CrORR

Croatian Operational Research Review

vol. 16 / no. 2 / 2025

ISSN (print): 1848 – 0225

ISSN (online): 1848 – 9931

CRORS

CROATIAN OPERATIONAL RESEARCH SOCIETY

ISSN: 1848–0225 Print
ISSN: 1848–9931 Online
UDC: 519.8 (063)
Abbreviation: Croat. Oper. Res. Rev.

Publisher:
CROATIAN OPERATIONAL RESEARCH SOCIETY

Co-publishers:
University of Zagreb, Faculty of Economics & Business
J. J. Strossmayer University of Osijek, Faculty of Economics in Osijek
University of Split, Faculty of Economics, Business and Tourism
J. J. Strossmayer University of Osijek, School of Applied Mathematics and Informatics
University of Zagreb, Faculty of Organization and Informatics

Croatian Operational Research Review

Volume 16 (2025)
Number 2
pp. 99-203

Zagreb, 2025

Croatian Operational Research Review is viewed/indexed in: Current Index to Statistics, DOAJ, EBSCO host, EconLit, Genamics Journal Seek, INSPEC, Mathematical Reviews, Current Mathematical Publications (MathSciNet), Proquest, SCOPUS, Web of Science Emerging Sources Citation Index (WoS ESCI), Zentralblatt für Mathematik/Mathematics Abstracts (CompactMath).

ISSN: 1848–0225 Print
ISSN: 1848–9931 Online
UDC: 519.8 (063)
Abbreviation: Croat. Oper. Res. Rev.
Edition: 40 copies

Croatian Operational Research Review

Croatian Operational Research Review (CRORR) is an open access scientific journal published bi-annually. Starting from 2014, Number 1, the main publisher of the CRORR journal is the Croatian Operational Research Society (CRORS). Co-publishers of the journal are:

- University of Zagreb, Faculty of Economics & Business,
- J. J. Strossmayer University of Osijek, Faculty of Economics in Osijek,
- University of Split, Faculty of Economics, Business and Tourism,
- J. J. Strossmayer University of Osijek, School of Applied Mathematics and Informatics and
- University of Zagreb, Faculty of Organization and Informatics.

Aims and Scope of the Journal

The aim of the Croatian Operational Research Review journal is to provide high quality scientific papers covering the theory and application of operational research and related areas, mainly quantitative methods and machine learning. The scope of the journal is focused, but not limited to the following areas: combinatorial and discrete optimization, integer programming, linear and nonlinear programming, multiobjective and multicriteria programming, statistics and econometrics, macroeconomics, economic theory, games control theory, stochastic models and optimization, banking, finance, insurance, simulations, information and decision support systems, data envelopment analysis, neural networks and fuzzy systems, and practical OR and application.
Occasionally, special issues are published dedicated to a particular area of OR or containing selected papers from the International Conference on Operational Research.
Papers should be written according to journal guidelines available at https://hrcak.srce.hr/ojs/index.php/crorr/about/submissions#authorGuidelines. Papers blindly reviewed and accepted by two independent reviewers are published in this journal.

Publication Ethics and Publication Malpractice Statement

CRORR journal is committed to ensuring ethics in publications and quality of articles. Conformance to standards of ethical behaviour is therefore expected of all parties involved: Authors, Editors, Reviewers, and the Publisher. Before submitting or reviewing a paper read our Publication ethics and Publication Malpractice Statement available at the journal homepage: http://www.hdoi.hr/crorr-journal.
The journal is financially supported by the Ministry of Science, Education and Sports of the Republic of Croatia, and by its co-publishers. All data referring to the journal with full papers texts since the beginning of its publication in 2010 can be found in the Hrcak database – Portal of scientific journals of Croatia (see http://hrcak.srce.hr./crorr).

Open Access Statement

Croatian Operational Research Review is an open access journal which means that all content is freely available without charge to the user or his/her institution. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles in the journal without asking prior permission from the publisher or the author. This is in accordance with the BOAI definition of open access. The journal has an open access to full text of all papers through our Open Journal System (http://hrcak.srce.hr/ojs/index.php/crorr) and the Hrcak database (http://hrcak.srce.hr/crorr).

Croatian Operational Research Review is viewed/indexed by:

- ◇ Current Index to Statistics
- ◇ Directory of Open Access Journals (DOAJ)
- ◇ EBSCO host
- ◇ EconLit
- ◇ Genamics Journal Seek
- ◇ INSPEC
- ◇ Mathematical Reviews, Current Mathematical Publications (MathSciNet)
- ◇ Proquest
- ◇ SCOPUS
- ◇ Web of Science Emerging Sources Citation Index (WoS ESCI)
- ◇ Zentralblatt für Mathematik/Mathematics Abstracts (CompactMath)

Republic of North Macedonia, e-mail: vcvetkoska@eccf.ukim.edu.mk
Maja Ćukušić, Faculty of Economics, Business and Tourism, University of Split, Croatia, e-mail: maja.cukusic@efst.hr
Anita Čeh Časni, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: aceh@efzg.hr
Mirjana Čižmešija, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: mcizmesija@efzg.hr
Ozren Despić, Aston Business School, College of Business and Social Sciences, UK, e-mail: o.despic@aston.ac.uk
Mateja Đumić, School of Applied Mathematics and Informatics, J. J. Strossmayer University of Osijek, Croatia, e-mail: mdjumic@mathos.hr
Margareta Gardijan Kedžo, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: mgardijan@efzg.hr
Danijel Grahovac, School of Applied Mathematics and Informatics, J. J. Strossmayer University of Osijek, Croatia, e-mail: dgrahova@mathos.hr
Tihomir Hunjak, Faculty of Organization and Informatics, University of Zagreb, Croatia, e-mail: tihomir.hunjak@foi.hr
Mario Jadrić, Faculty of Economics, Business and Tourism, University of Split, Croatia, e-mail: jadric@efst.hr
Nikola Kadoić, Faculty of Organization and Informatics, University of Zagreb, Croatia, e-mail: nikola.kadoic@foi.unizg.hr
Vedran Kojić, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: vkojic@efzg.hr
Ulrike Leopold-Wildburger, Institut für Statistik un Operations Research, Karl-Franzens-Universität Graz, Austria, e-mail: ulrike.leopold@uni-graz.at
Goran Lešaja, Department of Mathematical Sciences, Georgia Southern University, USA, e-mail: goran@georgiasouthern.edu
Zrinka Lukač, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: zlukac@efzg.hr
Robert Manger, Department of Mathematics, University of Zagreb, Croatia, e-mail: manger@math.hr
Branka Marasović, Faculty of Economics, Business and Tourism, University of Split, Croatia, e-mail: branka.marasovic@efst.hr
Milica Maričić, Faculty of Organisational Sciences, Serbia, University of Belgrade, e-mail: milica.maricic@fon.bg.ac.rs
Markéta Matulová, Department of Applied Mathematics and Computer Science, Faculty of Economics and Administration, Czech Republic, e-mail: Marketa.Matulova@econ.muni.cz
Josip Mesarić, Faculty of Economics in Osijek, J. J. Strossmayer University of Osijek, Croatia, e-mail: mesaric@efos.hr
Tea Mijač, Faculty of Economics, Business and Tourism, University of Split, Croatia, e-mail: tea.mijac@efst.hr
Luka Neralić, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: lneralic@efzg.hr
Ivan Papić, School of Applied Mathematics and Informatics, J. J. Strossmayer University of Osijek, Croatia, e-mail: ipapic@mathos.hr
Snježana Pivac, Faculty of Economics, Business and Tourism, University of Split, Croatia, e-mail: spivac@efst.hr
Krunoslav Puljić, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: kpuljic@efzg.hr
Jose Rui Figueria, Instituto Superior Tecnico, Technical University of Lisbon, Portugal, e-mail: figueira@ist.utl.pt

KRISTIAN SABO, School of Applied Mathematics and Informatics, J. J. Strossmayer University of Osijek, Croatia, e-mail: ksabo@mathos.hr
RUDOLF SCITOVSKI, School of Applied Mathematics and Informatics, J. J. Strossmayer University of Osijek, Croatia, e-mail: scitowsk@mathos.hr
MARC SEVAUX, Université de Bretagne-Sud, Lorient, France, e-mail: marc.sevaux@univ-ubs.fr
KENNETH SÖRENSEN, Department of Engineering Management, University of Antwerp, Belgium, e-mail: kenneth.sorensen@uantwerpen.be
PETAR SORIĆ, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: psoric@efzg.hr
GREYS SOŠIĆ, Information and Operations Management Department, The Marshall School of Business, University of Southern California, Los Angeles, CA, USA, e-mail: sosic@marshall.usc.edu
ALEMKA ŠEGOTA, Faculty of Economics, University of Rijeka, Croatia, e-mail: alemka.segota@efri.hr
TEA ŠESTANOVIĆ, University of Split, Faculty of Economics, Business and Tourism, Croatia, e-mail: tea.sestanovic@efst.hr
BLANKA ŠKRABIĆ PERIĆ, Faculty of Economics, Business and Tourism, University of Split, Croatia, e-mail: blanka.skrabic@efst.hr
KRISTINA ŠORIĆ, Rochester Institute of Technology Croatia, Croatia, e-mail: ksoric@zsem.hr
BEGOÑA VITORIANO, Complutense University of Madrid, Faculty of Mathematical Sciences, Spain, e-mail: bvitoriano@mat.ucm.es
SILVIJA VLAH JERIĆ, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: svlah@net.efzg.hr
RICHARD WENDELL, Joseph M. Katz Graduate School of Business, University of Pittsburgh, USA, e-mail: wendell@katz.pitt.edu
LIDIJA ZADNIK STIRN, Biotechnical Faculty, University of Ljubljana, Slovenia, e-mail: Lidija.Zadnik@bf.uni-lj.si
SANJO ZLOBEC, Department of Mathematics and Statistics, McGill University, Canada, e-mail: zlobec@math.mcgill.ca
NIKOLINA ŽAJDELA HRUSTEK, University of Zagreb, Faculty of Organization and Informatics, Croatia, e-mail: nikolina.zajdela@foi.unizg.hr
BERISLAV ŽMUK, Faculty of Economics & Business, University of Zagreb, Croatia, e-mail: bzmuk@efzg.hr

ISSN: 1848–0225 Print
ISSN: 1848–9931 Online
UDC: 519.8 (063)

Croatian Operational Research Review

Volume 16 (2025), Number 2

CONTENTS

Original scientific papers

Zagreb, 2025

A memetic NSGA-III for green flexible production with real-time energy costs and emissions

Sascha C Burmeister

Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany
E-mail: ⟨sascha.burmeister@upb.de⟩

Abstract. The use of renewable energies strengthens decarbonization strategies. To integrate volatile renewable sources, energy systems require grid expansion, storage capabilities, or flexible consumption. This study focuses on industries that adapt production to real-time energy markets, offering flexible consumption to the grid. Flexible production considers not only traditional goals like minimizing production time, but also minimizing energy costs and emissions, thereby enhancing the sustainability of businesses. However, existing research focuses on single goals, neglects the combination of makespan, energy costs, and emissions, or assumes constant or periodic tariffs instead of a dynamic energy market. We present a novel memetic NSGA-III to minimize makespan, energy cost, and emissions, integrating real energy market data, and allowing manufacturers to adapt energy consumption to current grid conditions. Evaluating it with benchmark instances from literature and real energy market data, we explore the trade-offs between objectives, showcasing potential savings in energy costs and emissions on estimated Pareto fronts.

Keywords: green flexible job shop scheduling, memetic nsga-iii, or in energy, sustainable production

Received: June 20, 2024; accepted: November 29, 2024; available online: February 5, 2025

DOI: 10.17535/crorr.2025.0009

Original scientific paper.

1. Introduction

Renewable energy is a key solution to the challenge of climate change. However, traditional energy systems face high penetration of volatile renewable energy sources and face new challenges in terms of power quality, reliability, or power system reliability [2]. To support the integration of renewable energy sources, energy systems need grid expansion, storage capabilities, or flexible loads. In the context of flexible loads, consumers are encouraged to shift energy demand away from peak times to relieve grid load and toward periods of high renewable generation, promoting sustainable energy consumption. For manufacturing industries as large energy consumers, flexible loads offer a high potential for cost savings [16]. Additionally, the sustainability of energy production can also be considered. Instead of pricing emitted emissions through CO_2 taxes and thus implicitly incorporating them into energy costs, it can be beneficial to explicitly account for emissions in production planning. This allows focused attention on a company's carbon footprint, which is advantageous for reasons such as anticipating or responding to regulatory pressures, attracting environmentally conscious customers, and maintaining legitimacy with external stakeholders [6]. This work focuses on the challenge manufacturers face in aligning their production schedule with fluctuations in energy prices and emissions of the energy market.

In the energy market, there are three classic types of energy tariffs: (1) fixed energy tariffs with constant energy prices, (2) time-of-use (TOU) tariffs, where energy prices are divided into several time periods (e.g., on-peak and off-peak prices), and (3) real-time pricing (RTP) tariffs, where the cost of electricity consumption is based on the electricity exchange and changes at least every hour. The latter allows manufacturers to align their production with the energy market, respond to price signals, adapt flexibly to the needs of the power system, and thereby reduce energy cost and emissions [13].

In the field of operations research, classical scheduling problems prioritize economic factors such as the schedule's makespan. Green Job Shop Scheduling Problems are an extension of classical scheduling problems, incorporating resource and environmental aspects [17]. As we show in Section 2, recent research places different focuses on economic goals such as the makespan or energy costs and ecological goals such as emissions. To the best of our knowledge, no research has combined the minimization of makespan, energy cost, and emissions while considering RTP tariffs. However, integrating these factors is essential for production planners to accurately calculate the potential reductions in energy costs and emissions achievable through improved production flexibility.

In this work, we focus on the effects of flexible production by exploring the research question: How do scheduling decisions that prioritize one objective — makespan, energy cost, or emissions — affect the others? Our contributions include (1) the formulation of a Flexible Job Shop Scheduling Problem (FJSP) with dynamic energy cost and emissions, (2) developing a memetic algorithm based on the Non-Dominated Sorting Algorithm III (NSGA-III), and (3) conducting computational experiments to evaluate the trade-offs among makespan, energy costs, and emissions. In our study, we continue the work of [4], who formulate a model for the bicriteria FJSP with respect to makespan and energy cost and present a memetic NSGA-II as a solution approach. Our mathematical model extends this framework to include emissions as a third objective. Additionally, our memetic NSGA-III enhances the existing memetic NSGA-II, enabling the calculation of multi-objective schedules aimed at minimizing makespan, energy cost, and emissions. This extension allows us to evaluate trade-offs among all three objectives in our computational experiments and quantify potential savings for practitioners.

As a solution approach, we opt for memetic NSGA-III due to its ability to represent schedules in a three-dimensional Pareto front, allowing decision-makers to balance trade-offs and select solutions based on their preferences. It also leads to favorable results in similar FJSP problems [20, 25] and has already been used to optimize schedules with respect to energy-related goals [23]. A notable aspect of our study is the evaluation of the scheduling problem using real energy costs and emissions data from the German energy market.

The remainder of this paper is organized as follows. Section 2 presents recent research in the area of green scheduling. Section 3 introduces the mathematical model, for which we present a memetic NSGA-III in Section 4. Section 5 shows our computational experiments and discusses our results. Section 6 summarizes our results and recommends directions for future research.

2. Recent research

In this section, we present research related to energy cost- and emission-aware scheduling. We discuss the respective objectives and solution approaches presented in the literature.

A schedule can be designed to minimize makespan, emissions, energy cost, or a combination of these. In order to take ecological and economic goals into account, one approach is to minimize makespan while limiting energy consumption [5]. Energy consumption can also be minimized as an objective, but is still independent of the underlying energy mix and its costs and emissions [18, 23]. Jiang and Wang [24] incorporate an economic view of energy consumption and minimize the makespan and energy costs. They divide a day into several periods with different energy prices based on a TOU tariff. A more detailed consideration of dynamic prices with RTP tariffs can be found in Abikarram et al. [1] and Fazli Khalaf and Wang [11]. They consider hourly-changing prices but neglect a minimization of the emissions and the makespan of the production schedule. Burmeister et al. [4] present a model for bicriteria optimization of makespan and energy cost, considering an RTP tariff.

Due to the NP-hard complexity of Job Shop Scheduling Problems [15], many studies favor metaheuristics over exact solution methods, allowing fast adaptation to fluctuating prices and emissions. Schulz et al. [22] use a multiphase iterated local search algorithm to determine a Pareto front regarding makespan, total energy costs and peak load. Most studies on energy-efficient scheduling reviewed in [14] employ swarm intelligence and evolutionary algorithms. Dong and Ye [10] design an improved hybrid salp swarm and NSGA-III algorithm to reduce carbon emissions and energy costs under the TOU tariff. Lu et al. [18] and Jiang and Wang [24] apply a multi-objective Genetic Algorithm, while Burmeister et al. [4] and Sun et al. [23] use multi-objective evolutionary metaheuristics for fast computation.

Recent research has focused on several areas of green scheduling. However, the reviewed works that include energy costs and emissions neglect the real-time energy market, while studies related to real-time energy markets do not include emissions. We aim to fill this research gap by combining economic and environmental perspectives, taking into account a dynamic energy market.

3. Mathematical model

In this section, we present the mathematical optimization model for the multi-objective FJSP with the objectives of minimizing makespan, energy cost, and emissions. The model is based on the research conducted by [4], which we extend to take into account emissions. Table 1 presents the notation for the mathematical model. The set $J = \{1, ..., \mu\}$ contains μ jobs to be processed, each of which is divided into v_i operations $O_i = \{(i, 1), ..., (i, v_i)\}$. The operations of a job must be processed in sequence. The set $O = \bigcup_{i \in J} O_i$ contains all operations. For each operation, τ_{ijk} specifies the duration with which an operation (i, j) can be processed on machine $k \in M$. Supplementary, η_{ijkt} and ζ_{ijkt} contain the energy cost and emissions for processing operation (i, j) on machine $k \in M$ beginning at time $t \in T$.

Notation	Description	Notation	Description
Sets		*Variables*	
J	Jobs, $i \in J$	c^{max}	Maximum makespan
O	Operations, $O = \bigcup\limits_{i \in J} O_i$,	p^{sum}	Sum of all energy cost
	$O_i = \{(i, 1), ..., (i, \nu_i)\}$	e^{sum}	Sum of all emissions
M	Machines, $k \in M$	s_{ijk}	Start time of operation (i, j) on machine k
T	Time steps, $t \in T$		
Parameters		c_{ijk}	End time of operation (i, j) on machine k
L	A large number		
τ_{ijk}	Processing time of operation (i, j) on machine k	x_{ijk}	Binary indicator, 1 iff operation (i, j) is allocated on machine k
η_{ijkt}	Energy cost for processing operation (i, j) on machine k when starting at time t	$y_{iji'j'k}$	Binary indicator, 1 iff operation (i, j) is predecessor of operation (i', j') on machine k
ζ_{ijkt}	Energy emissions for processing operation (i, j) on machine k when starting at time t	p_{ijkt}	Binary indicator, 1 iff operation (i, j) starts on machine k at time t

Table 1: *Notation for the math. formulation*

Objective function 1 minimizes the variables c^{max}, p^{sum}, and e^{sum}, which represent the maximum makespan, the sum of all energy cost, and the sum of all emissions, respectively. Constraint 2 reflects the final completion time across all operations $(i, j) \in O$ and machines $k \in M$. Constraints 3 and 4 sum the energy cost η_{ijkt} and emissions ζ_{ijkt}, respectively, for all operations $(i, j) \in O$ on all machines $k \in M$ at all time steps $t \in T$.

Constraints 5 to 10 are based on the MILP formulation for the general FJSP of [19]. They ensure that each operation is assigned to exactly one machine (Constraint 5), that operations of a job can only start if previously required operations have been completed (Constraints 6 to 8), and that operations of a machine cannot overlap (Constraints 9 and 10).

To consider the energy consumption in the model, we add a link from the allocation of the operations to their individual energy consumption: Constraint 11 forces the sum over binary indicators p_{ijkt} over all time steps t to be one, if operation (i, j) is assigned to machine k. Constraints 12 and 13 ensure that the binary indicator p_{ijkt} is set to one for the time t at which the operation (i, j) starts.

The model is composed of binary and continuous variables. Together with convex and linear constraints, it is classified as an MILP.

$$
\begin{aligned}
&\min\ (c^{max}, p^{sum}, e^{sum}) && && (1)\\
&\text{s.t.}\ \ c^{max} \geq c_{ijk} && \forall i, j, k && (2)\\
&\quad p^{sum} \geq \textstyle\sum_{i,j,k,t} \eta_{ijkt} p_{ijkt} && && (3)\\
&\quad e^{sum} \geq \textstyle\sum_{i,j,k,t} \zeta_{ijkt} p_{ijkt} && && (4)\\
&\quad \textstyle\sum_k x_{ijk} = 1 && \forall i, j && (5)\\
&\quad s_{ijk} + c_{ijk} \leq x_{ijk} L && \forall i, j, k && (6)\\
&\quad c_{ijk} \geq s_{ijk} + \tau_{ijk} - (1 - x_{ijk}) L && \forall i, j, k && (7)\\
&\quad \textstyle\sum_k s_{ijk} \geq \sum_k c_{i,j-1,k} && \forall i, j && (8)\\
&\quad s_{ijk} \geq c_{i'j'k} - y_{iji'j'k} L && \forall i, j, i', j', k && (9)\\
&\quad s_{i'j'k} \geq c_{ijk} - (1 - y_{iji'j'k}) L && \forall i, j, i', j', k && (10)\\
&\quad x_{ijk} = \textstyle\sum_t p_{ijkt} && \forall i, j, k && (11)\\
&\quad s_{ijk} - t \geq -(1 - p_{ijkt}) L && \forall i, j, k, t && (12)\\
&\quad s_{ijk} - t \leq (1 - p_{ijkt}) L && \forall i, j, k, t && (13)\\
&\quad c^{max}, p^{sum}, e^{sum}, s_{ijk}, c_{ijk} \in \mathbb{R}_+, && && \\
&\quad x_{ijk}, y_{iji'j'k}, p_{ijkt} \in \{0, 1\} \qquad \forall i, j, i', j', k, t && && (14)
\end{aligned}
$$

4. The memetic NSGA-III

In this section, we present the memetic NSGA-III for solving the green multi-objective FJSP. We divide the section into two parts, first introducing the representation of solutions in genotypes and phenotypes in Section 4.1, before explaining the algorithm of memetic NSGA-III in Section 4.2.

