
		(80,200]	0.50	0.20	0.00	
		(200,600]	0.00	0.00	0.50	
Difficult Receive = N Freq NFZ = >=5	<2000	[0,60]	1.00		1.00	1.00
		(60,80]	0.00		0.00	0.00
		(80,200]	0.00		0.00	0.00
		(200,600]	0.00		0.00	0.00
	2000–3000	[0,60]	1.00	1.00		
		(60,80]	0.00	0.00		
		(80,200]	0.00	0.00		
		(200,600]	0.00	0.00		
	3000–4000	[0,60]			1.00	
		(60,80]			0.00	
		(80,200]			0.00	
		(200,600]			0.00	
	>4000	[0,60]	1.00	0.50	0.00	
		(60,80]	0.00	0.00	1.00	
		(80,200]	0.00	0.50	0.00	
		(200,600]	0.00	0.00	0.00	

Table 5. Conditional probability table for the willingness to pay for a treatment series node (cont.)

	Salary (in PLN)	WTP Treatment	WTP Subscription			
			[0,120]	(120,400]	(400,660]	(660,930]
Difficult Receive = Y Freq NFZ = <=1	<2000	[0,60]			0.00	
		(60,80]			0.00	
		(80,200]			1.00	
		(200,600]			0.00	
	2000–3000	[0,60]			0.00	
		(60,80]			0.00	
		(80,200]			1.00	
		(200,600]			0.00	
	3000–4000	[0,60]	0.00	0.25	0.00	
		(60,80]	0.00	0.00	0.00	
		(80,200]	1.00	0.00	1.00	
		(200,600]	0.00	0.75	0.00	
	>4000	[0,60]	0.00	0.00	0.00	0.00
		(60,80]	0.00	0.00	0.50	0.43
		(80,200]	0.00	1.00	0.00	0.14
		(200,600]	1.00	0.00	0.50	0.43
Difficult Receive = Y Freq NFZ = 2–5	<2000	[0,60]	0.00	0.50	0.00	0.00
		(60,80]	0.00	0.00	0.00	1.00
		(80,200]	1.00	0.50	1.00	0.00
		(200,600]	0.00	0.00	0.00	0.00
	2000–3000	[0,60]	0.00	0.67	0.00	0.00
		(60,80]	0.00	0.00	0.00	0.00
		(80,200]	1.00	0.33	0.00	0.00
		(200,600]	0.00	0.00	1.00	1.00
	3000–4000	[0,60]	0.00			0.00
		(60,80]	0.00			0.00
		(80,200]	1.00			0.00
		(200,600]	0.00			1.00
	>4000	[0,60]	0.00	0.33	0.00	0.00
		(60,80]	0.17	0.00	0.00	0.80
		(80,200]	0.33	0.33	0.00	0.20
		(200,600]	0.50	0.33	1.00	0.00
Difficult Receive = Y Freq NFZ = >=5	2000–3000	[0,60]	0.00		1.00	
		(60,80]	1.00		0.00	
		(80,200]	0.00		0.00	
		(200,600]	0.00		0.00	
	3000–4000	[0,60]				1.00
		(60,80]				0.00
		(80,200]				0.00
		(200,600]				0.00
	>4000	[0,60]		1.00	0.00	
		(60,80]		0.00	1.00	
		(80,200]		0.00	0.00	
		(200,600]		0.00	0.00	

Table 5. Conditional probability table for the willingness to pay for a treatment series node (cont.)

	Salary (in PLN)	WTP Treatment	WTP Subscription			
			[0,120]	(120,400]	(400,660]	(660,930]
Difficult Receive = Y Freq NFZ = None	2000–3000	[0,60]		0.00		
		(60,80]		0.00		
		(80,200]		1.00		
		(200,600]		0.00		
	3000–4000	[0,60]	0.00			0.00
		(60,80]	0.00			0.00
		(80,200]	0.00			0.00
		(200,600]	1.00			1.00
	>4000	[0,60]	0.00	0.00	0.00	0.00
		(60,80]	0.50	0.00	0.33	0.00
		(80,200]	0.00	0.00	0.67	0.50
		(200,600]	0.50	1.00	0.00	0.50

Source: author's work based on data collected in Feb-March 2020.

Przegląd Statystyczny. Statistical Review, 2021, vol. 68, 4, 45–58 DOI: 10.5604/01.3001.0015.7799
Received: 10.12.2021, revised: 14.02.2022, accepted: 16.02.2022

The minimal-time growth problem and 'very strong' turnpike theorem

Emil Panek[a]

Abstract. This paper refers to the author's previous work, in which the 'weak' turnpike theorem in the stationary Gale economy was proved. This theorem states that each optimal growth process $\{y^*(t)\}_{t=0}^{t_1^*}$ that leads the economy in the shortest possible time t_1^* from the (initial) state of y^0 to the set of target/postulated states Y^1 almost always runs in the neighbourhood of the production turnpike, where the economy remains in a specific dynamic equilibrium (peak growth equilibrium). This paper presents a proof of the 'very strong' turnpike theorem in the stationary Gale economy, which states that if the optimal process (the solution to the minimal-time growth problem) reaches a turnpike in a certain period of time $\check{t} < t_1^* - 1$, then it stays on it everywhere else, except for, at most, final period t_1^*. The obtained result confirms the well-known Samuelson hypothesis about the specific turnpike stability of optimal growth paths in multiproduct/multisectoral von Neumann-Leontief-Gale-type models, also in the case where the growth criterion is not the (normally assumed) utility of production but the time needed by the economy to achieve the postulated target level or volume of production.