4.1. Representation of solutions

The NSGA-III is an evolutionary algorithm and represents solutions as individuals of a population that evolve over multiple generations. To represent solutions as individuals, we follow a decoder-based approach that encodes solutions as genotypes and uses phenotypes for decoding. We base our genotype on the work of [7], who choose a bipartite gene string for the sequence of operations and their machine assignment, and [4], who choose a tripartite gene string for the additional representation of the maximum allowable energy cost per operation. To account for emissions, we extend the latter approach and add a fourth gene string to this representation to indicate the maximum allowable emissions per operation.

Figure 1 shows an example genotype for three jobs to be assigned to two machines. The genotype is divided into four strings of genes, representing the sequence in which operations are allocated, the machine allocation, the tolerated energy costs, and the tolerated emissions. Genes linked to jobs 1, 2, and 3 are highlighted in gray, blue, and white, respectively. Each section has a length corresponding to the total number of operations. For decoding the genotype in Figure 1 as a schedule, Figure 2 illustrates the phenotype. The allocation of operations to the machines is plotted as a Gantt chart, with time steps noted on the x-axis and machines on the left y-axis. On the right y-axis are the energy cost values, shown as a dashed line, and the emissions values, shown as a dotted line. Energy cost is given in €/MWh and emissions are given in grams of carbon dioxide equivalent (gCO_2eq) per kWh. The sequence gene string specifies the order in which the jobs are placed, while the machine gene string reflects the machine to be selected. Thus, operation (1,1) is first assigned to machine 2. The operation is scheduled at the first possible time that satisfies both the associated maximum allowable energy cost of the energy cost gene string and the emissions of the emissions gene string. This means that operation (1,1) is scheduled at the first time that costs less than or equal to 1 € and less than or equal to 4 gCO_2eq, which is time step 4.

All jobs are scheduled in the order specified by the sequence gene string. If the time horizon considered is not sufficient to place an operation, e.g., because the allowed energy cost or emissions are too high, either the time horizon can be extended or the energy cost and emissions can be reduced. While the former tends to produce long schedules with favorable energy cost and emissions, the latter tends to produce fast and expensive schedules. To avoid bias and promote population diversification, we alternate the approach in each generation. Overall, the genotype and phenotype can represent any sequence of operations on any combination of machines, energy cost, and emissions, resulting in a complete representation of the solution space.

4.2. Algorithm

In this section, we explain our memetic NSGA-III. We base the algorithm on the NSGA-III [9], which uses a similar framework to its predecessor NSGA-II [8] while providing better diversity.

The NSGA-III is outlined in Alg. 1. For a population P_t of size N, NSGA-III generates an equally sized next generation P_{t+1} using recombination and mutation operators: For recombination, we use a two-point crossover to intensify the search in the solution space and to find better individuals based on existing ones. From two parents, the genes of the gene strings are swapped between two different random positions. The same positions for swapping are chosen for all gene strings. By swapping, the gene strings remain feasible for machine assignment, energy cost, and emissions. In the case of the sequence gene string, swapping may cause an infeasibility, e.g., if a job is no longer listed in the number of its operations. In this case, the defective children are repaired by replacing excess jobs with missing ones. For mutation, the

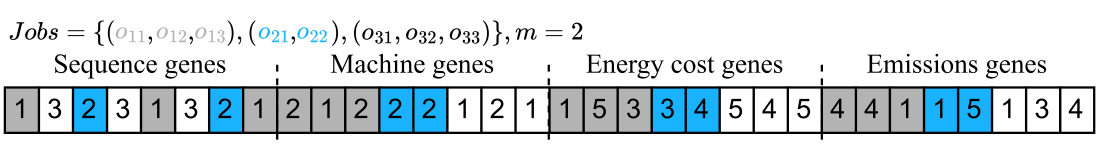


Figure 1: *Example genotype for solution encoding based on [4]*

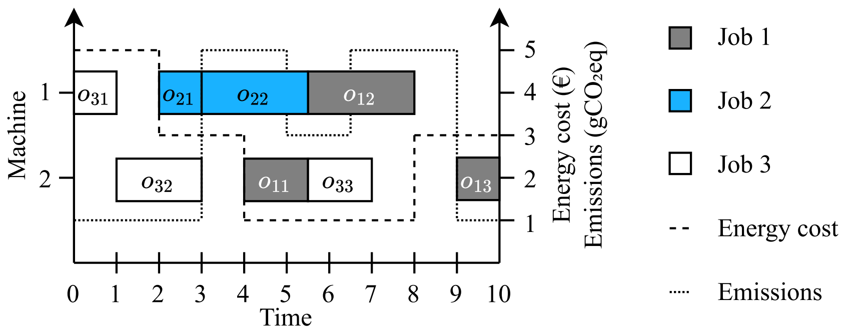


Figure 2: *Representation of the example genotype as a phenotype based on [4]*

algorithm randomly modifies gene strings by performing one of the following changes: swapping two genes within the sequence gene string, reassigning a gene in the machine gene string, or assigning new values to either the energy cost gene string or the emission gene string.

Memetic NSGA-III

Input:
Population P_t, Population size N
begin
$S_t = \emptyset, i = 1$
$R_t \leftarrow P_t \cup Recombination(P_t) \cup Mutation(P_t)$
$R_t \leftarrow LocalRefinement(R_t)$
$(F_1, F_2, ...) = NonDominatedSort(R_t)$
while $|S_t| < N$ **do**
$S_t = S_t \cup F_i$
$i \leftarrow i + 1$
end while
if $|S_t| = N$ **then**
$P_{t+1} \leftarrow S_t$
else
$P_{t+1} \leftarrow \bigcup_{j=1}^{i-1} F_j$
$k \leftarrow N - |P_{t+1}|$ // Remaining space in P_{t+1}
$Z^r \leftarrow Normalization(S_t)$
$Association(Z^r, S_t)$
$P_{t+1} \leftarrow P_{t+1} \cup Niching(F_i, k)$
end else

Algorithm 1: *Pseudocode of the memetic NSGA-III based on [9]*

We extend NSGA-III to a memetic algorithm and adopt local refinement from Burmeister et al. [4]. In their local refinement, a greedy approach improves solutions by adjusting the values of the gene strings for energy cost, ensuring that lower energy cost are achieved without increasing the makespan. We extend this approach to include emissions as shown in Figure 3. (1) First, it sorts all operations of the parent, based on their energy consumption, in descending order into a queue L. (2) Second, it calculates the lower and upper feasible start times l_{cij} and u_{cij} for each operation $(i, j) \in O$ based on the duration of previous and subsequent operations of the job for both children $c \in \{1, 2\}$. (3) Third, the greedy procedure selects the operation with the highest energy consumption, and (4) fourth, schedules the selected operation at the time of the cheapest energy cost (4a) and emissions (4b), respectively. (5) Fifth, it adjusts the earliest possible start and end times of the operations that are still to be sorted, and (6) dequeues the current operation from L. The procedure repeats steps (3) and (4) successively for the next operation until all operations in L have been scheduled.

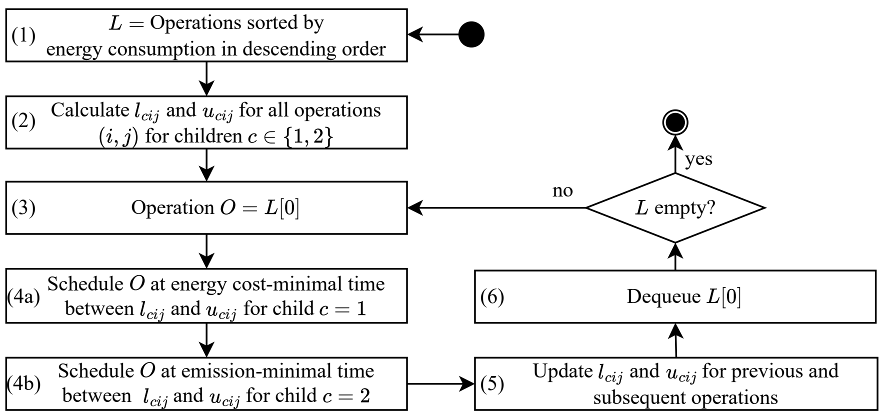


Figure 3: *Greedy local refinement*

After recombination, mutation, and local refinement, the resulting set R_t is sorted into several non-dominated fronts as described in [9]. Algorithm 1 performs a non-dominated sort of individuals on different fronts and successively adds the individuals of F_i to the set S_t until their cardinality is greater than or equal to the desired population size N. If $|S_t| = N$, the next generation P_{t+1} inherits all individuals from S_t. Otherwise, all previous fronts except the last front F_i are added to the new generation. To decide which k individuals remaining in F_i are included in P_{t+1}, the algorithm performs (1) normalization, (2) association, and (3) niching. The three procedures are explained in detail in [9] and are briefly outlined below: (1) First, the three values of the objective function of all individuals in S_t are normalized. (2) Then, the algorithm creates reference points Z^f and associates each individual from S_t with its nearest reference point. (3) Finally, the niching procedure successively adds the individuals from F_i to P_{t+1} with which the fewest individuals are associated. Niching ends when k individuals have been added to the new generation P_{t+1}, so $|P_{t+1}| = N$ holds. The memetic NSGA-III iterates until a termination criterion (e.g., generation or runtime limit) is satisfied. The first front F_1 of the last generation is then the estimate of the Pareto front.

5. Computational experiments

In this section, we present the computational experiments. Section 5.1 describes the experimental setup of scheduling problems in the presence of real-world energy market data. Sections 5.2 to 5.4 present the results and discuss the pairwise relationship of the objective function values, i.e., the relationship between makespan and energy cost, makespan and emissions, and energy cost and emissions. Section 5.5 formulates implications.

5.1. Instances and experimental setting

For the computational experiments, we use the benchmark set of [3]. Table 2 shows the set that contains 15 instances for the FJSP with jobs and their respective operations as well as machines, making it suitable for emulating production schedules. We select a 15-minute duration for each time step. Thus, a schedule covers multiple periods with varying energy prices and emissions.

Instance	Jobs	Machines	Operations per job	Operations in total	Time steps per operation
mk01	10	6	5-7	55	1-7
mk02	10	6	5-7	58	1-7
mk03	15	8	10	150	1-20
mk04	15	8	3-10	90	1-10
mk05	15	4	5-10	106	5-10
mk06	10	10	15	150	1-10
mk07	20	5	5	100	1-20
mk08	20	10	5-10	225	5-20
mk09	20	10	10-15	240	5-20
mk10	20	15	10-15	240	5-20
mk11	30	5	5-8	179	10-30
mk12	30	10	5-10	193	10-30
mk13	30	10	5-10	231	10-30
mk14	30	15	8-12	277	10-30
mk15	30	15	8-12	284	10-30

Table 2: *Benchmark instances by Brandimarte [3]*

For consideration of energy consumption, we add an energy demand of $500\frac{i}{|J|}$ kW for each operation of job $i \in J$, this means the demands of the operations range up to 500 kW. We enrich the instances with real data from the German energy market published by the [12]. Figure 4 shows both electricity prices from the German wholesale market and emission values in hourly resolution from February 1$^{\text{st}}$ to June 30$^{\text{th}}$, 2022. Energy prices are determined by the European Energy Exchange. Emission values are based on lifecycle emissions per generation technology, as outlined in [21]. The Pearson correlation coefficient for energy prices and energy emissions is 0.72.

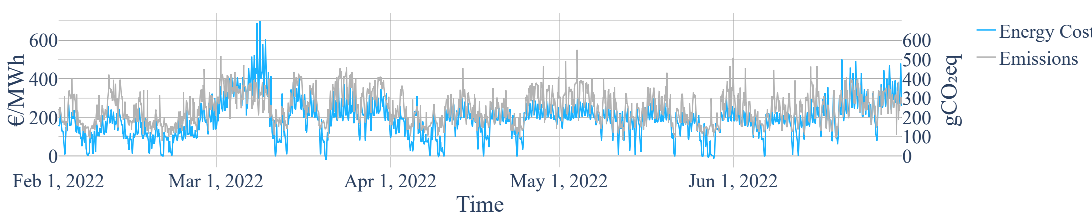


Figure 4: *Electricity prices and emissions of the German energy market [12]*

For parameterization, we follow the settings of [4]. We limit the runtime to 45 minutes to reflect the flexibility to respond to hourly price changes in the energy market. The memetic NSGA-III is implemented in C# 10 within the .NET 6 software framework. The problem is solved on a Red Hat Enterprise Linux 8.5 (Oopta) operating system with an Intel Xeon Gold 6148 CPU, 20x2.4GHz, and 190 GByte main memory.

5.2. Relation of makespan to energy cost

In this section, we focus on the relationship between makespan and energy cost without considering emissions. The first three columns of Table 3 show the instance, the minimum makespan found by the memetic NSGA-III, and the associated energy cost. The remaining columns show the percentage of energy cost that can be saved by increasing the makespan.

A 5% makespan increase saves between 0.1% (mk02) to 14.6% (mk07) of energy cost across all instances. For a 20% makespan increase, the savings range from 2.2% (mk02 and mk14) to 30.4% (mk13). For increases in makespan of 50% and 75%, the energy cost remain stagnant for instances mk03, mk05, mk08, mk11, mk12, and mk14, showing no further improvements. For other instances, energy cost continues to decrease, achieving savings of up to 45.0% (mk06). At a 75% makespan increase, instances mk06 and mk04 show the highest energy cost savings with 49.3% and 51.7%, while instances mk03 and mk14 show the lowest savings (2.5% and 2.2%).

The average savings in energy cost are 5.86%, 12.31%, 18.89%, and 22.35% for makespan increases of 5%, 20%, 50%, and 75%, respectively. The results indicate that small increases in makespan are more efficient for energy cost savings, as the relative increase in savings surpasses the increase in makespan for 9 out of 15 instances. However, efficiency decreases with larger increases in makespan. With a 20% increase in makespan, the savings in energy cost exceed the 20% mark in 2 out of 15 instances. Savings with an increase in makespan of 50% and 75% are below 50% and 75%, respectively, for all instances.

5.3. Relation of makespan to emissions

Table 4 shows the results of the reduction in emissions with increasing makespan and follows the structure of Table 3. A 5% makespan increase saves between 0.4% (mk02) and 10.5% (mk13) of the emissions. With a 20% makespan increase, the savings range from 1.0% (mk14) to 19.1% (mk13). For makespan increases of 50% and 75%, the emissions remain constant for the instances mk03, mk05, mk07, mk08, mk11, and mk14. Other instances show emissions savings of up to 23.7% (mk15) and 24.1% (mk09) with makespan increases of 50% and 75%, respectively.

The average emissions savings are 4.08%, 8.11%, 11.05%, and 12.59% for relative makespan increases of 5%, 20%, 50%, and 75%, respectively. Again, the most efficient savings are observed for smaller instances. For a 5% makespan increase, the relative increase in savings exceeds the increase in makespan for 6 of 15 instances.

The emissions savings are lower than the energy cost savings in Table 3. This can be attributed to the fact that the market values of the emissions have a positive value range, while the energy cost may have negative values. As a result, the algorithm has fewer opportunities to schedule jobs with high energy demands at favorable times when considering emissions.

5.4. Relation of energy cost to emissions

Table 5 shows the results of the emissions savings as energy cost increase without considering the makespan. It is structured similarly to Table 3 and 4. For instance mk02, the absolute value of the energy cost was used because the instance has few jobs that are processed for negative energy cost.

	min		**Increase of ms[1] (%)**			
Inst.	**ms[1]**	**ec[2]**	**5**	**20**	**50**	**75**
mk01	42	3965	2.0	6.5	17.0	22.3
mk02	29	3204	0.1	2.2	2.3	3.1
mk03	204	13954	1.7	2.5	2.5	2.5
mk04	66	6943	0.7	7.7	26.5	51.7
mk05	174	11791	5.0	6.9	6.9	6.9
mk06	77	7988	3.7	15.9	45.0	49.3
mk07	144	10658	14.6	15.9	15.9	15.9
mk08	523	37481	8.9	14.5	14.5	14.5
mk09	341	45041	5.5	16.7	32.5	38.6
mk10	263	40158	7.0	10.2	16.1	26.3
mk11	621	41038	8.6	10.1	10.1	10.1
mk12	524	49137	7.7	18.6	18.6	18.6
mk13	457	71059	11.8	30.4	36.7	36.7
mk14	694	63382	2.2	2.2	2.2	2.2
mk15	429	89821	8.4	24.3	36.6	36.6

[1] Makespan (time steps), [2] Energy cost (€)

Table 3: *Energy cost savings (in %) with increase in makespan*

min			Increase of ms[1] (%)				min			Increase of ms[1] (%)			
Inst.	**ms[1]**	**em[2]**	**5**	**20**	**50**	**75**	**Inst.**	**ms[1]**	**em[2]**	**5**	**20**	**50**	**75**
mk01	42	5.215	1.3	5.4	12.7	17.7	mk09	341	68.240	3.4	11.8	20.4	24.1
mk02	29	4.627	0.4	2.2	3.7	5.3	mk10	263	61.328	5.0	7.6	13.4	21.2
mk03	204	24.257	2.9	2.9	2.9	2.9	mk11	621	73.575	4.7	5.6	5.6	5.6
mk04	66	9.370	1.0	7.9	14.9	17.0	mk12	524	83.272	5.7	10.4	10.7	10.7
mk05	174	18.303	3.5	3.6	3.6	3.6	mk13	457	109.633	10.5	19.1	23.1	23.1
mk06	77	10.436	3.2	11.2	14.7	17.6	mk14	694	118.605	1.0	1.0	1.0	1.0
mk07	144	17.517	6.3	7.1	7.1	7.1	mk15	429	139.310	6.9	17.6	23.7	23.7
mk08	523	64.146	5.4	8.3	8.3	8.3							

[1] Makespan (time steps), [2] Emissions (tons of CO_2eq)

Table 4: *Emissions savings (in %) with increase in makespan*

A 5% energy cost increase saves between 0.7% (mk08) and 5.6% (mk06) of emissions. For increases in energy costs of 20, 50 and 75%, the maximum savings are 8.6 (mk06), 12.1 (mk06), and 14.2% (mk02), respectively, and the minimum savings are 1.4% (mk08). Except for the 5.6% reduction in emissions for a 5% deterioration in energy cost for the mk07 instance, the relative savings are less than the relative deterioration in all other cases. The average emissions savings are 1.89%, 4.52%, 5.92%, and 6.74% for relative energy cost increases of 5%, 20%, 50%, and 75%, respectively.

5.5. Implications

In this section, we draw implications from the results. Figure 5 summarizes the medians of the percentage savings in energy costs when increasing the makespan (solid line), emissions when increasing the makespan (dashed line), and emissions when increasing the energy cost (dotted line). The error bars show the minimum and maximum values.

First, extending the makespan can lead to savings in energy costs and emissions. The percentage values observed in Figure 5 indicate disproportionately high average savings in energy costs when the makespan increases by 5%. In contrast, increases in makespan beyond 5% and reductions in emissions show lower percentage savings compared to the initial increase in makespan. Energy costs show a higher savings potential than emissions, as the energy market can have negative prices, while emissions cannot be negative. Although an increase in makespan results in a disproportionately low increase in the percentage of savings in energy costs and emissions, the absolute savings remain considerable. This is because the underlying absolute energy costs range between €3,204 and €89,821, and emission values range from 4.63 to 139.31 tons of CO_2eq. We advise decision-makers to consider the benefits of enhancing production flexibility for potential savings.

Second, opting for an increase in energy cost to reduce emissions is less effective than

min			Increase of ec[1] (%)				min			Increase of ec[1] (%)			
Inst.	**ec[1]**	**em[2]**	**5**	**20**	**50**	**75**	**Inst.**	**ec[1]**	**em[2]**	**5**	**20**	**50**	**75**
mk01	5	3.249	0.3	1.7	5.4	6.2	mk09	10170	45.452	2.0	5.7	6.1	6.1
mk02[3]	-2	3.098	2.1	6.6	9.4	14.2	mk10	7182	40.082	1.8	5.0	6.7	7.3
mk03	1282	17.464	1.3	6.6	8.1	9.7	mk11	18933	60.929	1.4	2.5	2.5	2.5
mk04	71	6.343	1.6	5.7	8.0	8.9	mk12	16571	62.976	1.0	2.7	2.8	2.8
mk05	838	13.940	2.2	4.0	5.2	6.6	mk13	22309	74.076	3.4	3.6	3.6	3.6
mk06	132	7.506	0.2	8.6	12.1	13.5	mk14	32770	100.836	1.4	2.2	2.2	2.2
mk07	668	13.517	5.6	6.9	10.7	11.5	mk15	29438	93.441	3.3	4.6	4.6	4.6
mk08	13019	48.015	0.7	1.4	1.4	1.4							

[1] Energy cost (€), [2] Emissions (tons of CO_2eq), [3] Since the value is negative, we use the absolute value for the percentage increase.

Table 5: *Emissions savings (in %) with increase in energy cost*

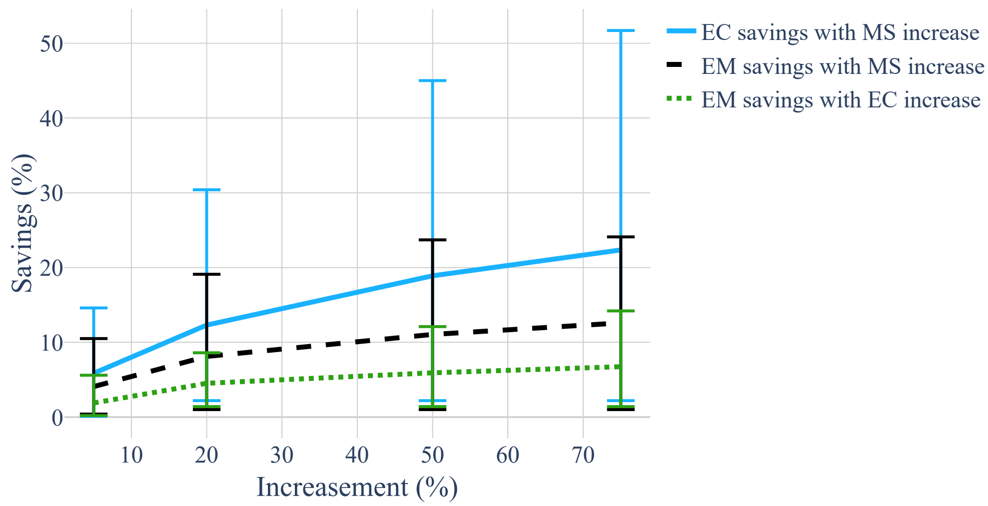


Figure 5: *Minimum, mean and maximum savings*

increasing makespan. The average emission reductions achieved by creating time flexibility with a longer makespan is higher than that resulting from the acceptance of higher energy cost for the purchase of renewable energy. If decision-makers are unable to consider an increase in makespan (e.g., due to deadlines), it may still be attractive for manufacturers to balance higher energy costs against emissions savings. Additional costs for reducing or avoiding emissions could be incorporated into adjusted product pricing, thereby promoting sustainable production without extending the makespan.

6. Conclusion

Our study addresses the challenge of managing flexible loads in manufacturing industries, with the aim of optimizing schedules with respect to makespan, energy cost, and emissions. Drawing from existing literature, we introduce a linear optimization model for a Green FJSP and develop a memetic NSGA-III approach as a three-objective solution method. Evaluating our method on benchmark instances enriched with real energy market data from Germany, we analyze potential savings in energy costs and emissions. This helps production planners assess the benefits of flexible production planning adapted to the energy market. We find that even a small extension of the makespan can lead to substantial savings in energy costs and emissions, with potential savings in energy costs being higher than those in emissions. In addition, emissions can be reduced at the expense of higher energy costs. However, while the potential savings from increased makespan are significant, they do not match the extent of savings achievable through flexible production. Our findings underscore the importance of considering the interplay between the considered objectives to strengthen sustainable production planning in manufacturing companies.

Based on the limitations and findings of this study, we advise avenues for further research. First, we recommend comparing the solutions obtained from our approach with those of other state-of-the-art methods. Given that this work focuses on exploring trade-offs between objectives, our results currently serve as lower bounds for potential savings, suggesting that improved solutions may also be possible. In future statistical analysis, solution quality could be examined by assessing differences in efficiency and solution quality, as well as calculating an optimality gap.

Second, we recommend extending our approach to account for data uncertainty, as the current model assumes deterministic knowledge of energy costs and emissions. In a dynamic variant, schedules could be developed using energy market data and forecasts, allowing real-time adjustments to accommodate evolving grid conditions and renewable energy availability. Investigating a dynamic model would provide insights into how decision-makers can adapt to fluctuations in the energy market.

Third, we recommend that our approach be analyzed on real production data. While we use real energy market data combined with scheduling problems from a benchmark set to analyze savings potentials for different-sized problem cases, we also recommend evaluating our approach with production orders from a real manufacturer. In this way, our goal is to achieve a better understanding of the trade-off between makespan, energy cost, and emissions.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the state of North Rhine-Westphalia, Germany, as part of the *progres.nrw* program area, in the framework of Re^{2}Pli (project number EFO 0127A) and the funding of this project by computing time provided by the Paderborn Center for Parallel Computing (PC^2).

References

[1] Abikarram, J. B., McConky, K., and Proano, R. (2019). Energy cost minimization for unrelated parallel machine scheduling under real time and demand charge pricing. Journal of cleaner production, 208, 232–242. doi: 10.1016/j.jclepro.2018.10.048.

[2] Basit, M. A., Dilshad, S., Badar, R., and Sami ur Rehman, S. M. (2020). Limitations, challenges, and solution approaches in grid-connected renewable energy systems. International Journal of Energy Research, 44 (6), 4132–4162. doi: 10.1002/er.5033.

[3] Brandimarte, P. (1993). Routing and scheduling in a flexible job shop by tabu search. Annals of Operations research, 41 (3), 157–183. doi: 10.1007/BF02023073.

[4] Burmeister, S. C., Guericke, D., and Schryen, G. (2023). A Memetic NSGA-II for the Multi-Objective Flexible Job Shop Scheduling Problem with Real-time Energy Tariffs. Flex Serv Manuf J. doi: 10.1007/s10696-023-09517-7.

[5] Carlucci, D., Renna, P., and Materi, S. (2021). A Job-Shop Scheduling Decision-Making Model for Sustainable Production Planning with Power Constraint. IEEE Transactions on Engineering Management. doi: 10.1109/TEM.2021.3103108.

[6] Dahlmann, F., Branicki, L., and Brammer, S. (2019). Managing carbon aspirations: The influence of corporate climate change targets on environmental performance. Journal of business ethics, 158, 1–24. doi: 10.1007/s10551-017-3731-z.

[7] Dai, M., Tang, D., Giret, A., and Salido, M. A. (2019). Multi-objective optimization for energy-efficient flexible job shop scheduling problem with transportation constraints. Robotics and Computer-Integrated Manufacturing, 59, 143–157. doi: 10.1016/j.rcim.2019.04.006.

[8] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6 (2), 182–197. doi: 10.1109/4235.996017.

[9] Deb, K. and Jain, H. (2014). An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. IEEE Transactions on Evolutionary Computation, 18 (4), 577–601. doi: 10.1109/TEVC.2013.2281535.

[10] Dong, J. and Ye, C. (2022). Green scheduling of distributed two-stage reentrant hybrid flow shop considering distributed energy resources and energy storage system. Computers & Industrial Engineering, 169, 108146. doi: 10.1016/j.cie.2022.108146.

[11] Fazli Khalaf, A. and Wang, Y. (2018). Energy-cost-aware flow shop scheduling considering intermittent renewables, energy storage, and real-time electricity pricing. International Journal of Energy Research, 42 (12), 3928–3942. doi: 10.1002/er.4130.

[12] Federal Network Agency Germany (2024). *SMARD market data*. online. URL: https://www.smard.de/en/downloadcenter/download-market-data.

[13] Finn, P. and Fitzpatrick, C. (2014). Demand side management of industrial electricity consumption: Promoting the use of renewable energy through real-time pricing. Applied Energy, 113, 11–21. doi: 10.1016/j.apenergy.2013.07.003.

[14] Gao, K., Huang, Y., Sadollah, A., and Wang, L. (2020). A review of energy-efficient scheduling in intelligent production systems. Complex & Intelligent Systems, 6, 237–249. doi: 10.1007/s40747-019-00122-6.

[15] Garey, M. R., Johnson, D. S., and Sethi, R. (1976). The complexity of flowshop and jobshop scheduling. Mathematics of operations research, 1 (2), 117–129. doi: 10.1287/moor.1.2.117.

[16] Keller, F., Schultz, C., Simon, P., Braunreuther, S., Glasschröder, J., and Reinhart, G. (2017). Integration and interaction of energy flexible manufacturing systems within a smart grid. Procedia CIRP, 61, 416–421. doi: 10.1016/j.procir.2016.11.250.

[17] Li, M. and Wang, G.-G. (2022). A review of green shop scheduling problem. Information Sciences, 589, 478–496. doi: 10.1016/j.ins.2021.12.122.

[18] Lu, C., Zhang, B., Gao, L., Yi, J., and Mou, J. (2021). A knowledge-based multiobjective memetic algorithm for green job shop scheduling with variable machining speeds. IEEE Systems Journal, 16 (1), 844–855. doi: 10.1109/JSYST.2021.3076481.

[19] Özgüven, C., Özbakır, L., and Yavuz, Y. (2010). Mathematical models for job-shop scheduling problems with routing and process plan flexibility. Applied Mathematical Modelling, 34 (6), 1539–1548. doi: 10.1016/j.apm.2009.09.002.

[20] Sang, Y. and Tan, J. (2022). Intelligent factory many-objective distributed flexible job shop collaborative scheduling method. Computers & Industrial Engineering, 164, 107884. doi: 10.1016/j.cie.2021.107884.

[21] Schlömer, S., Bruckner, T., Fulton, L., Hertwich, E., McKinnon, A., Perczyk, D., Roy, J., Schaeffer, R., Sims, R., Smith, P., et al. (2014). “Annex III: Technology-specific cost and performance parameters”. In: *Climate change 2014: Mitigation of climate change: Contribution of working group III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 1329–1356.

[22] Schulz, S., Neufeld, J. S., and Buscher, U. (2019). A multi-objective iterated local search algorithm for comprehensive energy-aware hybrid flow shop scheduling. Journal of cleaner production, 224, 421–434. doi: 10.1016/j.jclepro.2019.03.155.

[23] Sun, X., Wang, Y., Kang, H., Shen, Y., Chen, Q., and Wang, D. (2021). Modified multi-crossover operator NSGA-III for solving low carbon flexible job shop scheduling problem. Processes, 9 (1), 62. doi: 10.3390/pr9010062.

[24] Wang, L. et al. (2020). Multi-objective optimization based on decomposition for flexible job shop scheduling under time-of-use electricity prices. Knowledge-Based Systems, 204, 106177. doi: 10.1016/j.knosys.2020.106177.

[25] Wu, M., Yang, D., Zhou, B., Yang, Z., Liu, T., Li, L., Wang, Z., and Hu, K. (2021). Adaptive population NSGA-III with dual control strategy for flexible job shop scheduling problem with the consideration of energy consumption and weight. Machines, 9 (12), 344. doi: 10.3390/machines9120344.

On the closed-form solution of the monopolist long-run profit maximization problem with linear demand and Cobb-Douglas technology

Zrinka Lukač[1,*], **Mira Krpan**[1] and **Vedran Kojić**[1]

[1] *Faculty of Economics and Business, University of Zagreb*
Trg J. F. Kennedyja 6, 10000 Zagreb, Croatia
E-mail: ⟨{zlukac, mkrpan, vkojic}@efzg.hr⟩

Abstract. This paper provides an economic and mathematical analysis of the long-run profit maximization problem of a monopolist with linear demand and a two-input Cobb-Douglas technology to derive the conditions that guarantee the solution to this problem. In addition, the conditions under which the closed–form solution can be derived and the conditions under which the monopolist's profit is positive are discussed. Whenever the problem has a unique solution, the closed–form defines the profit function well for the given demand function. The closed–form solution to the problem depends on the returns to scale. The problem has a unique solution for decreasing returns to scale, for which in general no closed–form can be found. On the other hand, the closed–form of the unique solution can easily be found for constant returns to scale. However, three sub-cases are identified for the case of increasing returns to scale. The analysis is supported by economic interpretations and numerical illustrations.

Keywords: closed–form solution, Cobb-Douglas technology, linear demand function, long-run profit maximization, monopoly

Received: May 15, 2024; accepted: December 9, 2024; available online: February 5, 2025

DOI: 10.17535/crorr.2025.0010

Original scientific paper.

1. Introduction

The monopoly maximization problem is essential component of any microeconomic analysis concerning market power, serving as a benchmark in situations where there is no competition and the monopolist controls the entire market demand . It is also a standard element of the economics of regulation, as it represents a reasonable assumption in economic environments dominated by a single large player in single-product markets, even in multi-product markets where neither demand nor costs are linked across markets. In more broader contexts (e.g. the Cournot competition model), it is analyzed as a special case [2, 4, 5, 7, 9, 15].

This paper presents the theoretical analysis of the long-run profit maximization problem faced by a monopolist using linear demand functions and the Cobb-Douglas technology with two inputs. The aim of the paper is to find a closed-form solution to this problem and to analyze the conditions under which the monopolist can achieve positive profits in the long-run. The combination of mathematical analysis and economic interpretation enabled a unique theoretical analysis of the extremely important economic problem of a monopolist's long-run profit maximization at different returns to scale. The practical implications of this work lie in defining and elaborating the conditions under which a solution to the specified long-run

*Corresponding author.

profit maximization problem of a monopolist exists. In addition, the conditions under which a closed-form solution exists, allowing a comparative static analysis are identified.

In a monopoly market, there is only one active seller. In this paper, we assume that no inactive sellers exist in the market (for the theory of contestable markets, see [3]). We also assume that the monopolist is not subject to regulation (for more on regulation theory, see [7]). Moreover, there is no price discrimination in our model, meaning the monopolist charges the same price for each unit of the good. For more information on the monopolist's price discrimination strategies, see [16]. Under these assumptions, the monopolist can choose the profit-maximizing price and quantity combination. If he chooses the profit-maximizing quantity (as in our case), the demand function determines the price charged.

Profit maximization model of a monopolist with a linear inverse demand function is often analyzed in microeconomics theory [6]. However, the monopolist's technology is usually represented by the Cobb-Douglas production function with a specified scale elasticity parameter. To the best of our knowledge, the literature lacks a comprehensive mathematical and economic analysis of the long-run profit maximization problem of a monopolist with linear demand and a two-input Cobb-Douglas production function aimed at deriving the conditions that guarantee a solution to this problem. Additionally, the conditions under which a closed–form solution can be derived and those ensuring positive profit for the monopolist's profit are discussed (here, we define a formula to be in closed–form if it can be expressed using a finite number of operations, including addition, subtraction, multiplication, division, exponentiation to a natural power, and the extraction of roots of natural degree, with the given real coefficients.) Therefore the goal of this paper is to fill this gap. In the framework presented in [17], a perfectly competitive industry uses the monopolist's product as an input. The industry's technology is represented by a CES production function, which includes the substitution parameter. The industry's derived demand function for the monopolist's good is then used in the monopolist's long-run profit maximization problem to analyze how the substitution parameter affects the monopolist's profit margin. The monopolist's technology follows the Cobb-Douglas type.

It is known from the duality theory of microeconomic analysis that the technology of a firm can be equally well represented by its production function and its cost function under certain regularity conditions [8]. Since the monopolist is assumed to be a price taker in the input market, the profit maximization model of a monopolist can be represented in two equivalent ways. In the first model, the monopolist maximizes his profit, by choosing input quantities, with technology represented by the Cobb-Douglas production function. The optimal input quantities are then substituted into the production function to obtain the optimal quantity of output that the monopolist offers to the market given the linear inverse demand function. The starting point of the second model is the derived cost function for the Cobb-Douglas production function. In the profit maximization model, the choice variable is the quantity produced. It is shown that both approaches reduce the problem to solving the equation equating marginal cost and marginal revenue, a well-known necessary condition for profit maximization.

Moreover, the economic analysis and subsequent economic interpretation provide valuable insights into the relationship between returns to scale and the monopolist's profit. Finally, the dual approach to the long-run profit maximization model of a monopolist enriches the analysis and allows us to illustrate this relationship between the monopolist's profit and returns to scale within our framework.

The structure of the paper is as follows. After the introduction and notation, the long-run profit maximization problem of a monopolist with linear demand and a two-input Cobb-Douglas production function is presented. Whenever the problem has a unique solution, the closed–form defines the profit function for the given demand function. It is shown that the closed–form solution to the problem is conditional on economies of scale when demand is linear. This generalization of the monopolist's long-run profit maximization model with linear demand and Cobb-Douglas technology provides a thorough economic and mathematical analysis enriched by economic interpretations. Conditions under which the analyzed problem has a closed–form solution are identified, along with the market environment in which the monopolist continues producing in the long-run. Before the conclusions, some numerical examples are provided.

2. Model and main results

The model assumes that the monopolist faces a downward-sloping linear market demand function for the good it produces

$$p(q) = a - bq, \tag{1}$$

where $a, b > 0$, and p is the product price at which the monopolist can sell q units of the product he produces. The technology of a monopolist is represented by the Cobb-Douglas production function:

$$q = q(x_1, x_2) = x_1^{\alpha} x_2^{\beta}, \tag{2}$$

where q is the quantity of production, $x_1, x_2 \geq 0$ are input quantities, $\alpha > 0$ is the output elasticity of the first input, and $\beta > 0$ is the output elasticity of the second input.

It is assumed that the monopolist is the price taker in the input market, so the input prices $w_1, w_2 > 0$, are parameters of the model and the monopolist takes them as given.

The value of the elasticity of scale, ε, which is a measure of returns to scale for Cobb-Douglas technology, is $\varepsilon = \alpha + \beta$ [12]. Eatwell [10] defines returns to scale as follows: *"The production technique of a good y can be characterized as a function of the inputs x required: If all inputs are multiplied by a positive scalar t and the resulting output is represented as $t^s y$, then the value of s can be taken as an indicator of the level of returns to scale. If $s = 1$, then there are constant returns to scale: Any proportional change in all inputs results in an equally proportional change in output."* Elasticity of scale is greater than 1 when returns to scale are increasing, and less than 1 when returns to scale are decreasing [12].

The monopolist thus chooses the profit-maximizing quantity of production factors and the production quantity. In this context, the profit maximization model can be written in two ways, depending on whether we first want to derive the monopolist's input demand functions for a given demand function or determine the optimal production quantity directly. This distinction arises from duality results in production theory, where the producer's technology can be equivalently represented by either the production function or the cost function under certain regularity conditions. Both formulations are given below.

The long-run profit maximization model of a monopolist when the quantities of the factors of production are the choice variables is represented as:

$$\max_{x_1, x_2 \geq 0} p(q(x_1, x_2)) \cdot q(x_1, x_2) - w_1 x_1 - w_2 x_2, \tag{3}$$

where $w_1, w_2 > 0$ are input prices. In this approach, the technology of a monopolist is represented by the production function. Alternatively, the long-run profit maximization model of a monopolist can be represented in the following way:

$$\max_{q > 0} \Pi(q) = p(q)q - c(q), \tag{4}$$

where $c(q)$ is the cost function. It is known from microeconomic theory that the cost function is the result of an optimization problem in which the producer minimizes costs while taking into account a given level of production.

We demonstrate that both approaches reduce the discussion to the solvability of the equation equating marginal cost and marginal revenue, a well-known necessary condition for profit maximization. Additionally, the following result will be used in the remainder of the paper.

Theorem 1. *(Abel's Theorem) The generic algebraic equation of degree higher than four is not solvable by radicals, i.e., formulae do not exist for expressing roots of a generic equation of degree higher than four in terms of its coefficients by means of operations of addition, subtraction, multiplication, division, raising to a natural power, and extraction of a root of natural degree.*

The proof of Abel's theorem can be found in [1].

2.1. The first approach to solving the long-run profit maximization model of a monopolist

Below the first-order necessary conditions for the model (3) are derived:

$$\frac{\partial p}{\partial x_1}q + p\frac{\partial q}{\partial x_1} - w_1 = 0, \tag{5}$$

$$-b\frac{\partial q}{\partial x_1}q + p\frac{\partial q}{\partial x_1} - w_1 = 0, \tag{6}$$

and

$$\frac{\partial p}{\partial x_2}q + p\frac{\partial q}{\partial x_2} - w_2 = 0, \tag{7}$$

$$-b\frac{\partial q}{\partial x_2}q + p\frac{\partial q}{\partial x_2} - w_2 = 0. \tag{8}$$

Therefore,

$$\frac{\partial q}{\partial x_1}(p - bq) = w_1, \tag{9}$$

$$\frac{\partial q}{\partial x_2}(p - bq) = w_2. \tag{10}$$

Differentiation of the production function with respect to the quantities of inputs gives the marginal products of inputs:

$$\frac{\partial q}{\partial x_1} = \alpha \cdot \frac{q}{x_1}, \tag{11}$$

$$\frac{\partial q}{\partial x_2} = \beta \cdot \frac{q}{x_2}. \tag{12}$$

From (9) and (10) the equality between the input price ratio on the left-hand side and the marginal rate of substitution on the right-hand side is derived

$$\frac{w_1}{w_2} = \frac{\frac{\partial q}{\partial x_1}}{\frac{\partial q}{\partial x_2}} = \frac{\alpha\frac{q}{x_1}}{\beta\frac{q}{x_2}} = \frac{\alpha}{\beta} \cdot \frac{x_2}{x_1}. \tag{13}$$

From (13) the long-run production expansion path follows:

$$x_2 = \frac{\beta}{\alpha} \cdot \frac{w_1}{w_2}x_1. \tag{14}$$

Inserting (11) into (9) gives

$$x_1 = \frac{\alpha}{w_1} \cdot q(p - bq). \tag{15}$$

Due to the symmetry of the initial problem in (3), we get by analogy

$$x_2 = \frac{\beta}{w_2} \cdot q(p - bq). \tag{16}$$

From (1) the marginal revenue function is obtained:

$$MR(q) = \frac{d(p(q)q)}{dq} = \frac{d(aq - bq^2)}{dq} = a - 2bq = (a - bq) - bq = p - bq. \tag{17}$$

If (17) is inserted into (15) and (16), (15) and (16) become

$$x_1 = \frac{\alpha}{w_1} \cdot q(a - 2bq), \tag{18}$$

$$x_2 = \frac{\beta}{w_2} \cdot q(a - 2bq). \tag{19}$$

For the most interesting case where $x_1 > 0$ and $x_2 > 0$, from (18) and (19) it can be concluded that the following inequality has to be satisfied:

$$a - 2bq > 0 \Longleftrightarrow q \in \left\langle 0, \frac{a}{2b} \right\rangle. \tag{20}$$

Inserting (18) and (19) in (2) gives

$$q = \left(\frac{\alpha}{w_1}\right)^{\alpha} \left(\frac{\beta}{w_2}\right)^{\beta} q^{\alpha+\beta} (a - 2bq)^{\alpha+\beta}. \tag{21}$$

From (21) the following equation is obtained

$$a - 2bq = \rho q^{\frac{1}{\alpha+\beta} - 1}, \tag{22}$$

where the right hand side of equation (22), as we will see below, is the marginal cost of production function for Cobb-Douglas technology, $MC(q)$, and ρ is the marginal cost of 1 unit of output,

$$\rho = \left[\left(\frac{w_1}{\alpha}\right)^{\alpha} \left(\frac{w_2}{\beta}\right)^{\beta} \right]^{\frac{1}{\alpha+\beta}}. \tag{23}$$

Therefore, the term in (22) expresses the equality between marginal revenue (on the left) and marginal cost (on the right), $MR(q) = MC(q)$, the very well known first-order necessary condition for profit maximization. In the next subsection, the same equation is derived using a different approach, and the rest of the paper discusses its solvability.

2.2. The second approach to solving the long-run profit maximization model of a monopolist

In this second approach, the technology is represented by the cost function instead of the production function [8]. In the first step, the cost function is derived and then, in the second step, it is inserted into the long-term profit maximization model with the production quantity as the choice variable. In the following, the cost minimization model is solved to obtain the cost function for the Cobb-Douglas technology.