Keywords: stationary Gale economy, von Neumann equilibrium, minimum-time growth problem, turnpike effect
JEL: C62, C67, O41, O49

1. Introduction

There are several turnpike theorems (production, capital, consumption turnpikes, etc.) in the literature proved in various multi-product/multi-sector input-output models of economic dynamics – see e.g. Babaei (2019), Babaei et al. (2020), Giorgi and Zuccotti (2016), Jensen (2012), Khan and Piazza (2011), Majumdar (2009), Makarov and Rubinov (1977), Nikaido (1968, chap. 4), Panek (2003, part 2, chap. 5–6; 2014, 2015, 2019), Takayama (1985, chap. 6–7). An extensive bibliography on the turnpike theory is presented in McKenzie (2005), Mitra and Nishimura (2009), Panek (2011), and others. The role of the growth criterion is most frequently embraced by the production utility generated in the economy either in the last period of defined horizon $T = \{0, 1, \ldots, t-1\}$ or in all periods of the horizon. A paper by Panek (2021), however, presents a different approach – it proves a 'weak' turnpike theorem in the stationary Gale economy with n products and with a single production turnpike. In that paper, the time needed for an economy starting from a fixed initial state (production vector) of $y^0 = (y_1^0, \ldots, y_n^0) > 0$ to reach the

[a] University of Zielona Góra, Faculty of Economics and Management, Institute of Economics and Finance, ul. Podgórna 50, 65-246 Zielona Góra, e-mail: e.panek@wez.uz.zgora.pl,
ORCID: https://orcid.org/0000-0002-7950-1689.

desirable target set of states (production vectors)[1] $Y^1 = \{y \in R^n_+ | y \geqq y^1\}, y^1 > y^0$ assumes the role of the growth criterion. According to this theorem, almost all optimal growth processes[2] – regardless of the distance of the target set of states Y^1 from initial vector y^0 – take place in an arbitrarily close (in the angular measure sense) neighbourhood of the production turnpike, where the economy develops at the maximum rate and achieves the highest technological and economic efficiency. It is a state of a specific dynamic equilibrium (peak equilibrium of growth) in the Gale economy.

This paper refers to the aforementioned article and contains a proof of a 'very strong' turnpike theorem. It states that if optimal process $\{y^*(t)\}_{t=0}^{t_1^*}$ (the solution to the minimal-time growth problem) reaches the turnpike in a certain period of $\check{t} < t_1^* - 1$, it remains on it everywhere else, except for, at most, the last period of horizon $\{0, 1, \dots, t_1^*\}$. The potential precipitation of the economy from the turnpike in period t_1^* results from the necessity to reach the target set of states Y^1.

The paper further consists of Section 2, where a model of the stationary Gale economy is presented and selected properties of the production turnpike and the von Neumann equilibrium state are defined and discussed, Section 3, which presents the minimal-time growth issue and the conditions for the existence of a feasible stationary and optimal growth process, Section 4, which provides the formulation and proof of a 'very strong' turnpike theorem (Theorem 3), and Section 5, which features a certain specific version of the 'very strong' turnpike theorem in the stationary Gale economy with a single turnpike and a minimum-time growth criterion (Theorem 3'). The paper concludes with the author's indication of the possible directions for further development of the current research.

2. Technological and economic production efficiency. Von Neumann equilibrium[3]

In the economy we have $n < +\infty$ consumed and/or produced commodities. We consider a model with discrete time $t = 0, 1, \dots$. By $x = (x_1, \dots, x_n)$ we denote the input vector that is used in the economy in a specific unit of time, e.g. for a year (we also call it a production factors vector), and by $y = (y_1, \dots, y_n)$ the output vector that is produced in a unit of time (also called a production vector). If the technology at the disposal of the economy allows the achievement of production y from inputs x, then the pair (x, y) is said to create (describe) a technologically feasible

[1] If $a, b \in R^n$, then $a \geqq b$ means that $\forall i (a_i \geq b_i)$. The notation $a \geq b$ means that $a \geqq b$ and $a \neq b$. We define notation $a \leqq b$ similarly.

[2] Leading in the shortest time from initial state y^0 to set Y^1.