The cost function is derived from the cost minimization model subject to the given level of production [12]:

$$\begin{cases} \min\limits_{x_1, x_2 \geq 0} & w_1 x_1 + w_2 x_2 \\ \text{s.t.} & q(x_1, x_2) = x_1^{\alpha} x_2^{\beta} = q\,. \end{cases} \tag{24}$$

The solution of the model is the set of conditional input demand functions, $x_1(w_1, w_2, q)$ and $x_2(w_1, w_2, q)$, and the value function of the model is the cost function, $c(w_1, w_2, q) = w_1 x_1(w_1, w_2, q) + w_2 x_2(w_1, w_2, q)$.

The cost function for Cobb-Douglas technology is derived analytically in [17], from which we get the functional forms for the conditional demand functions for inputs and the cost function as given below:

$$x_1(w_1, w_2, q) = \left(\frac{\alpha}{\beta} \cdot \frac{w_2}{w_1}\right)^{\frac{\beta}{\alpha+\beta}} q^{\frac{1}{\alpha+\beta}}, \tag{25}$$

$$x_2(w_1, w_2, q) = \left(\frac{\beta}{\alpha} \cdot \frac{w_1}{w_2}\right)^{\frac{\alpha}{\alpha+\beta}} q^{\frac{1}{\alpha+\beta}}, \tag{26}$$

and

$$c(w_1, w_2, q) = (\alpha+\beta)\left[\left(\frac{w_1}{\alpha}\right)^{\alpha}\left(\frac{w_2}{\beta}\right)^{\beta}\right]^{\frac{1}{\alpha+\beta}} q^{\frac{1}{\alpha+\beta}}. \tag{27}$$

Taking into account the linear inverse demand function and the derived cost function, the profit maximization model (3) reduces to

$$\max_{q\geq 0} = (a - bq)q - (\alpha+\beta)\left[\left(\frac{w_1}{\alpha}\right)^{\alpha}\left(\frac{w_2}{\beta}\right)^{\beta}\right]^{\frac{1}{\alpha+\beta}} q^{\frac{1}{\alpha+\beta}}, \tag{28}$$

which can also be writen as

$$\max_{q\geq 0} = aq - bq^2 - c(w_1, w_2, 1)q^{\frac{1}{\alpha+\beta}}, \tag{29}$$

where

$$c(w_1, w_2, 1) = (\alpha+\beta)\left[\left(\frac{w_1}{\alpha}\right)^{\alpha}\left(\frac{w_2}{\beta}\right)^{\beta}\right]^{\frac{1}{\alpha+\beta}} > 0 \tag{30}$$

is the unit cost function, or the cost of 1 unit of output [12]. Given that $p(q) \geq 0$, $a - bq \geq 0$, it follows that $q \leq \frac{a}{b}$. Next, we derive the necessary first-order conditions by differentiating the objective function in terms of q and setting this derivative equal to zero,

$$\Pi'(q) = a - 2bq - c(w_1, w_2, 1)\frac{1}{\alpha+\beta}q^{\frac{1}{\alpha+\beta}-1} = 0. \tag{31}$$

Thus,

$$a - 2bq = c(w_1, w_2, 1)\frac{1}{\alpha+\beta}q^{\frac{1}{\alpha+\beta}-1} = 0. \tag{32}$$

If we recall the definition of ρ,

$$\rho = c(w_1, w_2, 1)\frac{1}{\alpha+\beta} \text{ and } \varepsilon = \alpha+\beta > 0, \tag{33}$$

the equation (32) can be written in the same form as in (22),

$$a - 2bq = \rho q^{\frac{1}{\varepsilon}-1}. \tag{34}$$

The left-hand side is the marginal revenue function, $MR(q)$, and the right-hand side is the marginal cost function, $MC(q)$. In general, the sufficient conditions for profit maximization of the monopolist reduce to the condition that the $\frac{dMR(q)}{dq} < \frac{dMC(q)}{dq}$, or intuitively that the slope of the marginal revenue curve at the stationary point is less than the slope of the marginal cost curve.

Due to the non-negativity of marginal costs, $\rho q^{\frac{1}{\varepsilon}-1} \geq 0$, marginal revenue cannot be negative either, which brings us back to the condition given by (20). This condition implies that the monopolist cannot maximize his profit on the inelastic part of the demand curve.

The closed–form solution for q depends on the explicit solvability of the equation

$$a - 2bq - \rho q^{\frac{1}{\varepsilon}-1} = 0, \tag{35}$$

which follows from (34). Bellow we comment the solvability of equation (35) as a function of the value of elasticity of scale, as a measure of returns to scale [12].

The exponent $\frac{1}{\varepsilon} - 1$ in equation (35) is a real number. However, since the set of rational numbers $\mathbb{Q}$ is dense in the set of real numbers $\mathbb{R}$, any real number can be approximated to any given degree of accuracy by a rational number. For practical and applicative reasons, we assume that $\varepsilon > 0$ from equation (33) can be written as a positive fraction, i.e.,

$$\varepsilon = \frac{m}{n},\ \gcd(m, n) = 1,\ m, n \in \mathbb{N}. \tag{36}$$

Accordingly, equation (35) becomes

$$\rho q^{\frac{n}{m}-1} = a - 2bq. \tag{37}$$

Theorem 2. *Equation (37) has a closed-form solution if and only if*

$$\varepsilon = \frac{m}{n} \in \left\{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, 1, \frac{3}{2}, 2\right\}. \tag{38}$$

Proof. Using the substitution $t = q^{\frac{1}{m}} > 0$, equation (37) becomes equivalent to

$$\rho t^{n-m} = a - 2bt^m. \tag{39}$$

Let us consider two cases. In the first case, if $n - m \geq 0$, then according to Abel's theorem (Theorem 1), equation (39) has a closed-form solution if and only if its degree is less than or equal to 4. Therefore, the following inequalities must hold:

$$\begin{cases} 1 \leq m \leq 4 \\ 0 \leq n - m \leq 4. \end{cases} \tag{40}$$

Since m and n are positive integers, it is straightforward to find all solutions of (40):

$$\begin{cases} m = 1 \Rightarrow n \in \{1, 2, 3, 4, 5\}, \\ m = 2 \Rightarrow n \in \{2, 3, 4, 5, 6\}, \\ m = 3 \Rightarrow n \in \{3, 4, 5, 6, 7\}, \\ m = 4 \Rightarrow n \in \{4, 5, 6, 7, 8\}. \end{cases} \tag{41}$$

Thus, in the first case, we have

$$\frac{m}{n} \in \left\{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, 1\right\}. \tag{42}$$

In the second case, where $n - m < 0$, we have $m - n > 0$. Multiplying equation (39) by t^{m-n}, it becomes equivalent to

$$\rho = at^{m-n} - 2bt^{2m-n}. \tag{43}$$

Note that $2m - n > m - n > 0$. Similarly to the first case, equation (43) has a closed-form solution if and only if its degree is less than or equal to 4. Thus, the following inequalities must hold:

$$\begin{cases} 1 \leq 2m - n \leq 4 \\ 1 \leq m - n \leq 4. \end{cases} \tag{44}$$

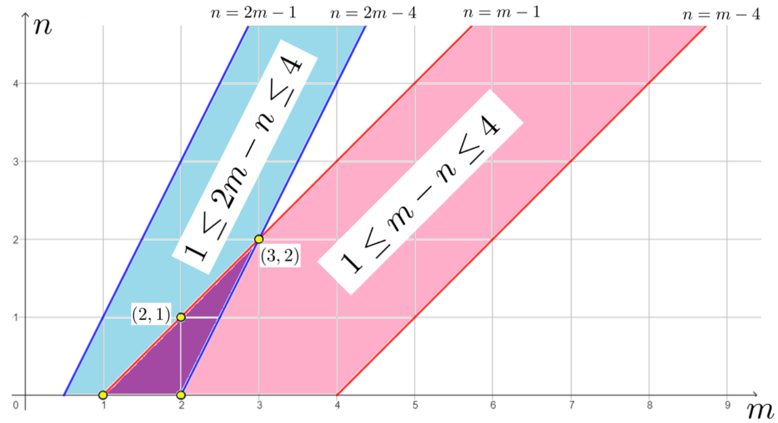


Figure 1: *Solution of the system (44).*

Using a graphical method (see Figure 1), it is easy to find the solutions of the system (44):

$$(m, n) \in \{(2, 1), (3, 2)\} \iff \frac{m}{n} \in \left\{\frac{3}{2}, 2\right\}. \tag{45}$$

The union of the sets in (42) and (45) yields (38), which completes the proof. □

The closed-form solution of equation (37) allows for the derivation of explicit formulas for the input demand functions and the profit function, which are crucial for sensitivity analysis and comparative statics in microeconomics. In the remainder of the paper, we will present several numerical examples to illustrate the solvability of model (4) and equation (37) depending on the value of ε. A full illustration of all values of the parameter ε from (38) is beyond the scope of this paper.

CASE I. ($\varepsilon = 1$ – **CONSTANT RETURNS TO SCALE**)

In the case of constant returns to scale, the elasticity of scale is equal to 1, $\varepsilon = \alpha + \beta = 1$, and equation (35) becomes

$$\Pi'(q) = a - 2bq - \rho = 0, \tag{46}$$

from which it follows that

$$q = \frac{a - \rho}{2b}, \tag{47}$$

where the condition $a > \rho$ must hold. The second order necessary condition reduces to

$$\Pi''(q) = -2b < 0. \tag{48}$$

Since marginal cost is constant in the case of global constant returns to scale, its derivative is intuitively zero. The slope of the marginal revenue function is negative.

Therefore, the supply function is derived for the given demand function:

$$q = q(a, b, w_1, w_2) = \frac{a - \rho}{2b}. \tag{49}$$

The following input demand functions and the profit functions for the given demand function are obtained:

$$x_1 = x_1(a, b, w_1, w_2) = \frac{\rho\alpha}{2bw_1}(a - \rho), \tag{50}$$

$$x_2 = x_2(a, b, w_1, w_2) = \frac{\rho\beta}{2bw_2}(a - \rho), \tag{51}$$

$$\begin{aligned}\pi(a, b, w_1, w_2) &= pq - w_1x_1 - w_2x_2 \overset{(3)}{=} (a - bq)q - w_1\frac{\rho\alpha(a-\rho)}{2bw_1} - w_2\frac{\rho\beta(a-\rho)}{2bw_2}\\ &= (a - bq)q - \underbrace{(\alpha+\beta)}_{=\varepsilon=1}\rho\underbrace{\left(\frac{a-\rho}{2b}\right)}_{=q} = (a - bq)q - \rho q = (a-\rho)q - bq^2 =\\ &= (a-\rho)\frac{a-\rho}{2b} - b\Big(\frac{a-\rho}{2b}\Big)^2 = \frac{(a-\rho)^2}{4b}.\end{aligned} \tag{52}$$

In the second approach, the input demand functions for the given demand function are derived by substituting the optimal output quantity (49) into the conditional input demand functions for the cost-minimising monopolist, as given by (25) and (26). In this way, the closed–form solutions for the input demand functions, the supply function and the profit function for the given demand function are obtained. Moreover, the economic profit in (52) is positive. Intuitively, in the case of constant returns to scale, both the average and marginal costs are constant and equal. In equilibrium, the marginal cost is equal to the marginal revenue and the price exceeds marginal revenue when the demand function is decreasing. As a result, the price is higher than the average cost, ensuring the profit is positive.

CASE II. ($\varepsilon < 1$ – **DECREASING RETURNS TO SCALE**)

In the case of decreasing returns to scale the elasticity of scale is less than 1, $\varepsilon < 1$. The function $q \mapsto \rho q^{\frac{1}{\varepsilon}-1}$, which represents the marginal cost function, is strictly increasing (since $\frac{1}{\varepsilon} - 1 > 0$), but the function $q \mapsto a - 2bq$, which represents the marginal revenue function, is strictly decreasing, so that their graphs intersect only once (Figure 2). Consequently, equation (35) has a unique solution and problem (4) has a unique stationary point.

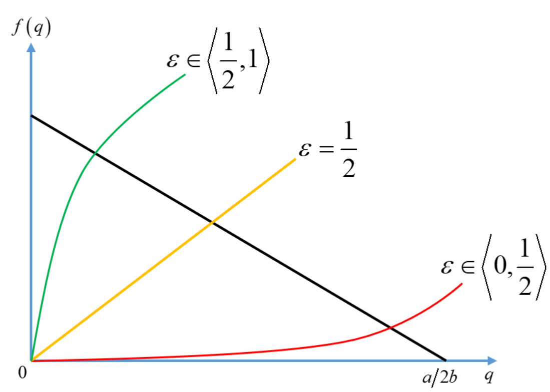


Figure 2: *The marginal revenue and the marginal cost functions in the case of decreasing returns to scale.*

For other values of ε, equation (35) can be solved by numerical methods. Moreover, the sufficient condition is always satisfied when $\alpha + \beta < 1$, since

$$\Pi''(q) = -2b - \rho\Big(\frac{1}{\varepsilon} - 1\Big)q^{\frac{1}{\varepsilon}-2} < 0. \tag{53}$$

The sufficient conditions for the local maximum are satisfied, and the unique stationary point of the problem (4) is the global maximum of the problem (4). Moreover, it can be shown that the global maximum is positive:

$$\begin{aligned} pq^* - w_1x_1^* - w_2x_2^* &= (a - bq^*)q^* - (\alpha + \beta)q^*(a - 2bq^*) \\ &= (1 - \alpha - \beta)(a - 2bq^*)q^* + b(q^*)^2 > 0. \end{aligned} \tag{54}$$

Intuitively, under globally decreasing returns to scale for a single product and perfect competition in markets for factors of production, both marginal and average costs increase as output rises, but with costs exceeding average costs. Since marginal cost equals marginal revenue in equilibrium, and price exceeds marginal revenue in a monopoly, profit must be positive in equilibrium because the price is higher than average cost.

Example 1. *We start with the value of the elasticity of scale equal to 0.5, i.e.* $\varepsilon = \frac{1}{2}$. *In this case* $\rho = \left(\frac{\alpha}{w_1}\right)^{-2\alpha}\left(\frac{\beta}{w_2}\right)^{-2\beta}$*.The equation (35) becomes* $\rho q = a - 2bq$*, and the supply function for the given demand function is derived,*

$$q = q(a, b, w_1, w_2) = \frac{a}{\rho + 2b}. \tag{55}$$

It is easy to see that $\frac{a}{\rho+2b} \in \langle 0, \frac{a}{2b}\rangle$*. The following input demand functions and profit functions for the given demand function are obtained:*

$$x_1 = x_1(a, b, w_1, w_2) = \frac{\alpha\rho}{w_1}\Big(\frac{a}{\rho + 2b}\Big)^2, \tag{56}$$

$$x_2 = x_2(a, b, w_1, w_2) = \frac{\beta\rho}{w_2}\Big(\frac{a}{\rho + 2b}\Big)^2, \tag{57}$$

$$\pi(a, b, w_1, w_2) = \Big(1 - b - \frac{\rho}{2}\Big)\Big(\frac{a}{\rho + 2b}\Big)^2. \tag{58}$$

Example 2. *Let* $\varepsilon = \frac{1}{3}$*. In this case* $\rho = \left(\frac{\alpha}{w_1}\right)^{-3\alpha}\left(\frac{\beta}{w_2}\right)^{-3\beta}$*. The equation (34) becomes* $\rho q^2 = a - 2bq$*, from which, due to* $q > 0$*, the supply function for the given demand function is obtained*

$$q = q(a, b, w_1, w_2) = \frac{1}{\rho}(\sqrt{b^2 + a\rho} - b). \tag{59}$$

It can be shown that $0 < \frac{1}{\rho}(\sqrt{b^2 + a\rho} - b) < \frac{a}{2b}$*. The following input demand functions and the profit functions for the given demand function are obtained:*

$$x_1 = x_1(a, bw_1, w_2) = \frac{\alpha}{\rho^2 w_1}\big(\sqrt{b^2 + a\rho} - b\big)^3, \tag{60}$$

$$x_2 = x_2(a, bw_1, w_2) = \frac{\beta}{\rho^2 w_2}\big(\sqrt{b^2 + a\rho} - b\big)^3, \tag{61}$$

$$\pi(a, b, w_1, w_2) = \frac{\sqrt{b^2 + a\rho} - b}{\rho}\Big(a - \frac{b}{\rho}\big(\sqrt{b^2 + a\rho} - b\big) - \frac{1}{3\rho}\big(\sqrt{b^2 + a\rho} - b\big)^2\Big). \tag{62}$$

CASE III. ($\varepsilon > 1$ – **INCREASING RETURNS TO SCALE**)

In the case of increasing returns to scale, where the scale elasticity is greater than 1, $\varepsilon = \alpha + \beta > 1$, the solvability of equation (35) is discussed below, which reducing the problem to finding the roots of the equation

$$\varphi(q) = \rho q^{\frac{1}{\varepsilon} - 1} + 2bq - a = 0. \tag{63}$$

Note that

$$\varphi(q) = -\Pi'(q). \tag{64}$$

It is easy to show that the function $q \mapsto \rho q^{\frac{1}{3}-1}$, which is the marginal cost function, is strictly decreasing (since $\frac{1}{\varepsilon} < 0$), strictly concave (due to $(\frac{1}{\varepsilon}-1)(\frac{1}{\varepsilon}-2) > 0$), as well as that the following holds: $\lim_{q\to 0+} \rho q^{\frac{1}{\varepsilon}-1} = +\infty$ and $\lim_{q\to +\infty} \rho q^{\frac{1}{\varepsilon}-1} = 0$. Moreover, the function $q \mapsto a - 2bq$, or the marginal revenue function, is strictly decreasing and linear.

Therefore, in search of the roots of the function $\varphi(q)$ three cases may occur: (a) the function φ has no roots ($\varphi > 0$ on the whole domain, Figure 3a, (b) the function φ has one root, Figure 3b, and (c) the function φ has two roots (see Figure 3c).

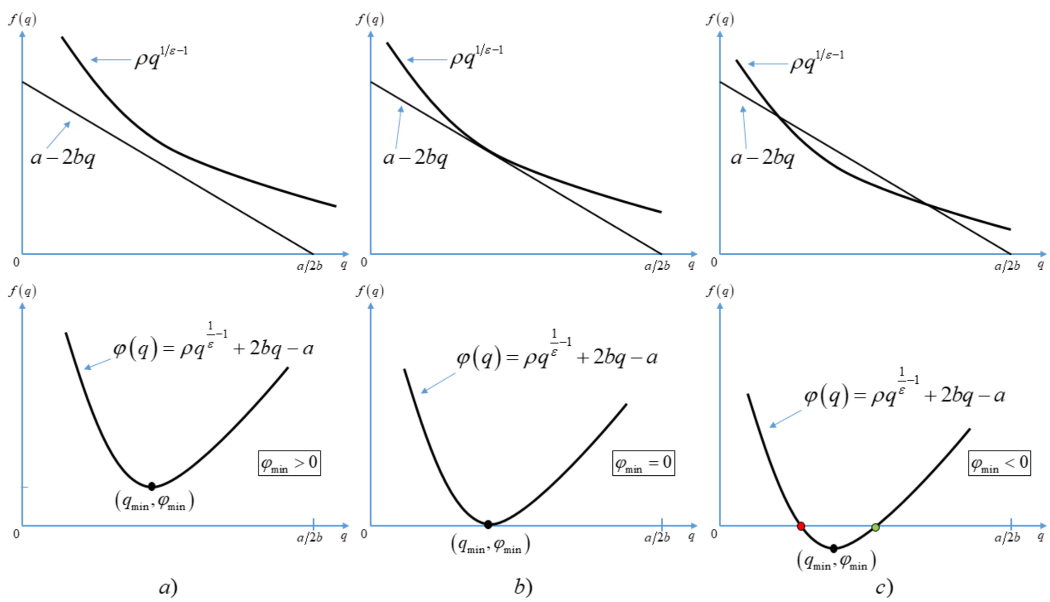


Figure 3: *The marginal revenue and the marginal cost functions in the case of increasing returns to scale.*

Subcase III.a) (the increasing returns to scale and $\varphi_{\min} > 0$)

Additionally, since $\varphi(q) = -\Pi'(q) > 0$, it follows that $\Pi'(q) < 0$. In the long-run, all costs are variable, $\Pi(0) = 0$, and the profit is negative for all $q \in \langle 0, \frac{a}{2b} \rangle$. For every unit of output, the monopolist would lose money since the marginal cost is higher than the marginal revenue.

Example 3. *Let us assume $a = b = \alpha = \beta = w_1 = w_2 = 1$. Then $\varepsilon = 2$, $\rho = 1$, and the problem (3) reduces to*

$$\max_{x_1, x_2 \geq 0} p(q) \cdot q - 1 \cdot x_1 - 1 \cdot x_2 = (1 - x_1 x_2) x_1 x_2 - x_1 - x_2. \tag{65}$$

The objective function (65) has no stationary points and it is negative on the whole domain (see Figure 4).

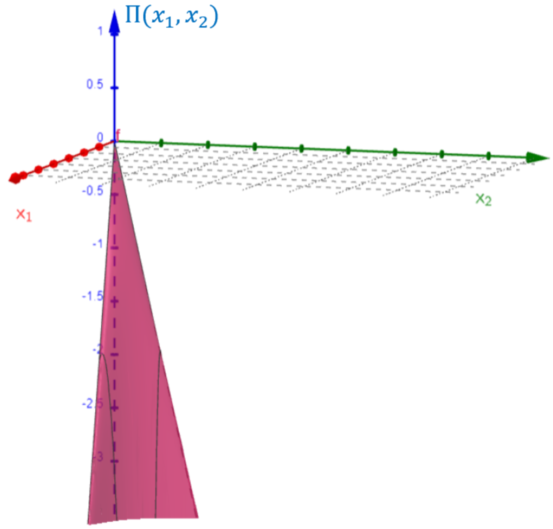


Figure 4: *Negative profit in the case of increasing returns to scale with the scale elasticity equal to 2.*

Subcase III.b) (the increasing returns to scale and $\varphi_{\min} = 0$)
In this case, $\varphi(q) = 0$. From the previous analysis, it is clear that $\varphi_{\min} = 0$ is the value of the strict global minimum of the function φ that is reached at the point $q_{\min}$. Therefore, $q_{\min}$ is both the null point of the function φ and the point of its strict global minimum. The value of $q_{\min}$ can be calculated. Namely, from $\varphi'(q_{\min}) = 0$ the following equation follows: $\left(\frac{1}{\varepsilon} - 1\right)\rho q^{\frac{1}{\varepsilon}-2} + 2b = 0$. Therefore,

$$q_{\min} = \left(\frac{2b}{\rho\left(1 - \frac{1}{\varepsilon}\right)}\right)^{\frac{\varepsilon}{1-2\varepsilon}}, \tag{66}$$

and

$$\varphi_{\min} = (2b)^{\frac{1-\varepsilon}{1-2\varepsilon}} \left(\frac{\varepsilon}{\rho(\varepsilon - 1)}\right)^{\frac{\varepsilon}{1-2\varepsilon}} \cdot \frac{2\varepsilon - 1}{\rho(\varepsilon - 1)} - a. \tag{67}$$

At this point, however, maximum profit is not obtained. Namely, from $\varphi(q_{\min} = 0)$ it follows that $\Pi'(q_{\min}) = 0$. Moreover, from $\varphi'(q_{\min}) = 0$ we have $\Pi''(q_{\min}) = 0$, and from $\varphi''(q_{\min}) > 0$ it follows that $\Pi'''(q_{\min}) \neq 0$. We can conclude that $q_{\min}$ is the inflection point.

Subcase III.c) (the increasing returns to scale and $\varphi_{\min} < 0$)
In this case,

$$\varphi_{\min} < 0 \Leftrightarrow a > (2b)^{\frac{1-\varepsilon}{1-2\varepsilon}} \left(\frac{\varepsilon}{\rho(\varepsilon - 1)}\right)^{\frac{\varepsilon}{1-2\varepsilon}} \cdot \frac{2\varepsilon - 1}{\rho(\varepsilon - 1)}. \tag{68}$$

The global minimum of the function φ is less than zero (negative), so the function φ has two null points, q_1 and q_2, where $q_1 < q_{\min} < q_2$ (Figure 3c). However, at q_1 maximim profit is not achieved. Naimly, from $\varphi(q_1) = 0$ it follows that $\Pi'(q_1) = 0$. It is clear that $\Pi'(q_1) < 0$ since φ is decreasing on $\langle 0, q_{\min}\rangle$, and $\varphi'(q_1) = -\Pi''(q_1) < 0 \Rightarrow \Pi''(q_1) > 0$. Therefore, in q_1 the function achieves local minimum profit.

On the other hand, it is clear that for q_2 it holds that $\varphi'(q_2) > 0$, so $\varphi'(q_2) = -\Pi''(q_2) > 0$ and $\Pi''(q_2) < 0$. Therefore, the function reaches local maximum profit in q_2.

Under conditions of increasing returns to scale, case c, (as specified in condition (68)), represents a necessary condition for maximizing profit. However, this condition applies to both minimum and maximum and says nothing about the positivity of the maximum profit.Consequently, this condition does not guarantee a positive profit. Furthermore, a closed–form solution for the profit function cannot be established. This result is illustrated by the following two examples.

Example 4. *Let us assume that* $a = 5, b = \alpha = \beta = w_1 = w_2 = 1$. *So* $\varepsilon = 2, \rho = 1$, *and the problem (3) becomes*

$$\max_{x_1,x_2\geq 0} p(q)q - 1\cdot x_1 - 1\cdot x_2 = (5 - x_1x_2)x_1x_2 - x_1 - x_2. \tag{69}$$

Condition (68) is satisfied. The solution to the problem (69) is the global maximum. Its value is approximately 3.19501, and it is achieved at point $(x_1, x_2) \approx (1.46962, 1.46962)$ *(Figure 5).*

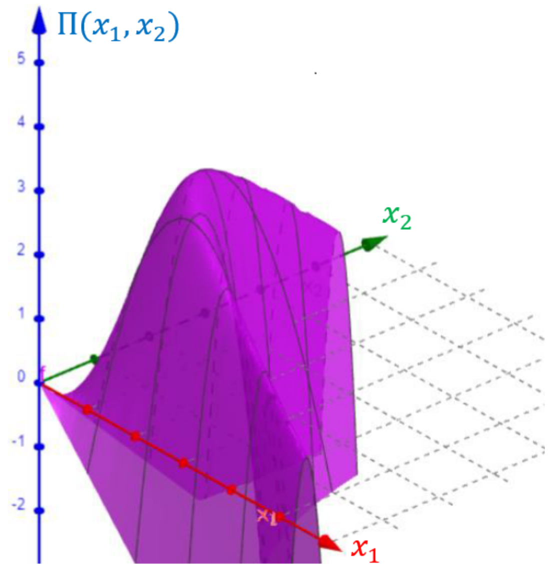


Figure 5: *The graph of the function (69).*

Example 5. *Let us assume that* $a = 5, b = 1, \alpha = \beta = w_1 = w_2 = 10$. *So* $\varepsilon = 20, \rho = 1$, *and the problem (3) becomes*

$$\max_{x_1,x_2\geq 0} p(q)q - 10\cdot x_1 - 10\cdot x_2 = (5 - x_1^{10}x_2^{10})x_1^{10}x_2^{10} - 10x_1 - 10x_2. \tag{70}$$

Condition (68) is satisfied. The solution to the problem (70) is the local maximum. Its value is approximately -14.6397, and it is achieved at point $(x_1, x_2) \approx (1.04185, 1.04185)$. *It can be concluded that the objective function of (68) is negative on the whole area (the exception is its value 0) and under these conditions the monopolist would operate at a loss.*

3. Conclusion

This paper examines the long-run profit maximization of a monopolist operating under linear demand and a two-input Cobb-Douglas production function. When the problem has a unique solution, the closed–form expression defines the profit function for the given demand function. It is shown that the closed–form solution to the problem is conditional on economies of scale in a linear demand setting. This generalization of the monopolist's long-run profit maximization model with linear demand and Cobb-Douglas technology provides a comprehensive economic and mathematical analysis enriched by economic interpretations. The conditions under which the problem yields a closed-form solution are identified, along with the market environment that enables the monopolist to sustain production in the long-run. Due to implications of duality results in production theory under certain regularity conditions, the monopolist's technology in the model is represented in two equivalent ways, by the Cobb-Douglas production function and the corresponding derived cost function. It is shown that both approaches reduce the discussion to the solvability of the equation equating marginal cost and marginal revenue, the well-known necessary condition for profit maximization. Furthermore, the economic analysis and subsequent economic interpretation offer interesting insights into the relationship between returns to scale and the monopolist's profit. The combination of mathematical analysis and economic interpretation enabled a unique theoretical analysis of the extremely important economic problem of the monopolist's long-run profit maximization under varying returns to scale. The practical implications of this work lie in defining and elaborating the conditions under which a solution to the specified long-run profit maximization problem of a monopolist exists. In addition, conditions are identified under which a closed-form solution is obtainable, enabling a comparative static analysis. Future research could focus on conducting sensitivity analysis and further comparative statics.

References

[1] Alekseev, V. B. (2004). *Abel's Theorem in Problems and Solutions*. New York: Springer. doi: 10.1007/1-4020-2187-9

[2] Auriol, E., Crampes, C. and Estache, A. (2021). *Regulating Public Services: Bridging the Gap between Theory and Practice*. Cambridge: Cambridge University Press. doi: 10.1017/9781108987660

[3] Baumol, W. J., Panzar, J. C. and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.

[4] Belleflamme, P. and Peitz, M. (2015). *Industrial organization: markets and strategies*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107707139

[5] Belleflamme, P. and Peitz, M. (2021). *The Economics of Platforms*. Cambridge: Cambridge University Press. doi: 10.1017/9781108696913

[6] Binger, B. R. and Hoffman, E. (1998). *Microeconomics with Calculus*. Harlow: Adison - Wesly.

[7] Decker, C. (2023). *Modern Economic Regulation: An Introduction to Theory and Practice*. Cambridge: Cambridge University Press.

[8] Diewert, W. E. (1982). Duality Approaches to Microeconomic Theory. In Intriligator, M. D. and Arrow, K. J. (Eds.) Handbook of Mathematical Economics (535–599). North-Holland.

[9] Ducci, F. (2020). *Natural Monopolies in Digital Platform Markets*. Cambridge: Cambridge University Press. doi: 10.1017/9781108867528

[10] Eatwell, J. (2008). Returns to Scale. In Durlauf, S. N. and Blume, L. E. (Eds.) The New Palgrave Dictionary of Economics (1–3). Palgrave Macmillan. doi: 10.1057/978-1-349-95121-5_1480-2

[11] Hicks, J. R. (1935). Annual Survey on Economic Theory: The Theory of Monopoly. Econometrica, 3(1), 1–20. doi: 10.2307/1907343

[12] Jehle, G. A. and Reny, P. J. (2011). *Advanced Microeconomic Theory*. Prentice Hall.

[13] Mas-Colell, A. and Whinston, M. D., Green, J. R. (1995). *Microeconomic Theory*. New York: Oxford University Press.

[14] Shepard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton: Princeton University Press.

[15] Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge: The MIT Press.

[16] Varian, H. R. (1992). *Microeconomic Analysis*. New York: Norton.

[17] Vrankić, I., Krpan, M. and Herceg, T. (2017). Monopolist's mark-up and the elasticity of substitution. Operational Research Review, 8(2), 377–390. doi: 10.17535/crorr.2017.0024

Numerical methods for checking the stability of gyroscopic systems

Ivana Kuzmanović Ivičić[1]

[1] *School of Applied Mathematics and Informatics, J. J. Strossmayer University of Osijek, Trg Ljudevita Gaja 6, Osijek, Croatia*
E-mail: ⟨*ikuzmano@mathos.hr*⟩

Abstract. Gyroscopic mechanical systems are modeled by the second-order differential equation

$$M\ddot{x}(t) + G\dot{x}(t) + Kx(t) = 0,$$

where $M \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix, $G \in \mathbb{R}^{n\times n}$ is a skew-symmetric ($G^T = -G$) matrix, and $K \in \mathbb{R}^{n\times n}$ is a symmetric matrix, representing the mass, gyroscopic, and stiffness matrices, respectively. The stability of such systems, which is the primary topic of this paper, is determined by the properties of the associated quadratic eigenvalue problem (QEP)

$$\mathcal{G}(\lambda)x = (\lambda^2 M + \lambda G + K)x = 0, \quad x \in \mathbb{C}^n,\ x \neq 0.$$

In this paper, we provide an overview of various linearizations of the QEP and propose numerical methods for checking the stability of gyroscopic systems based on solving the linearized problem. We present examples that demonstrate how the use of numerical methods provides a significantly larger stability region, which cannot be detected using the considered non-spectral criteria, or verify stability in cases where non-spectral criteria are not applicable, highlighting the advantages of numerical methods.

Keywords: generalized eigenvalues, gyroscopic systems, linearization, quadratic eigenvalue problem, stability

Received: October 2, 2024; accepted: January 8, 2025; available online: February 5, 2025

DOI: 10.17535/crorr.2025.0011

Original scientific paper.

1. Introduction

Gyroscopic mechanical systems originate from the study of rotating bodies and their dynamic behaviour. They play a crucial role in various disciplines, such as physics, engineering, and applied mathematics.

The dynamics of gyroscopic systems oscillating about an equilibrium position, with no external forces applied, is mathematically modelled by the equation

$$M\ddot{x}(t) + G\dot{x}(t) + Kx(t) = 0, \tag{1}$$

where the mass matrix $M \in \mathbb{R}^{n\times n}$ is symmetric and positive definite, the gyroscopic matrix $G \in \mathbb{R}^{n\times n}$, which represents the effect of gyroscopic internal forces, is skew-symmetric ($G^T = -G$), and the stiffness matrix $K \in \mathbb{R}^{n\times n}$ is symmetric. Here, $x = x(t)$ represents the time-dependent displacement vector.

The solutions of the equation (1) can be expressed in terms of the eigenvalues and eigenvectors of the quadratic matrix polynomial

$$\mathcal{G}(\lambda) = \lambda^2 M + \lambda G + K, \tag{2}$$

so the behaviour of the gyroscopic system (1) is fully characterized by the algebraic properties of the quadratic matrix polynomial (2) and the associated quadratic eigenvalue problem (QEP)

$$\mathcal{G}(\lambda)x = (\lambda^2 M + \lambda G + K)x = 0, \quad x \in \mathbb{C}^n, \quad x \neq 0. \tag{3}$$

One of the most important properties of mechanical systems, which is also the main topic of this paper, is stability. For gyroscopic systems, stability is achieved when the norm $\|x(t)\|$ of the solution of equation (1) is uniformly bounded for all $t > 0$ (see, for example, [1]). This condition is satisfied if and only if all eigenvalues of the quadratic matrix polynomial $\mathcal{G}$ are purely imaginary and semisimple, i.e., have equal geometric and algebraic multiplicities (see [1, 3]). It is well known that for all positive definite K (which physically corresponds to a system oscillating about a stable equilibrium point), the system (1) is stable. For negative definite K (corresponding to oscillation about an unstable equilibrium point) or indefinite K, there are no simple necessary and sufficient conditions for stability. Although there are some analytical 'non-spectral' stability criteria, they are often too restrictive and cannot be generalized to different types of system matrices. Rather than relying solely on analytical conditions, this paper explores numerical methods for checking the stability. This computational approach overcomes the limitations of theoretical results, offering a more comprehensive solution to the stability problem in gyroscopic mechanical systems (see [5, 6, 12, 18]).

More precisely, in this paper, we provide an overview of various linearizations of the QEP (3) and numerical methods for checking the stability of gyroscopic systems, which are based on solving the linearized problem. These methods do not require any prior information about the interconnections between the system matrices or the definiteness of the matrix K, which is usually needed for non-spectral criteria. We present an example often used to compare analytical criteria for determining the stability region of parameter-dependent problems and demonstrate that numerical methods provide a significantly larger stability region, which cannot be detected using the considered analytical criteria. Furthermore, we present examples where analytical criteria are difficult or impossible to apply while numerical methods provide relevant results.

System stability is often a fundamental requirement for analyzing other system properties. For example, in our previous work, we focused on perturbation analysis for stable gyroscopic systems, assuming stability in advance. In paper [14], we studied perturbation bounds for gyroscopic systems in motion about a stable equilibrium position. In particular, we studied perturbations in the eigenvalues and eigenvectors of stable gyroscopic systems with symmetric and positive definite stiffness matrix K. In [15], we analyzed gyroscopic systems in motion about an unstable equilibrium position. These systems are modelled as linear gyroscopic mechanical systems of the form (1), where the stiffness matrix K is symmetric and negative definite. In both papers, we derived new relative bounds for the perturbation of the eigenvalues and bounds of the $\sin\Theta$ - type for the associated eigenspaces based on results from [20].

The rest of the paper has the following structure: Section 2 provides a brief introduction to the stability of gyroscopic systems and presents related results used later in the paper. Section 3 examines the stability of gyroscopic systems in terms of the generalized eigenvalues of matrix pencils obtained through several different types of linearizations. Section 4 presents numerical methods for calculating system eigenvalues and their multiplicities. Section 5 includes numerical examples, and Section 6 provides an overview of the results.

Notation. Throughout this paper, I_n, $n \in \mathbb{N}$, denotes the $n \times n$ identity matrix. $\lambda(A)$ denotes the set of all eigenvalues of a matrix A. $A > 0$ ($A < 0$) denotes that the matrix A is positive (negative) definite, while $A \geq 0$ ($A \leq 0$) denotes that A is positive (negative) semidefinite.

2. Stability of gyroscopic systems

One of the most significant properties of the spectrum of the quadratic matrix polynomial $\mathcal{G}$ from (2) is its Hamiltonian symmetry, which means that it is symmetric with respect to both the real and imaginary axes (since M is nonsingular, the spectrum of (2) is always a bounded set in the complex plane, meaning that (2) has $2n$ finite eigenvalues). Specifically, if λ is an eigenvalue of $\mathcal{G}(\lambda)$, then $\overline{\lambda}$, $-\lambda$, and $-\overline{\lambda}$ are also eigenvalues, with associated pairs of left and right eigenvectors (y, x), $(\overline{y}, \overline{x})$, (x, y) and $(\overline{x}, \overline{y})$, respectively. Because of this, gyroscopic systems can only have the property of marginal stability, meaning they are considered stable if all eigenvalues of (2) are purely imaginary and semisimple. Therefore, a direct approach to check the stability of the system involves calculating all the eigenvalues of the quadratic matrix polynomial and verifying their algebraic and geometric multiplicities.

There are some very useful 'non-spectral' results on the stability of gyroscopic systems, expressed in terms of the properties of system matrices or their interconnections. The most significant are the results expressed only in terms of the definiteness of the matrix K. As already mentioned, a system with a positive definite K is always stable. On the other hand, if K is positive semidefinite, only zero eigenvalues can be defective (that is, not semisimple), and $\mathcal{G}(\lambda)$ has a defective eigenvalue if and only if K is singular and there exist nonzero vectors in $G(\operatorname{Ker} K) \cap \operatorname{Im} K$ (see [16, Theorem 2]). Furthermore, for systems with an indefinite K, it is known that if $\det K < 0$, then (1) is unstable (see [16, Theorem 3],[17, Theorem 6.3]). Also, result from [4] show that the system is stable if $K < 0$ and

$$4K - G^2 - \delta I > 0,$$

where $\delta = 2(\delta_1 - \delta_2)$, and δ_1 and δ_2 are the maximal and minimal eigenvalues of $-K$, respectively. Therefore, if information about the definiteness of the matrix K is available, some of the previously mentioned criteria can serve as necessary or sufficient conditions for the stability or instability of the system.

There are also criteria based on other system matrices and their properties or interconnections. For example, it is well known that the system $\ddot{x} + G\dot{x} + K = 0$ is unstable if

$$4K - G^2 < 0. \tag{4}$$

Unfortunaly, the converse is not true in the general case. However, in the case where K and G commute,

$$4K - G^2 > 0 \tag{5}$$

is both a necessary and sufficient condition for stability (see [10]). On the other hand, the condition

$$K - GM^{-1}G < 0$$

always implies instability of system (1) (see [7]).

The main disadvantage of all these criteria is that they are limited to specific types of systems, and some are difficult to verify.

3. Stability and generalized eigenproblem

The main goal of this paper is to investigate numerical methods for checking the stability of gyroscopic systems of the form (1), including the calculation of the eigenvalues of (2). A common approach to solving the QEP is to first linearize it into a linear eigenvalue problem. One possible class of linearizations of $\mathcal{G}$ is given by

$$L(\lambda) = \begin{bmatrix} 0 & X \\ -K & -G \end{bmatrix} - \lambda \begin{bmatrix} X & 0 \\ 0 & M \end{bmatrix}, \tag{6}$$

where $X \in \mathbb{R}^{n\times n}$ is an arbitrary regular matrix. Although the result is independent of X, the choice of suitable X may provide advantages in numerical methods. Specifically, we can use linearization (6) with $X = I$, resulting in

$$L_M(\lambda) = \begin{bmatrix} 0 & I \\ -K & -G \end{bmatrix} - \lambda \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} := A - \lambda B\,. \tag{7}$$

The generalized eigenvalue problem of the matrix pencil L_M and the QEP (3) are equivalent. This means that eigenvalues of $\mathcal{G}(\lambda)$ and the pencil $A - \lambda B$, i.e. the pair (A, B) are the same and that the associated eigenvectors of one problem can be calculated from those of the other. Since M is regular, instead of the generalized eigenvalues of the pair (A, B), we can consider a standard eigenvalue problem $B^{-1}Ax = \lambda x$, which involves calculating the eigenvalues of the matrix

$$\begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}G \end{bmatrix}. \tag{8}$$

If M is very ill-conditioned, such that $M^{-1}K$ and $M^{-1}G$ cannot be computed accurately, the QEP should not be replaced with a standard eigenproblem and must instead be solved as a generalized eigenproblem. However, the structure of M often allows for an efficient calculation of its inverse. Thus, if M^{-1} is given, or, as is common in theoretical analyses, the mass matrix M is assumed to be the identity matrix (see, for example, the criterion discussed in [1]), the standard eigenvalue problem with the matrix (8) may provide a more advantageous formulation. The main drawback of the linearization (7) is that the resulting eigenvalue problem is non-symmetric.

Linearization can also be chosen such that A and B preserve the special structures of the system matrices. There are some linearizations $A - \lambda B$ that reflect the property that the eigenvalues of a real gyroscopic system occur in quadruples $(\lambda, \overline{\lambda}, -\lambda, -\overline{\lambda})$, where one of the matrices A or B is Hamiltonian, and the other is skew-Hamiltonian. The matrix A is Hamiltonian if

$$(AJ)^T = AJ,$$

where

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

If

$$(AJ)^T = -AJ,$$

we say that A is skew-Hamiltonian. An example of such a skew-Hamiltonian / Hamiltonian linearization is given by (see [19])

$$\begin{bmatrix} K & 0 \\ G & K \end{bmatrix} - \lambda \begin{bmatrix} 0 & K \\ -M & 0 \end{bmatrix}. \tag{9}$$

On the other hand, a Hamiltonian/skew-Hamiltonian linearization is

$$\begin{bmatrix} 0 & -K \\ M & 0 \end{bmatrix} - \lambda \begin{bmatrix} M & G \\ 0 & M \end{bmatrix}, \tag{10}$$

which is preferred when K is singular (see [19]).

There are also linearizations that result in a Hermitian eigenvalue problem. To obtain such a linearization, from the matrix polynomial $\mathcal{G}$ from (2), we define a new matrix polynomial

$$\mathcal{H}(\lambda) = -\mathcal{G}(-i\lambda) = \lambda^2 M + \lambda H - K \tag{11}$$

where i is the imaginary unit, and

$$H := iG$$

is a complex Hermitian matrix. The benefit of this transformation lies in the fact that all matrices in (11) are Hermitian, and the eigenvalues of $\mathcal{G}$ and $\mathcal{H}$ can simply be computed from each other. More specifically, if λ is an eigenvalue of $\mathcal{G}$, then $i\lambda$ is an eigenvalue of $\mathcal{H}$, and conversely, both sharing the same eigenvector. As stated previously, the gyroscopic system (1) is stable if and only if all the eigenvalues of $\mathcal{G}$ are purely imaginary and semisimple. This equivalently means that all eigenvalues of $\mathcal{H}$ are real and semisimple. In the case where $K > 0$, the QEP associated with (11) is a hyperbolic eigenvalue problem, which is known to have real and semisimple eigenvalues. This is consistent with the statement that gyroscopic systems with $K > 0$ are stable (see [19]).