[3] The notation used further in this paper refers to Panek (2021).

production process.[4] Non-empty set $Z \subset R_{+}^{2n}$ of all technologically feasible production processes is called the Gale production space (or the technological set) if the following conditions are met:

(G1) $\forall (x^1, y^1) \in Z \ \forall (x^2, y^2) \in Z \ \forall \lambda_1, \lambda_2 \geq 0 \ (\lambda_1 (x^1, y^1) + \lambda_2 (x^2, y^2) \in Z)$
(inputs/outputs proportionality condition and the additivity of production processes),

(G2) $\forall (x, y) \in Z \ (x = 0 \Rightarrow y = 0)$
('no cornucopia' condition),

(G3) $\forall (x, y) \in Z \ (x' \geqq x \Longrightarrow (x', y) \in Z) \ \& \ \forall (x, y) \in Z 0 \leqq y' \leqq y \Longrightarrow (x, y') \in Z$
(possibility of wasting the inputs/outputs),

(G4) production space Z is a closed subset of R_{+}^{2n}.

The Gale production set is a closed cone in R_{+}^{2n} with a vertex at 0. If $(x, y) \in Z$ and $x = 0$, then, according to (G2), also $y = 0$. We are only interested in processes $(x, y) \in Z \backslash \{0\}$. The number

$$\alpha(x, y) = \max\{\alpha | \ \alpha x \leqq y\}$$

is called the index of the technological efficiency of process $(x, y) \in Z \backslash \{0\}$. It follows from the definition that function $\alpha(\cdot)$ is non-negative and positively homogeneous of degree 0 on $Z \backslash \{0\}$.

□ Theorem 1. If conditions (G1)–(G4) are satisfied, then a solution to the problem exists:

$$\max_{(x,y) \in Z \backslash \{0\}} \alpha(x, y) = \alpha(\bar{x}, \bar{y}) \geq 0.$$

For proof, see Panek (2022, th.1), Takayama (1985, th. 6.A.1). ■

The number α_M is called the optimal indicator of the technological production efficiency. Process $(\bar{x}, \bar{y}) \in Z \backslash \{0\}$ is called the optimal production process. In the stationary Gale economy it is determined with the accuracy of a multiplication

[4] Due to the technology that the economy has at its disposal.

by a positive constant (with a structure accuracy); if $\alpha(\bar{x},\bar{y}) = \alpha_M$, then $\forall \lambda > 0(\alpha(\lambda\bar{x},\lambda\bar{y}) = \alpha_M)$.

We are interested in an economy where in optimal production process $(\bar{x},\bar{y})$ all commodities are produced and the production of the commodities exceeds (on all coordinates) the inputs.

This is ensured by the following condition:
(G5) $\exists(\bar{x},\bar{y}) \in Z\backslash\{0\}(\alpha(\bar{x},\bar{y}) = \alpha_M > 1 \,\&\, \bar{y} > 0)$.

An economy that satisfies condition (a) is called regular, and an economy which meets condition (b) is called productive. If condition (G5) is met, then due to (G3):

$$\exists(\bar{x},\bar{y}) \in Z\backslash\{0\}(\bar{y} = \alpha_M\bar{x} > 0).$$

Everywhere else, when we talk about optimal process $(\bar{x},\bar{y})$, we mean the production process that meets the above-mentioned condition. We say that vector

$$\bar{s} = \frac{\bar{y}}{\|\bar{y}\|}$$

represents the production structure in optimal process $(\bar{x},\bar{y})$[5]. Ray

$$N = \{\lambda\bar{s}|\lambda > 0\} \subset R^n_+$$

is called the production turnpike (the von Neumann ray) in the stationary Gale economy.

By $p = (p_1, \dots, p_n) \geq 0$ we denote the commodity price vector in the Gale economy. Let $(x,y) \in Z\backslash\{0\}$. Then $\langle p,x\rangle = \sum_{i=1}^n p_i x_i$ is the inputs value and $\langle p,y\rangle = \sum_{i=1}^n p_i y_i$ the production value in process (x,y). The number

$$\beta(x,y,p) = \frac{\langle p,y\rangle}{\langle p,x\rangle}$$

$(\langle p,x\rangle \neq 0)$ is called the index of the economic efficiency of process (x,y). Let $(\bar{x},\bar{y}) \in Z\backslash\{0\}$ be the optimal production process in the Gale economy. Then

$$\alpha_M\bar{x} = \bar{y} > 0. \tag{1}$$

[5] If $a \in R^n_+\backslash\{0\}$, then $\|a\| = \sum_{i=1}^n a_i$ and $\frac{a}{\|a\|} = \left(\frac{a_1}{\|a\|}, \dots, \frac{a_n}{\|a\|}\right)$.