Due to the desirable properties of $\mathcal{H}(\lambda)$, instead of the original QEP (3), we can consider the QEP

$$\mathcal{H}(\lambda)x = 0, \quad x \in \mathbb{C}^n, \quad x \neq 0. \tag{12}$$

One possible Hermitian linearization of $\mathcal{H}$ is

$$L_{\mathcal{H}}(\lambda) = \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} - \lambda \begin{bmatrix} H & M \\ M & 0 \end{bmatrix} := A - \lambda B\,. \tag{13}$$

The generalized eigenvalues of the original problem can be obtained from the eigenvalues of $A - \lambda B$ by multiplying them by $-i$. Since $A, B \in \mathbb{C}^{n\times n}$ are Hermitian matrices, the matrix pair (A, B) is said to be Hermitian. Note that the linearizations discussed above involved real matrices, while in this case, B is a complex matrix.

If $K < 0$, the linearization (13) can be further simplified. By transforming (13), one of the matrices in the pair (A, B) can be reduced to a sign matrix J. Let L_M and L_C be the Cholesky factors of M and $C = -K$ (C is positive definite), such that $M = L_M L_M^*$ and $K = -C = -L_C L_C^*$. Instead of the pair (A, B), we can consider the equivalent pair

$$(L^{-1}AL^{-*}, L^{-1}BL^{-*}) := (J, B_L),$$

where

$$L^{-1} = \begin{bmatrix} L_C^{-1} & 0 \\ 0 & L_M^{-1} \end{bmatrix}$$

and

$$J = \begin{bmatrix} -I_n & 0 \\ 0 & I_n \end{bmatrix}, \quad B_L = \begin{bmatrix} L_C^{-1}HL_C^{-*} & L_C^{-1}L_M \\ L_M^* L_C^{-*} & 0 \end{bmatrix}. \tag{14}$$

This specific linearization was used in [15] to obtain relative perturbation bounds.

3.1. Stability and definite matrix pairs

A Hermitian matrix pair (A, B) is said to be definite if there exists a real linear combination $\alpha A + \beta B$ of the matrices A and B that is positive definite. If either A or B is a definite matrix, the pair (A, B) is trivially definite. Definite matrix pairs have real and semisimple eigenvalues. Therefore, if (A, B) from (13) is definite, the corresponding gyroscopic system is stable (see [19]).

Using the ideas from [1, 8] applied to the QEP (12), we obtain that for the definiteness of the pair (A, B) from (13), we have

$$\alpha A + \beta B = \begin{bmatrix} \alpha K + \beta H & \beta M \\ \beta M & \alpha M \end{bmatrix} = \tag{15}$$

$$= \begin{bmatrix} I & -\frac{\beta}{\alpha} I \\ 0 & I \end{bmatrix} \begin{bmatrix} \alpha K - \beta H - \frac{\beta^2}{\alpha} M & 0 \\ 0 & \alpha M \end{bmatrix} \begin{bmatrix} I & 0 \\ -\frac{\beta}{\alpha} & I \end{bmatrix}, \tag{16}$$

which implies that $\alpha A + \beta B$ is congruent to

$$\alpha \begin{bmatrix} -\mu^2 M - \mu H + K & 0 \\ 0 & M \end{bmatrix},$$

where $\mu = -\frac{\beta}{\alpha}$. Since M is positive definite, the pair (A, B) is definite if and only if

$$-\mu^2 M - \mu H + K$$

is positive definite for some μ. Equivalently, this condition holds if and only if

$$\mu^2 M + \mu H - K$$

is negative definite for some μ. Therefore, if we have information about the negative definiteness of this matrix, we can conclude that the system is stable. Notice that when K is positive definite, then A is also positive definite, and the pair (A, B) is trivially definite. This corresponds to the well-known fact that gyroscopic systems with positive definite K are always stable. However, while the positive definiteness of the pair (A, B) implies that the system is stable, the converse is not true. A system can still be stable even if (A, B) is not definite.

4. Numerical methods for detecting stability

In this section, we discuss numerical methods for detecting stability in gyroscopic systems. These methods involve the computation of eigenvalues of the pair (A, B), which is derived from the gyroscopic system through the linearization of the corresponding QEP. One approach to obtaining the eigenvalues of (A, B) is by reducing the pair to its generalized Schur form

$$Q^* A Z = S, \quad Q^* B Z = T, \tag{17}$$

where $Q, Z \in \mathbb{C}^{2n \times 2n}$ are unitary matrices, and $S, T \in \mathbb{C}^{2n \times 2n}$ are upper triangular. The generalized eigenvalues are then given by

$$\frac{s_{ii}}{t_{ii}}, i = 1, \ldots, 2n,$$

where t_{ii} and s_{ii} are the diagonal elements of T and S, respectively. For general matrix pairs, it is possible that $t_{ii} = 0$ for some i, corresponding to an infinite eigenvalue. However, gyroscopic systems with $M > 0$ have only finite eigenvalues. If a real linearization (A, B) is chosen, then a real generalized Schur form exists, where Q and Z are orthogonal, and T and S are upper quasi-triangular. The method described above is the QZ algorithm, a numerically stable method for determining generalized eigenvalues (see [13]).

The QZ algorithm does not take advantage of the structure of the system matrices or the special spectral properties of gyroscopic systems. If (A, B) is chosen to be one of the skew-Hamiltonian/Hamiltonian or Hamiltonian/skew-Hamiltonian linearizations given in (9) and (10), the problem can be reformulated as a Hamiltonian eigenvalue problem. Such problems can be solved using specialized algorithms, such as the one described in [2].

A concise overview of numerical methods for the generalized eigenvalue problem, along with their properties, is provided in [19].

Another numerical challenge in stability analysis is determining whether the eigenvalues are semisimple. Simple eigenvalues are always semisimple, but for multiple eigenvalues (those with an algebraic multiplicity greater than 1), it is necessary to verify their geometric multiplicity. For a generalized eigenvalue λ of the matrix pair (A, B), the geometric multiplicity is defined as the dimension of the null space of $A - \lambda B$. To compute the geometric multiplicity, we can calculate the rank of $A - \lambda B$. This method is particularly effective for small to medium-sized matrices and when the eigenvalues are well-separated.

5. Numerical examples

In this section, we demonstrate the advantages of using a numerical approach for stability analysis of gyroscopic systems compared to algebraic necessary/sufficient conditions. In the following example, we demonstrate how the numerical approach can provide a significantly larger stability region in a parameter-dependent problem.

Example 1. *Let's consider a widely used example of a gyroscopic system commonly found in literature (e.g., see [4, 9, 11]). This example represents a simplified model of a disk mounted on a weightless, non-circular rotating shaft subjected to a constant axial compression force. The system can be described by the following differential equation:*

$$M\ddot{x}(t) + G\dot{x}(t) + Kx(t) = 0,$$

where

$$M = I, \quad G = 2\omega \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad K = \begin{bmatrix} k_1 - \eta - \omega^2 & 0 \\ 0 & k_2 - \eta - \omega^2 \end{bmatrix}. \tag{18}$$

The parameter ω represents the shaft's angular velocity, and its value in this example is $\omega = 2$. Furthermore, η represents the axial force, with a value of $\eta = 3$. The parameters $0 < k_1 < 7$ and $0 < k_2 < 7$ denote the elastic rigidities in the two principal directions.

The paper [4] analyzes what stability area various analytical criteria give for this example. In that paper, the best result gives Bulatovic's criterion, which is applicable only to gyroscopic systems with negative definite K:

Theorem 5.1. *(see [4]) If $G^T = -G$ and $K^T = K < 0$, the system $\ddot{x} + G\dot{x} + Kx = 0$ is stable if*

$$4K - G^2 - \delta I > 0 \tag{19}$$

where $\delta = 2(\delta_1 - \delta_2)$, where δ_1 and δ_2 are the maximal and the minimal eigenvalues of $-K$, respectively.

Previous criterion gives that the system is stable if (see [4])

$$2(k_1 - 3) - |k_1 - k_2| > 0 \quad \text{and} \quad 2(k_2 - 3) - |k_1 - k_2| > 0,$$

which reduces to

$$k_1 > 3 \quad \text{and} \quad (6 + k_1)/3 < k_2 < -3 + 2k_1,$$

and stability area is illustrated in Figure 1(a).

On the other hand, the following Barkwell and Lancaster (see [1]) stability criterion gives better results for this example.

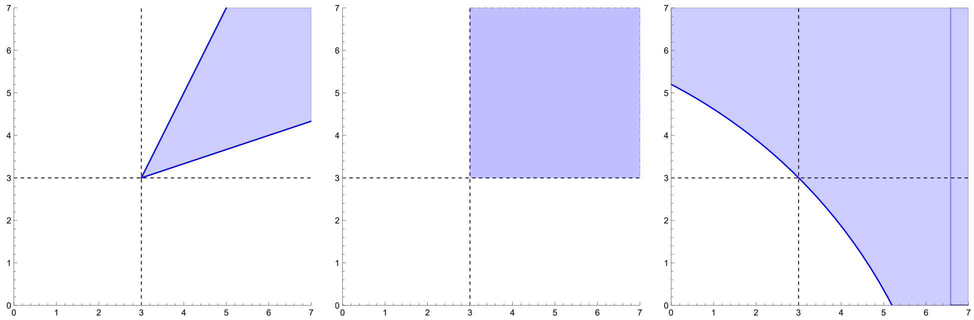
Theorem 5.2. *[1, Theorem 5] If G is a real skew-symmetric nonsingular matrix, and K is a real negative definite matrix, then the matrix polynomial $\mathcal{H}(\lambda) = \lambda^2 I + \lambda iG - K$ has all real and semisimple eigenvalues whenever*

$$|G| \geq kI - k^{-1}K \tag{20}$$

holds for some $k > 0$, where $|G|$ denotes the positive square root of $-G^2$.

For this particular example, criterion '$|G| > kI - k^{-1}K$, holds for some $k > 0$' means that

$$k_1, k_2 > k^2 - 4k + 7 \text{ for some } k > 0. \tag{21}$$



(a) (b) (c)

Figure 1: (a) Stability area from [4] (b) Stability area from [1] (c) Numerical stability area

Since

$$\min_{k>0}(k^2 - 4k + 7) = 3,$$

this stability criterion gives a wider stability area, so criterion (20) *is better then* (19) *for the observed example (see Figure 1(b)). Using our direct approach, it is easy and very straightforward (it can also be calculated by hand because of such small system matrices) to see that some parts of the stability area are not included in the previous two results. An example of this was obtained when we calculated the eigenvalues of the system for* $k_1 = 2.3$ *and* $k_2 = 3.65$ *(which is not included in any stability region defined by previous results). The eigenvalues are* $\pm 2.052i$ *and* $\pm 1.9337i$*, which are purely imaginary and semisimple, implying stability. Similarly, the numerical method gives many more configurations that ensure the stability of the system, and the whole area is illustrated in Figure 1(c). We can see that the stability area is much wider than the analytical criteria can confirm.*

Example 2. *In this example, we examine a system similar to the one in the previous example, but with two discs. Specifically, we consider a simplified model consisting of two discs mounted on a non-circular, weightless, elastic shaft that rotates at a constant angular velocity* ω*. An example of the system matrices for this type of system is*

$$M = I, \quad G = 2\omega \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 2-\omega^2 & 1 & 0 & 0 \\ 1 & 4-\omega^2 & 0 & 0 \\ 0 & 0 & 2-\omega^2 & 1 \\ 0 & 0 & 1 & 4-\omega^2 \end{bmatrix}.$$

We observe the stability of this system with respect to $\omega \geq 0$*. This results in* K*, which can be positive definite (e.g., for* $\omega = 1$*), indefinite (e.g., for* $\omega = 2$*), or negative definite (e.g., for* $\omega = 3$*). This means that none of the analytical criteria that rely on the assumption of the definiteness of* K *can be applied to all three cases (such as those given in Theorems 5.1 and 5.2 from the previous example). If we notice that* K *and* G *commute, we can apply the analytical criterion* (5) *and verify the positive definiteness of* $4K - G^2$ *to conclude that the system is stable in all three cases. Indeed, numerical methods also confirm stability (the obtained eigenvalues are given in Table 1), but without requiring verification of any conditions on the system matrices, which gives them an advantage in practical applications.*

Example 3. *In this example, we analyze the stability of a gyroscopic system of dimension* $n = 10$*, where the mass matrix is a random diagonal matrix with diagonal elements in* $[0, 1]$*, the gyroscopic matrix is a full random antisymmetric matrix with elements from* $[-100, 100]$*, and the stiffness matrix is a random symmetric matrix with eigenvalues in the range* $[-100, 1]$

ω	eigenvalues
1	$\pm 3.1010\,i$, $\pm 2.2593\,i$, $\pm 0.2593\,i$, $\pm 1.1010\,i$
2	$\pm 4.1010\,i$, $\pm 3.2593\,i$, $\pm 0.7407\,i$, $\pm 0.1010\,i$
3	$\pm 5.1010\,i$, $\pm 4.2593\,i$, $\pm 1.7407\,i$, $\pm 0.8990\,i$

Table 1: Eigenvalues of QEP from Example 2 for $\omega \in \{1, 2, 3\}$

(hence, K can be negative (semi)definite or indefinite). We repeated the experiment 100 times. In six instances, the generated system had an indefinite K, and all corresponding systems were unstable. In the remaining 94 *cases, the generated system had a negative definite K and was stable in* 19 *instances. This analysis does not rely on any prior assumptions about the interconnections between the system matrices, and the most effective approach to analyze it is through numerical procedures.*

6. Conclusions

In this paper, the stability problem of gyroscopic systems is considered. Stability is related to the eigenvalues of the associated quadratic problem and their multiplicities. Many papers provide analytical conditions for stability verification, which are simpler than computing generalized eigenvalues, but their application is very restrictive, so they often do not give the results. On the other hand, numerical methods always provide results, and when stable algorithms are used, obtained results are reliable. In the examples in Chapter 5, it can be seen how the numerical criterion provides a significant advantage compared to some algebraic methods in checking the stability or in determining the stability region for a specific parameter-dependent problems. These results indicate that numerical methods should be used in practice, especially when analytical criteria are not applicable.

References

[1] Barkwell, L., and Lancaster, P. (1992). Overdamped and gyroscopic vibrating systems. Journal of Applied Mechanics 59(1), 176–181. doi: 10.1115/1.2899425

[2] Benner P., Mehrmann, V. and Xu, H. (1998). A numerically stable, structure preserving method for computing the eigenvalues of real Hamiltonian or symplectic pencils. Numerische Mathematik 78(3), 329–358. doi: 10.1007/s002110050315

[3] Bulatovic, R. (1997). On the Lyapunov stability of linear conservative gyroscopic systems. Comptes Rendus de l'Académie des Sciences- Series IIB Mechanics-Physics-Chemistry-Astronomy 324, 679–683. doi: 10.1016/S1251-8069(97)83173-9

[4] Bulatovic, R. (1998). A nonspectral stability criterion for linear conservative gyroscopic systems. Journal of Applied Mechanics 65(2), 539–541. doi: 10.1115/1.2789091

[5] Gladwell, G., Khonsari, M. and Ram, Yosef. (2003). Stability Boundaries of a Conservative Gyroscopic System. Journal of Applied Mechanics 70(4), 561-567. doi: 10.1115/1.1574062

[6] Guo C-H. (2004) Numerical solution of a quadratic eigenvalue problem, Linear Algebra and its Applications 385, 391-406. doi: 10.1016/j.laa.2003.12.010

[7] Hagedorn, P. (1975) Über die instabilität konservativer systeme mit gyroskopischen kräften. Archive for Rational Mechanics and Analysis 58, 1–9. doi: 10.1007/BF00280151

[8] Higham, N. J., Tisseur, F. and Van Dooren, P. M. (2002). Detecting a definite Hermitian pair and a hyperbolic and elliptic quadratic eigenvalue problem, and associated nearness problem. Linear Algebra and its Applications 351–352, 455–474. doi: 10.1016/S0024-3795(02)00281-1

[9] Huseyin, K. (1976). Standard Forms of the Eigenvalue Problems Associated with Gyroscopic Systems. Journal of Sound and Vibration 45(1), 29–37. doi: 10.1016/0022-460X(76)90665-9

[10] Huseyin, K., Hagedorn, P. and Teschner, W. (1983) On the stability of linear conservative gyroscopic systems. Zeitschrift für Angewandte Mathematik und Physik (ZAMP) 34, 807–815. doi: 10.1007/BF00949057

[11] Inman, D. J. (1988). A sufficient condition for the stability of conservative gyroscopic systems. Journal of Applied Mechanics 55(4), 895–898. doi: 10.1115/1.3173738

[12] Jia, Z. and Wei, M. (2011). A Real-Valued Spectral Decomposition of the Undamped Gyroscopic System with Applications. SIAM Journal on Matrix Analysis and Applications 32(2), 584-604. doi: 10.1137/100792020

[13] Kressner, D. (2005). Numerical Methods for General and Structured Eigenvalue Problems. Springer Berlin, Heidelberg. doi: 10.1007/3-540-28502-4

[14] Kuzmanović Ivičić, I. and Miodragović, S. (2023). Perturbation bounds for stable gyroscopic systems, BIT Numerical Mathematics 63(1). doi: 10.1007/s10543-023-00943-5

[15] Kuzmanović Ivičić, I. and Miodragović, S. (2024) Perturbation bounds for stable gyroscopic systems in motion about an unstable equilibrium position, Journal of Computational and Applied Mathematics 451, doi: 10.1016/j.cam.2024.116061

[16] Lancaster, P. (2013). Stability of linear gyroscopic systems: A review. Linear Algebra and its Applications 439(3), 686–706. doi: 10.1016/j.laa.2012.12.026

[17] Merkin, D. R. (1997). Introduction to the Theory of Stability. Springer, New York. doi: 10.1007/978-1-4612-4046-4

[18] Sui, Yf., Zhong, Wx. (2006). Eigenvalue problem of a large scale indefinite gyroscopic dynamic system. Appl Math Mech 27, 15–22. doi: 10.1007/s10483-006-0103-z

[19] Tisseur, F. and Meerbergen, K. (2001). The quadratic eigenvalue problem. SIAM review 43(2), 235–286. doi: 10.1137/S0036144500381988

[20] Truhar N. and Miodragović S. (2015). Relative perturbation theory for definite matrix pairs and hyperbolic eigenvalue problem. Applied Numerical Mathematics 98, 106-121. doi: 10.1016/j.apnum.2015.08.006

CRORR **16:2**(2025), 137–150

An application of an environmentally extended input-output model to the assessment of CO2 emissions of Croatian economic sectors

Davor Mikulić[1], **Katerina Fotova Čiković**[2] and **Damira Keček**[2,*]

[1] *The Institute of Economics Zagreb, Trg J. F. Kennedyja 7, 10000 Zagreb, Croatia*
E-mail: ⟨dmikulic@eizg.hr⟩

[2] *University North, Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia*
E-mail: ⟨{kcikovic, dkecek}@unin.hr⟩

Abstract. The adverse effects of pollution and the impacts of different productive sectors on the degradation of the environment have become important issues in economic development analysis. This study aims to estimate the overall carbon footprint of different economic sectors of the Croatian economy. Total CO2 emissions are decomposed by direct and indirect emissions of productive sectors and final demand by applying the environmentally extended input-output analysis (EEIO). EEIO approach enables the decomposition of the total CO2 emissions of specific sectors into emissions related to the production of final goods and services and the production of intermediate inputs delivered to other economic sectors. The highest level of direct emissions has been generated by sectors producing energy products and transportation. As the output of those sectors is required as intermediate inputs used by the other sectors, the EEIO model redistributes emissions to other sectors, delivering final products to domestic users and abroad. Besides transport and electricity, Croatia's highest level of total direct and indirect emissions is found in the construction sector, production of non-metallic mineral products and food industry.

Keywords: CO2 emission, environmentally extended input-output model, Republic of Croatia, total effects

Received: October 7, 2024; accepted: December 16, 2024; available online: February 5, 2025

DOI: 10.17535/crorr.2025.0012

Original scientific paper.

1. Introduction

In the ever-increasing changes and dynamics of changes in all spheres of peoples' lives, the topical issue of sustainability and sustainable growth is inevitable. Moreover, considering how climate change and environmental degradation impact the world in a myriad of ways, it is apparent that policymakers, government bodies, entrepreneurs and corporations are highly interested in ways to identify what are the productive sectors that cause the most damage and are the least sustainable. The environmentally extended input-output analysis (hereafter EEIO) represents a potent technique used to evaluate the influence of certain economic activities on environmental degradation, depletion of natural resources and emission of pollutants. Moreover, EEIO provides "a simple and robust method for evaluating the linkages between economic consumption activities and environmental impacts, including the harvest and degradation of natural resources" that assesses the relationship between economic activities and sectors and downstream environmental and ecological impacts [16]. The underlying concept of EEIO methodology is

*Corresponding author.

to develop and replace previously used production and diagram trees with an input-output matrix that shows flows, usually monetary, but sometimes physical, between economic sectors [16], therefore developing a quantitative framework for the identification of sustainable and unsustainable national industries, in terms of direct and indirect environmental and ecological impact [9].

The main objective of this study is to estimate the overall carbon footprint of productive sectors of the Croatian economy and to distribute the total CO2 emissions to individual components of final demand. The EEIO approach regarding the Croatian economy has been applied to study wastewater discharges, revealing that industrial sectors generate significantly more wastewater than in comparable European countries [7]. The EXIOBASE 3 database provides a comprehensive time series of EE MRIO tables for global economic and environmental analysis, including data for Croatia [31]. Input-output analysis has also been used to assess the economic impact of the transportation sector in Croatia, showing significant multiplicative effects, particularly in air transport [34]. These studies demonstrate the versatility of EEIO methods in examining various aspects of the Croatian economy, from environmental impacts to sectoral interdependencies, providing valuable insights for policymakers and researchers. However, to the best of the authors' knowledge, this article is the first to quantify Croatian productive sectors' direct and indirect CO2 emissions using the EEIO approach.

The structure of the paper takes the form of five parts. Section 2 provides a literature review of the EEIO method and greenhouse gas emissions (GHG) emissions. Section 3 unveils the methodology. The research results are presented in Section 4, and a discussion and conclusion are provided in the last section.

2. Literature review

A literature review is focused on previous studies that use the EEIO model and are relevant to the objective of this study. Thus, it includes the literature developing EEIO methodology, empirical studies for other economies that identify industries related to the most severe environmental degradation, and literature that provides policy recommendations for efficiently reducing CO2 emissions proposed in the previous empirical studies for other economies. Findings from previous studies could be helpful in formulating the optimal model for assessing sources of CO2 emissions in Croatia, identifying similarities and differences in the economic structure, and formulating adequate policy recommendations based on specific national conditions.

Modifications to standard IO tables with ecological indicators were first introduced in [3] and [18], where authors incorporated pollution variables into the standard input-output (IO) model. The main idea behind the EEIO model is to augment the standard technical coefficients matrix with additional rows reflecting the pollutants' emissions [24]. Thus, by application of the assumption of the standard IO model, where final demand through the technological connection of economic sectors induces economic effects not only for units producing final products but along the overall supply chain, direct and indirect emissions of pollutants can also be calculated [24]. Beside analyse of GHG, the same methodological background can be used in many different applications and environmental extensions to the EEIO method, such as studies that incorporate water pollution [5, 21], carbon emissions [8, 35], waste [13] and food waste [1, 22].

Empirical studies are conducted for many economies to identify the sources of CO2 emissions based on the EEIO model. The findings of recent empirical studies are summarized in Table 1. In most empirical studies, construction, agriculture, electricity produced from fossil fuels, the metal industry, transport, and the production of construction materials are sectors with the most significant environmental footprints. At the same time, clean-electric power generation, public services, education, health and social work are associated with the lowest emissions. The negative impact of economic activity on the environment is not limited to traditional sectors.

Moon, Yun and Lee (2020) [25] concluded that the sectors with the most CO2 emissions in the Korean economy produce goods and services of the fourth industrial revolution.

Author(s) of empirical study	Country/region	Sectors with the highest CO2 emissions
Liu et al. (2020) [19]	Province Saskatchewan in Canada	Construction; crop and animal production and forestry and logging; electricity from fossil fuels
Shmelev and Brook (2021) [30]	Austria, China, France, Germany, Sweden and the USA	Production of electricity from gas, coal, oil, and other petroleum products; extraction of metals and other ores; production of cement, lime, gypsum, and air transport
Alcantara and Padilla (2006) [2]	Spain	Electricity and gas: land transport; manufacture of basic metals; manufacture of non-metallic mineral products, and manufacture of chemicals
He et al. (2021) [14]	China	Electricity/heat production; Construction
Imansyah and Putranti (2017) [15]	Indonesia	Crude oil and natural gas; commerce; thermal power; coal; non-metallic ores mining; petroleum refinery products and road transport
Othman and Jafari (2016) [26]	Malaysia	Energy industry and transportation
Moon, Yun and Lee (2020) [25]	South Korea	Manufacture of computers, electronic and optical products; chemicals and pharmaceutical products; motor vehicles, trailers and semi-trailers and construction
Beidari, Lin and Lewis (2017) [4]	South Africa	Electricity, other industries; transportation, and residential sectors

Table 1: *The results of recent empirical EEIO studies on the sectors with the highest CO2 emissions (Source: Authors' systematization)*

EEIO analysis on the sources of environmental degradation enables governments and regional leaders to assess productive sectors that harm the environment and to consider this information when developing new regulations, policies and laws. Empirical results on the sources of CO2 pollution are essential for formulating guidelines towards sustainable development and achieving the Paris Climate Agreement goals. Alcantara and Padilla (2006) [2] suggest that reducing emissions would be made possible by introducing specific measures and incentives such as increased use of renewable energy sources, innovations in the production of vehicles, and improvement of the ecological quality of fuel. He et al. (2021) [14] concluded that emission reduction policies should promote the decarbonisation of the electricity industry and improve energy efficiency by controlling carbon emissions in other productive sectors. Moon, Yun and Lee (2020) [25] pointed out that it is necessary to urgently reorganise production in modern sectors in a more environmentally friendly way without limiting their production and the possibility of losing competition. Beidari, Lin and Lewis (2017) [4] found that direct CO2 emissions are more significant than indirect CO2 emissions and proposed that the government should invest in expanding and maintaining national energy capacities and reduce coal and oil imports.

Some authors tried to answer the sustainability question: Is economic growth necessarily related to increased emissions of pollutants? Firdaus (2020) [12] compares the Indonesian economy's output and CO2 emission multipliers. The author pointed out that electricity and gas supply are sectors with a high output multiplier but also a relatively high CO2 emission multiplier, which means that increased environmental destruction may accompany economic growth. On the other hand, the rubber and plastic products sector has a high output multiplier and a very low CO2 multiplier, so it would be desirable to increase that sector's production to reduce CO2 emissions.

3. Research methodology

This research estimates direct and indirect CO2 emissions of economic sectors and final demand based on the EEIO methodology. In general, IO analysis makes it possible to determine the interdependence of productive sectors in the national economy [17, 24, 33]. IO tables describe the flow of goods and services among productive sectors. Readers can learn more about the standard IO approach [24, 33].

The standard IO table can be expanded to show the impact of individual activities on the environment, that is, the role of each economic sector in generating harmful substances that pollute the environment [24, 27, 28, 32]. A fundamental problem in models that include effects on the environment is the application of an appropriate unit of measure for the emission of harmful substances [23, 24].

In general form, the EEIO table can be presented by Table 2. The upper part is the standard IO table presenting the structure of deliveries of economic sectors (by rows) and the use of intermediate inputs and components of final demand (by columns). Intersectoral deliveries of products produced by sector i, which are used as intermediate inputs of sector j are presented by $X_{ij}, (i, j = 1, \ldots, n)$. The total output of each sector $(X_i, i = 1, \ldots, n)$, besides intermediate inputs, includes goods and services used by final consumers $(Y_i, i = 1, \ldots, n)$. Final uses can be further disaggregated into main components: personal consumption, government consumption, gross capital formation and exports, which are here, because of clarity and simplicity, not presented in Table 2.

Columns of IO table present the structure of intermediate inputs produced domestically (X_{ij}) and imported intermediate inputs (M_i). In the fundamental IO model, the matrix A, square $n \times n$ matrix, expresses a set of technical coefficients $a_{ij} = \frac{X_{ij}}{X_j}$, $i, j = 1, \ldots, n$. Each sector uses intermediate inputs from other economic sectors to deliver the required final output. An increase in the final demand for the output of one sector thus indirectly stimulates economic activities in other sectors depending on the technological connections between sectors. The lower part of the table includes set of environmental variables and their distribution by sectors and final uses.

		Economic sectors	Final uses	Output
		$1 \dots j \dots n$		
Economic sectors	1	$X_{11} \dots X_{1j} \dots X_{1n}$	Y_1	X_1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	i	$X_{i1} \dots X_{ij} \dots X_{in}$	Y_i	X_i
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n	$X_{n1} \dots X_{nj} \dots X_{nn}$	Y_n	X_n
Imports		$M_1 \dots M_j \dots M_n$	M_Y	M
Net taxes on products		$T_1 \dots T_j \dots T_n$	T_Y	T
GVA		$V_1 \dots V_j \dots V_n$		V
Output		$X_1 \dots X_j \dots X_n$	Y	X
Set of environmental variables				
Use of energy in MJ				
Electricity		$B_{11} \dots B_{1j} \dots B_{1n}$	B_{1Y}	$B1$
Crude oil and derivatives		$B_{21} \dots B_{2j} \dots B_{2n}$	B_{2Y}	$B2$
Gas		$B_{31} \dots B_{3j} \dots B_{3n}$	B_{3Y}	$B3$
Greenhouse gas emissions (in tonnes)				
CO2		$D_{11} \dots D_{1j} \dots D_{1n}$	D_{1Y}	$D1$
CH4		$D_{21} \dots D_{2j} \dots D_{2n}$	D_{2Y}	$D2$
SO2		$D_{31} \dots D_{3j} \dots D_{3n}$	D_{3Y}	$D3$
Use of water and waste disposal				
Water (in litres)		$W_{11} \dots W_{1j} \dots W_{1n}$	W_{1Y}	$W1$
Waste (in tonnes)		$W_{21} \dots W_{2j} \dots W_{2n}$	W_{2Y}	$W2$

Table 2: *Environmentally extended input-output model, a general case (Source: Authors systematization according to [6] and [24])*

The standard IO model can be extended by environmental variables usually available in physical units as presented in the lower part of Table 2. Analysis can include a broad set of environmental variables disaggregated by industries and final demand: use of specific energy products, GHG emissions, water use, and waste disposal. As each economic sector uses intermediate inputs delivered by other sectors, total emissions related to the production process of one sector also induce indirect emissions related to the production of required inputs. The environmental emission matrix D represents the environmental impacts related to GHG emissions associated with cross-sectoral interactions, which includes direct effect coefficients. Theoretically, matrix D is $k \times n$ matrix where k is the number of rows determined by the number of indicators of interest for environmental protection, and n is the number of economic sectors of the standard IO model [23, 24]. For each type of GHG, direct emission intensity can be calculated for each sector [6]. Elements $d_{ij} = \frac{D_{ij}}{X_j}, i = 1, ..., k, j = 1, \dots, n$ of matrix D are expressed as the ratio of a specific indicator of environmental effects or the quantity of emissions to the gross output of a particular economic sector. This study is focused on CO2 emissions only. Thus, matrix D is a row vector whose elements present the ratios of CO2 emissions to the output of each sector. Croatian IO table is available for 64 mutually exclusive sectors; thus matrix D in this study is 1×64 row vector.

Using the EEIO framework, both direct and indirect emissions of pollutants can be calculated as they relate to the production of a specific sector and the emission of pollutants in the entire value-added chain of a particular industry. The emission multiplier thus includes direct and indirect emissions of a specific sector along the supply chain. It estimates the total (direct and indirect) CO2 emissions required to produce the unit value of each sector (CO2 per EUR 1 million of sectoral output). In this study, the focus is exclusively on the CO2 emissions and matrix

$$X^{e*} = D \cdot L \tag{1}$$

1×64 row vector, where $L = (I - A)^{-1}$, square 64×64 matrix, is the Leontief inverse matrix [24], whose elements show the sum of each sector's direct and indirect output required per unit of the output produced and delivered by another industry to the final user.

Information on total emissions related to the production of individual sectors is helpful, especially for those economic sectors that do not emit significant amounts of pollutants but use intermediate inputs produced by other industries that could be essential pollutants. However, the disadvantage of this approach is potential double counting if the output of each sector is not separated into the production of final products and intermediate inputs. The CO2 emissions allocated to the output of, for example, the electricity sector could be allocated to the emissions of all other sectors producing final goods and services where energy is an essential input. The problem of double counting in the distribution of total emissions to economic industries can be solved by multiplying sectoral emission multipliers by a vector of final demand instead of a vector presenting sectoral output. In that way, emissions related to the production of intermediate inputs are redistributed from the input producer to the sector which produces the final product.

If the total pollution matrix is marked with X^* then the total emission of pollution related to the production processes necessary for the delivery of goods and services according to the given final demand equals:

$$X^* = D^* \cdot L \cdot Y \tag{2}$$

where D^* is a diagonal matrix derived from a row vector D. Matrix Y is $n \times 1$ column vector of final demand, which can be further disaggregated into components: final demand of households, government consumption, gross fixed capital formation, and exports.

Unlike the direct emissions emitted by the production sectors, the responsibility for indirect emissions, i.e. those that occur in the production chain, can be considered from different aspects. According to [29], this paper uses an approach in which emissions are allocated to industries producing final demand, with exports included. The above implies that the emissions of products intended for export abroad, where they are used both for final and intermediate consumption, are added to the total national emissions. Unfortunately, statistical data on the exact purpose of exported products (intermediate or final use of Croatian products in the country where products are exported) is unavailable. Most empirical studies focused on individual countries applied the same approach and treated total exports as the final demand. Separating exported products for intermediate or final use in the importing economy would require an abundant data set on sectoral emissions for all trade partners and the application of a world IO table.

Besides economic sectors, total emissions of CO2 from productive sectors analysed in this paper are divided into components of final demand, i.e. personal consumption, government consumption, investments, and exports. To distribute domestic emissions into components of final demand for each productive sector, formula (2) was applied where Y presents individual components of final demand.

IO data for different periods should be expressed at constant prices to estimate trends in the CO2 emission dynamics relative to actual economic activities. IO tables for 2015 and 2020 are officially published at current prices and prices of the previous year, while IO tables at constant prices of the base year are not published. Thus, the transformation of IO into constant 2020 prices is based on the approach proposed by [20] and the application of the double deflation method. The chain-linking technique estimates the total price changes over the analysed period. Based on this approach, IO data for 2015 are converted to constant 2020 prices. The ratio of the current price to the base-year price level for sector i is expressed as p_i; the crucial point in the calculation is diagonal $n \times n$ matrix P (in this study 64×64) that contains the elements p_i (cumulative price index for each sector in period 2015-2020) on the main diagonal and off-diagonal entries are zero. The following formula is applied to transform IO 2015 data to IO at constant 2020 prices:

$$X_r = A_r X_r + Y_r \tag{3}$$

where $X_r = P^{-1}X$ is $n \times 1$ output matrix, $Y_r = P^{-1}Y$ is $n \times 1$ final demand matrix and $A_r = P^{-1}AP$ is $n \times n$ matrix of IO coefficients all expressed at constant 2020 prices.

IO tables for the Croatian economy in current prices and prices of the previous period for 2015 and 2020 were downloaded from [10], and CO2 emissions per Croatian sectors from [11].

4. Research results

Results present direct, indirect and total CO2 emissions distributed to economic sectors and final demand for 2015 and 2020. Sectoral distribution and trends in direct and indirect emissions of pollutants provide an analytical background for formulating policy recommendations oriented toward reducing the negative environmental impact of economic growth.

Sector code	Sector	2015	2020	Index 2020/2015	Structure 2020
D	Electricity, gas, steam and air conditioning	3,708	3,290	88.7	19.3
C23	Other non-metallic mineral products	2,469	2,640	106.9	15.5
C19	Coke and refined petroleum products	1,409	858	60.9	5.0
H49	Land transport services and transport services via pipelines	847	909	107.4	5.3
A01	Products of agriculture, hunting and related services	645	718	111.3	4.2
C20	Chemicals and chemical products	642	604	94.2	3.5
B	Mining and quarrying	489	495	101.1	2.9
H50	Water transport services	432	257	59.4	1.5
F	Constructions and construction work	279	328	117.7	1.9
C10-C12	Food, beverages and tobacco products	273	253	92.6	1.5
Ten sectors with the highest emissions		11,194	10,355	92.5	60.8
Other economic sectors		1,804	1,809	100.2	10.6
Total emissions by producers		12,999	12,163	93.6	71.4
Heating/cooling activities by households		1,036	1,036	1,018	6.0
Transport activities by households		3,871	3,871	3,560	20.9
Other activities by households		303	303	300	1.8
Total CO2 emissions		18,208	17,041	93.6	100.0

Table 3: *Direct emissions of CO2 in 000 tonnes as recorded by official statistics, Republic of Croatia (Source: Eurostat database)*

Total CO2 emissions, as available from official statistical sources (Table 3), are recorded separately for production activities and pollution emitted directly by final consumers, who use energy products for household heating, cooling, and transportation. In 2020, the production processes of domestic companies were responsible for 71.4 percent of total CO2 emissions. It can be noticed that pollution is concentrated in a few critical sectors. Thus, electricity production is responsible for almost 20 percent of total CO2 emissions, followed by producing other non-metallic mineral products (mainly construction materials), refined petroleum products and land transport. The ten sectors with the highest emissions presented in Table 3 produce over 60 percent of total emissions. CO2 emissions related to non-productive activities are mostly related to use of personal vehicles of citizens and energy consumed for heating or cooling.

Direct emissions intensity $d_{ij} = \frac{D_{ij}}{X_j}$ is a relative indicator defined as the ratio of sectoral emissions (in physical quantities) and gross output (measured in monetary terms). It is usually expressed as Kg CO2 emitted by the production of output valued at 1 thousand EUR and based on data directly observable from official statistical data sources. On average, output valued at 1 thousand EUR in Croatia generated 144 kg CO2 (in 2020), which is significantly lower compared to 2015 when the same indicator was 171. The direct pollution intensity is the highest in producing other non-metallic mineral products, electricity, transport, petroleum and chemical products (Figure 1). Production processes used in those sectors require significant amounts of energy from fossil sources, which transformation pollutes the environment. Sectors with the best results in reducing direct CO2 emission intensity were water transport, production of coke and refined petroleum products, mining and quarrying and production of other non-metallic mineral products. Interestingly, because of the decrease in economic activity of domestic energy producers, direct emission intensity per output in electricity production increased despite the reduction in total emissions.

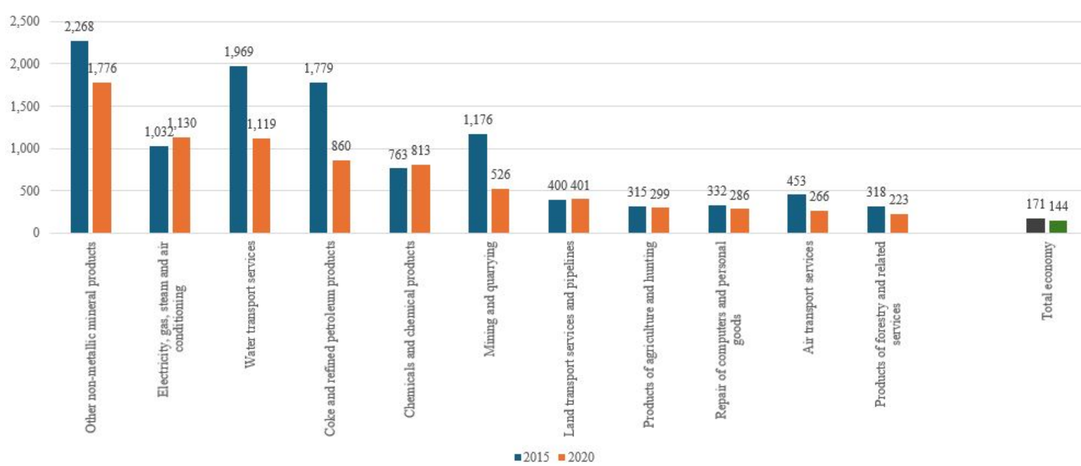


Figure 1: *Direct emissions intensity, Kg CO2 emitted by the output valued at 1 thousand EUR. (Source: Authors calculation based on Eurostat data on CO2 emissions and Croatian IO table.)*

As presented in the methodological part of the paper, the production of final products, besides direct, includes indirect input requirements for intermediary products. Thus, specific sectors do not emit significant amounts of pollutants directly but could require inputs from other sectors with much higher emissions. Thus, those sectors' final product demand contributes to the total carbon footprint along the supply chain. Total CO2 emissions, including direct and indirect emissions along the supply chain, are calculated using the formula (1) and are to be interpreted as the quantity of CO2 emitted to produce a sectoral output of 1 thousand EUR (Figure 2). Besides the sectors mentioned above with high direct emissions, the list of important pollutants includes construction, fishery and operation of travel agencies that directly record low emissions but use intermediate inputs related to higher emissions.

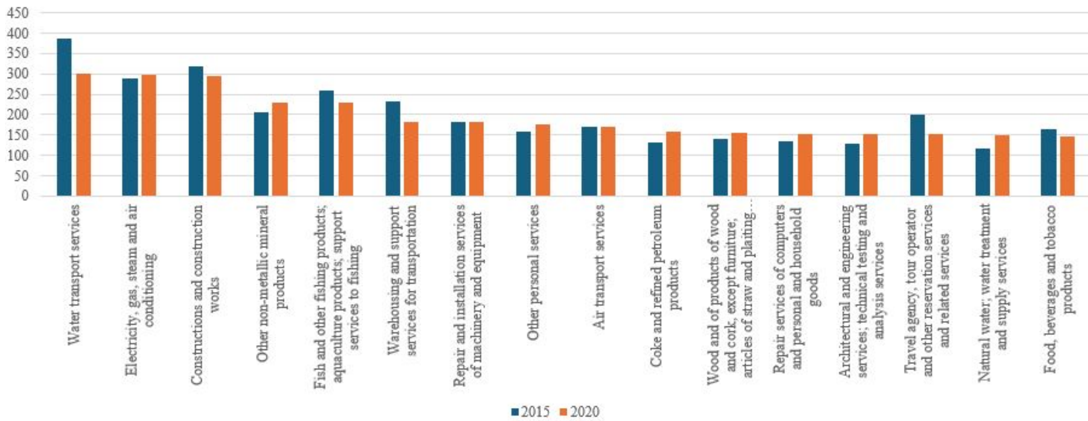


Figure 2: *Total emissions of CO2 in Kg emitted by the output valued at 1 thousand EUR. (Source: Authors calculation based on Eurostat data on CO2 emissions and Croatian IO table.)*

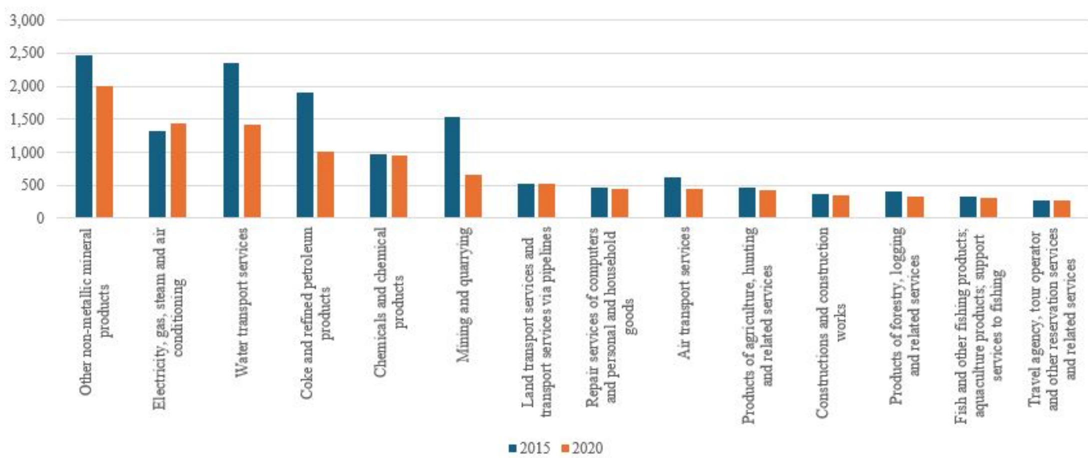


Figure 3: *Indirect emissions intensity, Kg CO2 emitted by the output valued 1 thousand EUR. (Source: Authors calculation based on Eurostat data on CO2 emissions and Croatian IO table.)*

Figure 3 presents the intensity of indirect emissions for the economic sectors with the highest indirect CO2 emissions (difference between total and direct emissions). Water transport, electricity, and other non-metallic mineral products are sectors that, besides direct emissions, also have significant requirements for inputs from different sectors with a high carbon footprint. Besides those sectors where both direct and indirect emissions are high, specific sectors such as construction, warehousing and supporting services, repair and installation of equipment, wood industry, water supply and food industry, despite low direct emissions, through the requirements for intermediate inputs, significantly contribute to the overall pollution.