□ Theorem 2. If conditions (G1)–(G5) are satisfied, then such a price vector $\bar{p} \geq 0$ exists that

$$\forall(x, y) \in Z(\langle\bar{p}, y\rangle \leq \alpha_M \langle\bar{p}, x\rangle). \tag{2}$$

For proof, see e.g. Panek (2003; chap. 5, th. 5.4). ■

Since in optimal process $(\bar{x}, \bar{y})$ the production vector is positive and the price vector is at least semi-positive, then

$$\langle\bar{p}, \bar{y}\rangle > 0. \tag{3}$$

From (1)–(3) it follows that

$$\beta(\bar{x}, \bar{y}, \bar{p}) = \frac{\langle\bar{p},\bar{y}\rangle}{\langle\bar{p},\bar{x}\rangle} = \max_{(x,y)\in Z\backslash\{0\}} \beta(x, y, \bar{p}) = \alpha_M.$$

We say that the triple $\{\alpha_M, (\bar{x}, \bar{y}), \bar{p}\}$ represents the (optimal) von Neumann equilibrium state in the stationary Gale economy. Price vector $\bar{p}$ is called the von Neumann price vector. In the equilibrium state, the technological production efficiency matches its economic efficiency (at the maximum possible level of α_M that can be achieved by the economy).

In the von Neumann equilibrium state production process $(\bar{x}, \bar{y})$ and price vector $\bar{p}$ are determined with a structure accuracy (multiplication by a positive constant).

To ensure the uniqueness of turnpike N, we assume that the economy satisfies the following condition:

(G6) $\forall(x, y) \in Z\backslash\{0\}(x \notin N \Rightarrow \beta(x, y, \bar{p}) < \alpha_M)$.

Condition $x \notin N = \{\lambda\bar{s} | \lambda > 0\}$ holds if and only if $\frac{x}{\|x\|} \neq \bar{s}$. Therefore, if in a certain production process the inputs structure differs from the turnpike structure, then according to (G6), the economic efficiency of such a process is lower than optimal.[6]

□ Lemma 1. If conditions (G1)–(G6) are satisfied, then

[6] If conditions (G1)–(G6) are satisfied, then not only input vector $\bar{x}$ and output vector $\bar{y}$, but also the von Neumann equilibrium price vector $\bar{p}$ is positive.

$$\forall \varepsilon > 0 \; \exists \delta_\varepsilon \in (0, \alpha_M) \; \forall \, (x, y) \in$$
$$\in Z \left(\left\| \frac{x}{\|x\|} - \bar{s} \right\| \geq \varepsilon \Rightarrow \beta(x, y, \bar{p}) = \frac{\langle \bar{p}, y \rangle}{\langle \bar{p}, x \rangle} \leq \alpha_M - \delta_\varepsilon \right).$$

For proof, see: Radner (1961), Takayama (1985; chap. 7), Panek (2003; chap. 5, lemma 5. 2). ■

3. Dynamics. Feasible, stationary and optimal growth processes

We assume that time t is discrete, $t = 0, 1, \dots$. We denote the input vector (or the production factors vector) that is used in the economy in period t by $x(t) = \big(x_1(t), \dots, x_n(t)\big)$, and the output vector (or the production vector) that is produced in period t by $y(t) = \big(y_1(t), \dots, y_n(t)\big)$. From the assumption that $\big(x(t), y(t)\big) \in Z \backslash \{0\}$ it follows that it is possible to produce production vector $y(t)$ from input vector $x(t)$ in period t. The economy is closed in the sense that inputs $x(t+1)$ (that are incurred in the next period) come from production $y(t)$ (produced in the previous period), i.e.

$$x(t+1) \leqq y(t), \quad t = 0, 1, \dots.$$

Hence, according to (G3), it follows that

$$\big(y(t), y(t+1)\big) \in Z \backslash \{0\}, \quad t = 0, 1, \dots. \tag{4}$$

Let y^0 be a given (positive) initial production vector (produced in period $t = 0$),

$$y(0) = y^0 > 0 \tag{5}$$

and

$$Y^1 = \{y \in R^n_+ | y \geqq y^1\}, \tag{6}$$

be a fixed target set of the desired states (production vectors); $y^1 > y^0$.

Each production vector sequence $\{y(t)\}_{t=0}^{t_1}$ satisfying conditions (4), (5) and the following condition:

$$y(t_1) \in Y^1 \tag{7}$$

is called (y^0, Y^1, t_1) – the feasible growth process.

Feasible process $\{y^*(t)\}_{t=0}^{t_1^*}$ is called (y^0, Y^1, t_1^*) − optimal, if it is a solution to the following minimal-time growth problem:

$$\begin{gathered} \min t_1 \\ \text{subject to (4), (5), (7),} \end{gathered} \tag{8}$$

in which vector y^0 and set Y^1 are fixed.

In problem (8), production vectors $y(1), \dots, y(t_1)$ and time t_1 are the decision variables. This problem has a solution if assumptions (G1)–(G6) are met (Panek, 2021).

If conditions (G1)–(G6) are satisfied, and particularly $y^0 = \bar{y} \in N$, then a growth process exists (satisfying conditions (4), (5)) of the following form:

$$\bar{y}(t) = \alpha_M^t \bar{y}, \quad t = 0, 1, \dots, \tag{9}$$

which is called the stationary growth process with the α_M rate.[7] Since in such a process the following condition is satisfied:

$$\forall t \left(\bar{s}(t) = \frac{\bar{y}(t)}{\|\bar{y}(t)\|} = \frac{\alpha_M^t \bar{y}}{\|\alpha_M^t \bar{y}\|} = \frac{\bar{y}}{\|\bar{y}\|} = \bar{s} \right),$$

we therefore say that it is characterised by a constant (turnpike) production structure. Each stationary growth process lies on turnpike N. The production of all the commodities in such a process increases at a maximum rate of $\alpha_M > 1$ achievable by the economy. This fact still plays an important role in the proof of the 'very strong' turnpike theorem in the next section.

4. 'Very strong' turnpike effect

Let us introduce the following (angular) distance measure of production vector $y(t)$ from turnpike $N = \{\lambda \bar{s} | \lambda > 0\}$:

$$d(y(t), N) = \left\| \frac{y(t)}{\|y(t)\|} - \bar{s} \right\|.$$

[7] The stationary growth process exists if and only if condition $(\bar{y}, \alpha_M \bar{y}) \in Z \backslash \{0\}$ is satisfied. This condition is fulfilled in our model.

In a paper by Panek (2021, Th. 4), we proved the 'weak' turnpike theorem which states that if conditions (G1)–(G6) apply, and (*) such a number $M < +\infty$ exists that regardless of the distance between target states set Y^1 and initial state y^0, each vector $y^1 > y^0$ determining the shape of this set (see (6)) satisfies condition $\frac{\max_i y_i^1}{\min_i y_i^1} \leq M$, then – regardless of the distance between target states set Y^1 from initial state y^0 – the production structure in each optimal growth process (y^0, Y^1, t_1^*), i.e. the solution to problem (8), almost always[8] differs slightly, in an arbitrary way, from the turnpike production structure on which the economy develops at its maximum rate, achieving the highest possible technological and economic efficiency. According to condition (*), y^1 is any production vector (greater than initial vector y^0) in which with $\|y^1\| \to +\infty$, the distance (range) between the values of its coordinates does not increase 'too rapidly' (i.e. no faster than linearly).[9]

We will now trace the trajectory of optimal growth process $\{y^*(t)\}_{t=0}^{t_1^*}$, which in a certain time period of $\check{t} < t_1^* - 1$ reaches turnpike N, when condition (*) is replaced with the following condition:[10]

(G7) vector $y^1 > y^0$, on which the set of target states Y^1 depends, satisfies

$$\frac{\max_i \frac{y_i^1}{\bar{s}_i}}{\min_i \frac{y_i^1}{\bar{s}_i}} \leq \alpha_M.$$

For the proof of Theorem 3, the following lemma will be necessary.

□ Lemma 2. Let us assume that $(y^0, Y^1, t_1^*) -$ optimal growth process $\{y^*(t)\}_{t=0}^{t_1^*}$ and the solution to problem (8), in a certain period $\check{t} < t_1^* - 1$, reaches the turnpike:

$$y^*(\check{t}) \in N.$$

If conditions (G1)–(G7) are satisfied, then there exists such a $(y^0, Y^1, t_1) -$ a feasible process $\{\tilde{y}(t)\}_{t=0}^{t_1}$:

---

[8] Except for a number of time periods, independent of Y^1 or t_1^*.

[9] For example, if we have sequence of problems (8) with the sets of target states $Y^{1,i} = \{y \in R_+^n | y \geqq y^{1,i} > y^0\}$, $\|y^{1,i}\| \underset{i}{\to} +\infty$, then condition (G7) excludes the situation in which for a certain k-th coordinate, $y_k^{1,i} \underset{i}{\to} \bar{y}_k^1 < +\infty$, and for a certain (different) j-th coordinate, $y_j^{1,i} \underset{i}{\to} +\infty$.

[10] By vector y^1 we may also understand any production vector greater than initial vector y^0, in which with $\|y^1\| \to +\infty$ the distance between the values of its coordinates, relativised with respect to the coordinates of turnpike production structure vector $\bar{s}$, increases no faster than linearly with coefficient $M = \alpha_M$.

$$\tilde{y}(t) = \begin{cases} y^*(t), t = 0, 1, \dots, \check{t}, \\ \alpha_M^{t-\check{t}} y^*(\check{t}), t = \check{t} + 1, \dots, t_1 \end{cases}, \tag{10}$$

$t_1 \geq t_1^*$, that

$$\tilde{y}(t_1) \geqq y^1 \text{ and } \tilde{y}(t_1 - 1) \leqq y^1.$$

Proof.[11] If assumptions (G1)–(G7) are satisfied, then there exists (y^0, Y^1, t_1) – a feasible growth process (10), in which $\tilde{y}(t_1) \geqq y^1$, and t_1 is the smallest natural number which satisfies the condition

$$t_1 \geq \tau_1, \tag{11}$$

where $\tau_1 = \check{t} + \frac{\ln A_1}{\ln \alpha_M}$, $A_1 = \max_i \frac{y_i^1}{y_i^*(\check{t})} = \sigma^{-1} \max_i \frac{y_i^1}{\bar{s}_i} > 1$, $\sigma = \|y^*(\check{t})\| > 0, \bar{s} = \frac{y^*(\check{t})}{\|y^*(\check{t})\|}$.

In this process, $\tilde{y}(\check{t}) = y^*(\check{t})$ and $\forall t \in \{\check{t}, \check{t} + 1, \dots, t_1\}$ $(\tilde{y}(t) \in N)$. Let us denote by l_{y^1} the smallest number (not necessarily natural) for which the following inequality holds:

$$\alpha_M^{\tau_1 - \check{t} - l_{y^1}} y^*(\check{t}) \leqq y^1. \tag{12}$$

Such a number exists (since $y^1 > 0$ and $\alpha_M > 1$) and

$$l_{y^1} = \tau_1 - \check{t} - \frac{\ln A_2}{\ln \alpha_M} \geq 0,$$

where $A_2 = \min_i \frac{y_i^1}{y_i^*(\check{t})} = \sigma^{-1} \min_i \frac{y_i^1}{\bar{s}_i} > 0$ $(A_2 \leq A_1)$. From (G7) it follows that

$$\frac{A_1}{A_2} \leq \alpha_M, \text{ or } \frac{\ln A_2}{\ln \alpha_M} \geq \frac{\ln A_1}{\ln \alpha_M} - 1,$$

and hence, according to (11), we come to the conclusion that

$$l_{y^1} = \tau_1 - \check{t} - \frac{\ln A_2}{\ln \alpha_M} = \check{t} + \frac{\ln A_1}{\ln \alpha_M} - \check{t} - \frac{\ln A_2}{\ln \alpha_M} \leq 1.$$

[11] The proof is partially based on the proof of Lemma 3 from the paper by Panek (2021). The sequence elements (10) starting from $t = \check{t}$ belong to stationary growth process (9) with the α_M rate and initial production vector $\bar{y}(0) = \bar{y} = \alpha_M^{-\check{t}} y^*(\check{t}) \in N$.

Thus, each target production vector y^1 that satisfies condition (G7) corresponds to such a non-negative number $l_{y^1} \leq 1$ that condition (12) is satisfied. Particularly,

$$\tilde{y}(t_1 - 1) = \alpha_M^{t_1 - \check{t} - 1} y^*(\check{t}) \leqq \alpha_M^{\tau_1 - \check{t} - l_{y^1}} y^*(\check{t}) \leqq y^1,$$

and this concludes the proof. ■

In Theorem 3 we prove that if (y^0, Y^1, t_1^*) – optimal growth process in a certain time period $\check{t}$ meets condition $\check{t} < t_1^*$, then everywhere else, except possibly in the last period t_1^*, it remains on the turnpike.

□ Theorem 3. Suppose the following applies:

- conditions (G1)–(G7) are satisfied;
- (y^0, Y^1, t_1^*) – optimal growth process in a certain time period $\check{t} < t_1^* - 1$ reaches the turnpike:
  $y^*(\check{t}) \in N$,
- the solution to problem (8) is unique,
  then

$$\forall t \in \{\check{t} + 1, \dots, t_1^* - 1\}(y^*(t) \in N). \tag{13}$$

Proof. Let us consider any such (y^0, Y^1, t_1^*) – optimal process $\{y^*(t)\}_{t=0}^{t_1^*}$ where $t_1^* > \check{t} + 1$. From (2) and from the definition of the optimal growth process we have

$$\langle \bar{p}, y^*(t+1) \rangle \leq \alpha_M \langle \bar{p}, y^*(t) \rangle, t = 0, 1, \dots, t_1^* - 1,$$

and hence, in particular,

$$\langle \bar{p}, y^*(t_1^*) \rangle \leq \alpha_M^{t_1^* - \check{t}} \langle \bar{p}, y^*(\check{t}) \rangle. \tag{14}$$

Let us assume that $y^*(\tau) \notin N$ for a certain $\tau \in \{\check{t} + 1, \dots, t_1^* - 1\}$. Then

$$\exists \varepsilon > 0 \left(d(y^*(\tau), N) = \left\| \frac{y^*(\tau)}{\|y^*(\tau)\|} - \bar{s} \right\| = \varepsilon > 0 \right).$$

According to Lemma 1, such a $\delta_\varepsilon \in (0, \alpha_M)$ number exists that

$$\langle \bar{p}, y^*(\tau + 1) \rangle \leq (\alpha_M - \delta_\varepsilon) \langle \bar{p}, y^*(\tau) \rangle. \tag{15}$$

From (14), (15) we obtain the upper limit of the value of the outputs produced in period t_1^*

$$\langle \bar{p}, y^*(t_1^*)\rangle \leq \alpha_M^{t_1^*-\check{t}-1}(\alpha_M - \delta_\varepsilon)\langle \bar{p}, y^*(\check{t})\rangle. \tag{16}$$

On the other hand, according to Lemma 2, there exists such a $(y^0, Y^1, t_1)-$ feasible process following (9) that satisfies the condition

$$\tilde{y}(t_1 - 1) = \alpha_M^{t_1-\check{t}-1} y^*(\check{t}) \leqq y^1.$$

Then

$$\langle \bar{p}, y^*(t_1^*)\rangle \geq \langle \bar{p}, y^1\rangle \geq \langle \bar{p}, \tilde{y}(t_1 - 1)\rangle = \alpha_M^{t_1-\check{t}-1}\langle \bar{p}, y^*(\check{t})\rangle. \tag{17}$$

From (16), (17) (according to $t_1 \geq t_1^*$), we obtain the following inequality:

$$\alpha_M - \delta_\varepsilon \geq \alpha_M^{t_1-t_1^*}. \tag{18}$$

If $t_1 = t_1^*$, then process $\{\tilde{y}(t)\}_{t=0}^{t_1}$ is (y^0, Y^1, t_1^*) $-$ optimal.[12] For $t_1 = t_1^* + 1$, we have $\alpha_M - \delta_\varepsilon \geq \alpha_M$, therefore $\delta_\varepsilon \leq 0$, in contradiction to our assumption. If $t_1 - t_1^* = k \geq 2$, then from (18) we get $\alpha_M - \delta_\varepsilon > \alpha_M^k$, which contradicts condition $\alpha_M > 1$. The obtained contradictions close the proof. ■

5. Final remarks

The necessity to leave the turnpike by optimal process $\{y^*(t)\}_{t=0}^{t_1^*}$, i.e. the solution to minimal-time growth problem (8), in final period t_1^* results simply from the postulate that the economy should reach the set of target states Y^1. In a particular case, when target production vector y^1 that determines the form of target state set (6) is located on the turnpike, $y^1 \in N$, condition (13) holds also for $t = t_1^*$.

The following version of Theorem 3 without the postulate of uniqueness of the solution also remains true:

□ Theorem 3'. Suppose the following applies:

- conditions (G1)–(G7) are satisfied;

[12] We deal with a similar situation in the paper by Panek (2021; Th. 3). When $t_1 = t_1^*$, then $\delta_\varepsilon < \alpha_M - 1$, which cannot be excluded, because $\alpha_M > 1$. Therefore, one of the assumptions of this theorem is the condition of the solution unique.

- a certain (y^0, Y^1, t_1^*) – optimal growth process in period $\check{t} < t_1^* - 1$ reaches the turnpike,

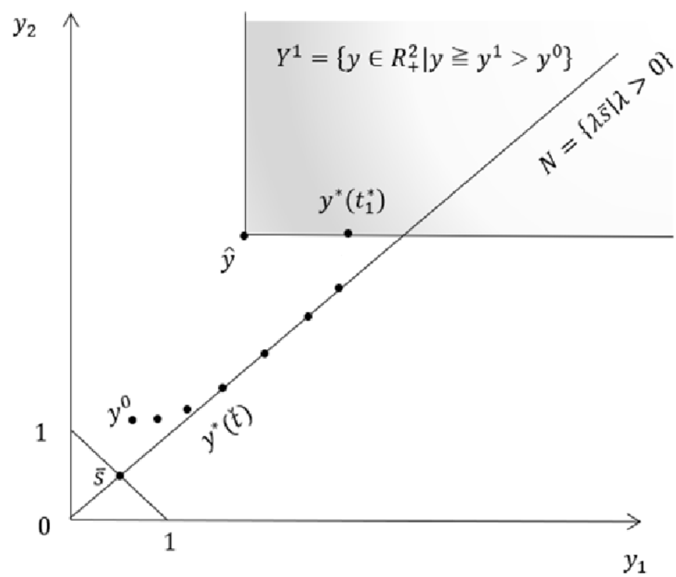
then there also exists such a (y^0, Y^1, t_1^*) – optimal growth process $\{y^*(t)\}_{t=0}^{t_1^*}$ that

$$\forall t \in \{\check{t} + 1, \dots, t_1^* - 1\}(\ y^*(t) \in N).$$

The proof here is exactly the same as the proof of Theorem 3. ■

An example trajectory of (y^0, Y^1, t_1^*), i.e. the optimal growth process in $Z \subset R_+^2$ satisfying the conditions of Theorem 3 is illustrated in the Figure.

Figure Illustration to Theorem 3. The trajectory of the (y^0, Y^1, t_1^*) optimal growth process and the solution to problem (8) in the neighbourhood of turnpike $N = \{\lambda\bar{s}|\lambda > 0\} \subset \subset Z \subset R_+^2$.



Source: author's work.

6. Conclusions

In many papers devoted to the asymptotic/turnpike properties of the optimal growth processes in von Neumann-Gale-Leontief economies, production utility is assumed to be the growth criterion. The novelty of the approach proposed in this article, like in the earlier paper by Panek (2021), consists in replacing the utility of production as the standard quality criterion of economic growth processes by a minimum-time growth criterion (minimising the time needed by the economy to reach the postulated/desired target state). It was proven that changing the growth criterion does not deprive the Gale economy of its asymptotic/turnpike properties.

It would be interesting to study the turnpike properties of the solutions to the minimal-time growth problems of type (8) also in a non-stationary Gale economy with changing technology and multilane production turnpike, especially in the Gale economy with an investment mechanism (see Panek, 2022).

Probably the solution to minimal-time growth problem (8) is also characterised by a 'strong' turnpike effect (as in the case of many other optimal growth processes in the Gale economy with the maximisation of the production utility criterion).

References

Babaei, E. (2019). *Von Neumann-Gale Dynamical Systems with Applications in Economics and Finance* [Doctoral dissertation]. University of Manchester. https://www.research.manchester.ac.uk/portal/files/159166728/FULL_TEXT.PDF.

Babaei, E., Evstigneev, I. V., & Schenk-Hoppé, K. R. (2020). Log-optimal and rapid paths in von Neumann-Gale dynamical systems. *Journal of Mathematical Analysis and Application*, *481*(2), 1–30. https://doi.org/10.1016/j.jmaa.2019.123489.

Giorgi, G., & Zuccotti, C. (2016). *Equilibrium and Optimality in Gale-von Neumann Models* (DEM Working Papers No. 118). http://economiaweb.unipv.it/wp-content/uploads/2017/06/DEMWP0119.pdf.

Jensen, M. K. (2012). Global Stability and the "Turnpike" in Optimal Unbounded Growth Models. *Journal of Economic Theory*, *147*(2), 802–832. https://doi.org/10.1016/j.jet.2010.07.010.

Khan, M. A., & Piazza, A. (2011). An overview of turnpike theory: towards the disconunted deterministic case. In S. Kusuoka & T. Maruyama (Eds.), *Advanced in Mathematical Economics* (vol. 14, pp. 39–67). Springer. https://doi.org/10.1007/978-4-431-53883-7_3.

Majumdar, M. (2009). Equilibrium and optimality: Some imprints of David Gale. *Games and Economic Behavior*, *66*(2), 607–626. https://doi.org/10.1016/j.geb.2009.04.018.

Makarov, V. L., & Rubinov, A. M. (1977). *Mathematical Theory of Economic Dynamics and Equilibria*. Springer-Verlag. https://doi.org/10.1007/978-1-4612-9886-1.

McKenzie, L. W. (2005). Optimal Economic Growth, Turnpike Theorems and Comparative Dynamics. In K. J. Arrow & M. D. Intriligator (Eds.), *Handbook of Mathematical Economics* (2nd edition, vol. 3), 1281–1355.

Mitra, T., & Nishimura, K. (Eds.). (2009). *Equilibrium, Trade, and Growth. Selected Papers of Lionel W. McKenzie*. The MIT Press. https://doi.org/10.7551/mitpress/7848.001.0001.

Nikaido, H. (1968). *Convex Structures and Economic Theory* (vol. 51). Academic Press.

Panek, E. (2003). *Ekonomia matematyczna*. Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Panek, E. (2011). O pewnej prostej wersji „słabego" twierdzenia o magistrali w modelu von Neumanna. *Przegląd Statystyczny*, *58*(1–2), 75–87. https://ps.stat.gov.pl/PS/2011/1-2/2011_58_1-2_075-087.pdf.

Panek, E. (2014). Model gospodarki Gale'a ze zmienną technologią, rosnącą efektywnością produkcji i szczególną postacią kryterium wzrostu. „Słaby" efekt magistrali. *Przegląd Statystyczny*, *61*(4), 325–334. https://ps.stat.gov.pl/PS/2014/4/2014_61_4_325-334.pdf.

Panek, E. (2015). A turnpike theorem for a non-stationary Gale economy with limit technology. A particular case. *Eonomics and Business Review*, *15*(4), 3–13. https://doi.org/10.18559/ebr.2015.4.1.

Panek, E. (2019). A Non-stationry Gale economy with Limit Technology, Multilane Turnpike and General Form of Optimality Criterion. *Argumenta Oeconomica Cracoviensia*, (1), 9–22. https://doi.org/10.15678/AOC.2019.2001.

Panek, E. (2021). The minimal-time growth problem and turnpike effect in the stationary Gale economy. *Economics and Business Review*, *21*(1), 7–25. https://doi.org/10.18559/ebr.2021.1.2.

Panek, E. (2022). Gale Economy with Investments and Limit Technology. *Central European Journal of Economic Modelling and Econometrics*, *14*(1), 57–80. https://doi.org/10.24425/cejeme.2022.140512.

Radner, R. (1961). Paths of Economic Growth that are Optimal with Regard only to Final States: A Turnpike Theorem. *Review Economic Studies*, *28*(2), 98–104. https://doi.org/10.2307/2295707.

Takayama, A. (1985). *Mathematical economics* (2nd edition). Cambrige University Press.