An EEIO model is convenient for calculating the effects of final demand on sectoral CO2 emissions. To avoid double counting when calculating the national carbon footprint, sectoral emission intensity, as presented in Figure 2, should not be multiplied by the total value of sectoral output but by the value of the sectoral final uses. Total emissions induced by final demand for a particular sector are calculated using formula (2). By that approach, the emissions of industries that deliver intermediate inputs are distributed by EEIO analysis to sectors that produce final products. When households buy final products, such as food, it increases the output of the food industry and the revenues in sectors producing intermediate inputs, some of which, such as agricultural products or energy, could have high CO2 emission intensity. In Table 4, emissions of CO2 for electricity include only emissions incorporated in final electricity expenditures (electricity bought by households, government, and non-profit organisations). The emissions related to the production of energy, which are used as intermediate input of another sector (for example, the food industry), are in Table 4, distributed to the other sectors producing the final goods and services demanded by the final consumer.

Comparing Table 3 and Table 4, one can see that approximately one-half of direct CO2 emitted by electricity producers is generated by final deliveries of (1,605 thousand tonnes in 2020, as presented by Table 4, compared to 3,290 of direct emissions presented by Table 3). The other half of electricity emissions result from the production of intermediate inputs required by other industries. On the contrary, the construction sector emits only 328 thousand tonnes of CO2 directly, while total emissions related to final deliveries of construction services are estimated at 1,636 thousand tonnes. Thus, the construction sector is responsible for the most sizeable proportion of the total CO2 emissions. This is because of the high proportion of inputs required from other non-metallic mineral products (mostly construction materials), transport, and energy.

	2015	2020	Index 2020/2015	Structure 2020
Constructions	1,563	1,636	104.7	9.6
Electricity, gas, steam	2,366	1,605	67.8	9.4
Land transport services, pipelines	777	956	123.0	5.6
Other non-metallic mineral products	814	801	98.4	4.7
Food, beverages and tobacco products	805	779	96.8	4.6
Products of agriculture, hunting	428	518	121.1	3.0
Wholesale trade services	451	509	112.8	3.0
Retail trade services	421	481	114.2	2.8
Coke and refined petroleum products	912	423	46.4	2.5
Chemicals and chemical products	356	374	105.1	2.2
Public administration and defence	469	335	71.4	2.0
Human health services	188	298	159.1	1.8
Accommodation and food services	391	271	69.4	1.6
Imputed rents of owner-occupied dwellings	240	263	109.6	1.5
The sum of the most important sectors	10,178	9,249	90.9	54.3
Other sectors	2,820	2,914	103.3	17.1
Total emissions induced by producers	12,999	12,163	93.6	71.4
Heating and transport of households	5,209	4,878	93.6	28.6
Total emissions	18,208	17,041	93.6	100.0

Table 4: *Total emissions induced by final demand for a particular sector, CO2 in 000 tonnes, Republic of Croatia. (Source: Authors calculation based on Eurostat data on CO2 emissions and Croatian IO table.)*

Like electricity, in the case of land transport and production of non-metallic mineral products, the share of emissions induced by final demand is significantly lower than the sector's total emissions. On the contrary, sectors such as food, beverages, tobacco products, wholesale and retail trade, human health services, and accommodation and food services recorded significantly higher CO2 emissions if indirect pollution is included.

Although total emissions induced by producers decreased by 6.4 percent in the analysed period, the increase in total emissions is estimated for construction, land transport, agriculture, trade, chemicals and chemical products, human health services and imputed rents of owner-occupied dwellings. The highest reduction in total emissions is found in producing coke and refined products and electricity. In the case of the output of refined products, it is probably the result of the decreasing production of oil derivatives because of the downsizing of the domestic output in the Croatian company INA. Instead of two refineries, only one refinery (located in Rijeka) is expected to be modernised. In contrast, the production of oil derivates in the other refinery in Sisak was reduced and recently fully terminated. Reducing CO2 emissions in electricity primarily results from promoting renewable energy plants, such as wind and solar energy, which increase their share in the total electrical energy supply.

Total emissions could be decomposed into the individual components of the final demand (personal consumption, government consumption, gross fixed capital formation and exports) using formula (2). As personal consumption is the crucial component of final demand, its carbon footprint is also the highest (Table 5). In 2020, of total emissions, 29.3 percent were induced by the production of output delivered to households, while an additional 28.6 percent was emitted by household energy for heating and transport. After personal consumption, the highest contribution in total emissions is estimated for exported goods and services (19.4 percent of total emissions in 2020), while the share of government consumption and investment is slightly above 10 percent. In relative terms (emissions induced by 1 million EUR of expenses), pollution intensity is the highest for exports and investments. It is the consequence of the structure of items where the share of various manufacturing sectors is higher than that of services where production processes are more environmentally friendly.

Component of the final demand	Emissions induced by		Structure		Index 2020/2015	Emissions	
	2015	2020	2015	2020		2015	2020
Personal consumption	6,390	4,986	35.1	29.3	78.0	269.3	217.5
Government consumption	1,158	1,726	6.4	10.1	149.0	128.0	140.0
Investment	1,829	2,151	10.0	12.6	117.6	269.5	256.1
Exports	3,621	3,301	19.9	19.4	91.1	385.6	318.5
Total emissions induced by domestic output	13,001	12,165	71.4	71.4	93.6	255.1	217.1
Heating and transport	5,209	4,878	28.6	28.6	93.64		
Total	18,208	17,041	100.0	100.0	93.59		

Table 5: *Total emissions induced by individual components of final demand, CO2 in 000 tonnes, Republic of Croatia. (Source: Authors calculation based on Eurostat data on CO2 emissions and Croatian IO table.)*

5. Conclusion

An EEIO approach is applied in this research to allocate CO2 emissions in the Croatian economy to components of final demand and economic sectors which deliver final goods and services.

According to the results presented above, the production of construction materials, electricity, transport, agriculture and construction services are responsible for the most sizable proportion of the total CO2 emissions. Significant emission intensity in producing energy products and transport is usually found in all previous studies for many other economies, as presented in [2, 4, 30] and mainly include direct emissions. In the case of construction, Croatian results align with the results for Canada [19], where indirect emissions are more important than direct emissions. Construction services require intermediary inputs produced by sectors which emit significant amounts of CO2: construction materials, transport, and energy. The proportion of indirect emissions is also high in sectors such as food, beverages, tobacco products, wholesale and retail trade, human health services, and accommodation and food services. Although personal consumption, in absolute terms, is responsible for the highest share of CO2 emissions, in relative terms (emissions induced by 1 million EUR of expenses), pollution intensity is higher for exports and investments.

A decrease in the total CO2 emissions was found during the analysed period. The energy sector and the production of refined oil products and electricity are the sectors that contribute to overall ecological improvements. In the case of electricity, reducing CO2 emissions is a consequence of promoting renewable sources, such as wind or solar energy. The downsizing reduces emissions in the domestic production of oil derivatives by the INA, a leading Croatian company in the oil industry. Contrary to the general trend, the increase in total emissions is estimated for construction, land transport, agriculture, trade, chemicals and chemical products, human health services and imputed rents of owner-occupied dwellings.

The results of this research can represent guidelines for green economy policymakers in creating further measures and strategies to reduce CO2 emissions not only in production sectors with high direct CO2 emissions but also in those with a significant indirect effect on the environment. In the case of electricity, the Croatian government, by introducing specific measures and incentives, should further promote the transformation of energy production from fossil to renewable energy sources, such as wind, water and solar energy. Positive examples from other countries could be used to expand national energy capacities based on renewable sources and reduce imported fossil energy products [4]. As proposed in previous studies [25], policy measures should be introduced to improve energy efficiency in other sectors where energy is an important input. Promoting investments in modern equipment and technologies is crucial in improving energy efficiency in manufacturing. Innovations in the industry that produce motor vehicles and improvements in the ecological quality of fuel are critical in reducing carbon footprints in transport [2]. In the Croatian case, where the manufacture of motor vehicles is less developed, developing a regulatory framework favouring environmentally more efficient vehicles is more important than regulating vehicle producers.

Besides its advantages, the methodology applied in this research has some limitations. The disadvantage is related to the treatment of total exports as a final demand. However, a certain proportion of exported products is intended for intermediate consumption abroad and should be allocated to the emissions for producing final products in the importing economy. The problem of potentially double-counting CO2 emissions could arise if multipliers are applied to total sectoral output due to overlapping production chains of different sectors. As in previous studies, the problem of double counting in this research is solved by the multiplication of emission multipliers by final demand. In this way, emissions related to the production of intermediary inputs are transferred from the original producer to the producer of the final product.

For future research, the authors plan to apply the described approach to the comparative analysis of CO2 emissions between European Union economies. Also, the environmental input-output analysis framework is expected to investigate CO2 emissions embodied in international trade.

References

[1] Albizzati, P., Rocchi, P., Cai, M., Tonini, D. and Astrup, T. F. (2022). Rebound effects of food waste prevention: Environmental impacts. Waste management, 153, 138-146. doi: 10.1016/j.wasman.2022.08.020

[2] Alcantara, V. and Padilla, E. (2006). An input output analysis of "key" sectors in CO2 emissions from a production perspective: an application to the Spanish economy. Universitat Autonama of Barcelona, Spain. Working papers. url: https://ddd.uab.cat/pub/estudis/2006/hdl_2072_2143/wpdea0601.pdf [Accessed 12/11/2024]

[3] Ayres, R. and Kneese, A. (1969). Production, consumption and externalities. The American Economic Review, 59(3), 282-297.

[4] Beidari, M., Lin, S. J. and Lewis, C. (2017). Multiplier Effects of Energy Consumption and CO2 Emissions by Input-Output Analysis in South Africa. Aerosol Air Qual. Res. 17, 1666-1678. doi: 10.4209/aaqr.2017.04.0150

[5] Chapagain, S. K., Mohan, G., Rimba, A. B., Payus, C., Sudarma, I. M. and Fukushi, K. (2022). Analyzing the relationship between water pollution and economic activity for a more effective pollution control policy in Bali Province, Indonesia. Sustainable Environment Research, 32(5), 1-14. doi: 10.1186/s42834-021-00115-6

[6] Csutora, M. and Vetőné Mózner, Z. (2024). A fosszilis inputok és outputok teljes költsége az ágazati kapcsolatok mérlege alapján: Kína példája= The total cost of fossil inputs and outputs based on input-output tables–the example of China. Statisztikai Szemle, 102(2), 158-186. doi: 10.20311/stat2024.02.hu0158

[7] Čegar, S., Denona Bogović, N., and Jugović, A. (2022). Impact of intersectoral dependencies in national production on wastewater discharges: An extended input–output study of the Croatian economy. Water, 14(13), 2122. doi: 10.3390/w14132122

[8] Demeter, C., Lin, P-C., Sun, Y-Y. and Dolnicar, S. (2022). Assessing the carbon footprint of tourism businesses using environmentally extended input-output analysis. Journal of Sustainable Tourism, 30(1), 128-144. doi: 10.1080/09669582.2021.192418

[9] Dobos, I. and Tóth-Bozó, B. (2025). Calculation of ecological land-footprint – based on the input-output model and focusing on the imported commodities. Croatian operational research review, 16(1), 65-72, doi: 10.17535/crorr.2025.0006

[10] Eurostat. (2024a). Supply, use and input-output tables. doi: https://ec.europa.eu/eurostat/databrowser/view/naio_10_cp1700 [Accessed 14/05/2024]

[11] Eurostat. (2024b). Air emissions account by NACE Rev. 2 activity. doi: https://ec.europa.eu/eurostat/databrowser/view/env_ac_ainah_r2 [Accessed 14/05/2024]

[12] Firdaus, Z. F. (2020). Environmentally-Extended Input Output Tables – Integrated Analysis for Low Carbon Development in Indonesia. 2020 Asia–Pacific Statistics Week. Bangkok, Thailand. url: https://www.unescap.org/sites/default/files/APS2020/SC1_16_Environmentally-Extended_Input-Output_Tables_Low_Carbon.pdf [Accessed 12/11/2024]

[13] He, H., Reynolds, C. J., Zhou, Z. X., Wang, Y. and Boland, J. (2019). Changes of waste generation in Australia: Insights from structural decomposition analysis. Waste management. 83, 142-150. doi: 10.1016/j.wasman.2018.11.004

[14] He, F., Yang, Y., Liu, X., Wang, D., Ji, J. and Yi, Z. (2021). Input–Output Analysis of China's CO2 Emissions in 2017 Based on Data of 149 Sectors. Sustainability 13, 4172. doi: 10.3390/su13084172

[15] Imansyah, M. H. and Putranti, T. (2017). The Identification of Key Sector in CO2 Emissions in Production Perspective of Indonesia: An Input Output Analysis. International Journal Sustainable Future for Human Security, 5(2), 21-29. doi: 10.24910/jsustain/5.2/2129

[16] Kitzes J. (2013). An Introduction to Environmentally-Extended Input-Output Analysis. Resources, 2(4), 489-503. doi: 10.3390/resources2040489

[17] Keček, D., Mikulić, D. and Fotova Čiković, K. (2022). Economic contribution and integration of Croatian ICT sectors. Croatian operational research review, 32(2), 161-172. doi: 10.17535/crorr.2022.0012

[18] Leontief, W. (1970). Environmental repercussions and the economic structure: an input-output approach. Review of Economics and Statistics, 52(3), 262-271.

[19] Liu, L., Huang, G., Baetz, B., Cheng, G., Pittendrigh, S. M. and Pan, S. (2020). Input-output modeling analysis with a detailed disaggregation of energy sectors for climate change policy-making: A case study of Saskatchewan, Canada. Renewable Energy, 151, 1307-1317, doi: 10.1016/j.renene.2019.11.136

[20] Llop, M. (2017). Changes in energy output in a regional economy: A structural decomposition analysis. Energy, 128, 145–151. doi: 10.1016/j.energy.2017.04.003

[21] Lu, P., Xiangzheng, D. and Zhihui, L. (2023). An extended input-output analysis of links between industrial production and water pollutant discharge in the Yangtze River Economic Belt, Journal of Cleaner Production, 390, 136115. doi: 10.1016/j.jclepro.2023.136115

[22] Meshulam, T., Font-Vivanco, D., Blass, W. and Makov, T. (2023). Sharing economy rebound: The case of peer-to-peer sharing of food waste. Journal of Industrial Ecology, 27(3), 882-895. doi: https://doi.org/10.1111/jiec.13319

[23] Mikulić D. (2018). *Osnove input-output analize s primjenom na hrvatsko gospodarstvo*. Zagreb: Ekonomski institut. url: https://www.eizg.hr/publikacije/knjige/osnove-input-output-analize-s-primjenom-na-hrvatsko-gospodarstvo/2837 [Accessed 12/11/2024]

[24] Miller, R. E. and Blair, P. D. (2009). *Input-output analysis: Foundations and extensions*. 2nd Edition, Cambridge: Cambridge University Press.

[25] Moon, J., Yun, E. and Lee, J. Identifying the Sustainable Industry by Input–Output Analysis Combined with CO2 Emissions: A Time Series Study from 2005 to 2015 in South Korea. Sustainability, 12, 6043. doi: 10.3390/su12156043

[26] Othman, J. and Jafari, Y. (2016). Identification of the key sectors that produce CO2 emissions in Malaysia: application of input–output analysis. Carbon Management, 7(1–2), 113–124. doi:10.1080/17583004.2016.1178451

[27] Schäfer, D. and Stahmer, C. (1989). Input-output model for the analysis of environmental protection activities. Economic Systems Research, 1(2), 203–228. doi: 10.1080/09535318900000015

[28] Schäfer, D. and Stahmer, C. (1990). Conceptual considerations on satellite systems. Review of Income and Wealth, 36(2), 167–176. doi:10.1111/j.1475-4991.1990.tb00278.x

[29] Schoenaker, N. and Delahaye, R. (2017). Environmental Input-Output Analyses for the Netherlands. Project and report commissioned by the European Union, Project of Directorate E, Eurostat, European Commission, Grant Agreement Number 05121.2016.001-2016.274. url: https://www.cbs.nl/-/media/_pdf/2018/06/290118-report-io-analysis.pdf [Accessed 12/11/2024]

[30] Shmelev, S. and Brook, H. R. (2021). Macro Sustainability across Countries: Key Sector Environmentally Extended Input-Output Analysis. Sustainability, 13, 11657. doi: 10.3390/su132111657

[31] Stadler, K. et al. (2018). EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables. Journal of Industrial Ecology, 22(3), 502-515. doi: 10.1111/jiec.12715

[32] Stahmer, C. (2000). The magic triangle of input-output tables. url: https://ideas.repec.org/h/elg/eechap/1735_7.html [Accessed 12/11/2024]

[33] Raa, T. (2005). The economics of input-output analysis. Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511610783

[34] Vukić, L., Mikulić, D. and Keček, D. (2021). The impact of transportation on the Croatian economy: the input–output approach. Economies, 9(1), 7. doi:10.3390/economies9010007

[35] Zha, D. L., Chen, Q. and Wang, L. J. (2022). Exploring carbon rebound effects in Chinese households' consumption: A simulation analysis based on a multi-regional input-output framework. Applied energy, 313, 118847, doi:10.1016/j.apenergy.2022.118847

Evaluating upskilling tools for digital, green, and entrepreneurial competences: A multicriteria approach to advancing Europe's green and digital transition

Ivana Ninčević Pašalić[1,*], **Maja Ćukušić**[1] and **Ljiljana Najev Čačija** [1]

[1] *Faculty of Economics, Business and Tourism, University of Split, Cvite Fiskovića 5, 21000 Split, Croatia*
Corresponding author e-mail: ⟨ivana.nincevic.pasalic@efst.hr⟩

Abstract. Advancing green and digital transition is a key topic on Europe's agenda, aligned with the UN's SDG goals. In this context, this study investigates existing support tools and programs for upskilling digital, green, and entrepreneurial (DGE) competences with the goal to select and reuse the best ones available within the project partner group (11 partners from 9 European countries). The data collection process started with the identification and thorough elaboration of 90 upskilling tools/programs, each differing in their contributions to DGE competences, and in other varying features. The collected database contains more than 30 qualitative and quantitative features per upskilling tool/program. An expert focus group then defined the tool selection criteria and their grading, highlighting the importance of learning outcomes recognition and the development of multiple DGE competence areas, in particular. After the grading process, alternative assessments were analysed using the PROMETHEE II method, resulting in a ranking of the 15 most useful DGE tools/programs. In addition to validating the framework set up for the MCDM that takes into account the complexities of (international) training needs of diverse target groups, the study demonstrated significant application potential. Namely, these tools, both the initial list of 90 and the selected 15, all having confirmed their effectiveness and success in previous projects addressing SDG 4 and SDG 9, are now promoted to (new) relevant stakeholders and prioritised in capacity-building actions in international settings. As such, the collective exercise based on strong scientific foundations had clear practical value.

Keywords: PROMETHEE, SDG4 Quality education, SDG 9 Industry, innovation and infrastructure, upskilling

Received: July 29, 2024; accepted: December 4, 2024; available online: February 5, 2025

DOI: 10.17535/crorr.2025.0013

Original scientific paper.

1. Introduction

The transition towards a green and digital economy has become a central focus on Europe's agenda, reflecting global commitments such as the United Nations' Sustainable Development Goals (SDGs). As economies and societies strive to meet these goals, the demand for effective upskilling tools and programs has never been more pressing. SDG 4, which emphasises quality education, and SDG 9, which focuses on industry, innovation, and infrastructure, emphasise the need for developing competences in digital, green, and entrepreneurial (DGE) areas, in particular. [13] raises the importance of competence frameworks in shaping the future workforce through education and estimating the future performance of existing business operations. Similarly, [17] in their study of sustainable entrepreneurship recognise digital technology support as

*Corresponding author.

beneficial in avoiding and exploiting the potential of environmental and social challenges. Still, the body of knowledge addressing digital and sustainable entrepreneurship is too dispersed and fragmented to serve as the foundation for development (ibid).

While the fourth industrial revolution era is based on digitalisation, with economic growth and social progress depending on technological innovation, the fifth industrial revolution can be seen as green – merging ecology, technology, sociology, science, math, and economics fields [16]. Eventually, it will lead to a more sustainable future. The role of education in this transition is of great importance. To achieve its potential, educational practices need to be transformed into learning on matters related to sustainable development [25] and genuine commitment to educational paradigm change. Accordingly, existing competence frameworks brought forth by the European Commission – DigComp, GreenComp, and EntreComp, can be valuable tools in estimating the current situation and clarifying future paths.

This study seeks to address the gap between current educational offerings and the competencies required for sustainable and innovative growth. Building on the expertise of DGE training and support, the primary objective was to identify and rank the most effective tools and programs available to a collaborative network of 11 partners from nine countries. The scope of this investigation covers a diverse range of 90 tools and programs, each varying in its approach and contribution to DGE competence development. The important questions for businesses, individuals, education institutions and business support organisations are related to the selection of DGE tools and programs that will enhance their (individual) competences, but at the same time bring value to their respective organisations. These questions clearly point to a multi-criterial problem faced by decision-makers, which is investigated in this occasion using a Multi-Criteria Decision-Making (MCDM) method as a framework for scientifically sound decision support in an effort to effectively upskill future and existing workforce, in particular their DGE competences.

With regards to the contribution, by systematically analyzing a diverse range of tools and programs, the paper aims to provide a comprehensive portfolio that not only enhances individual competences but also adds significant value to organizations and educational institutions. The contribution lies not only in the identification and ranking of these tools but also in the establishment of a robust methodological framework that incorporates expert evaluations and stakeholder insights. This framework is designed to facilitate the effective upskilling of the current and future workforce, particularly in the realms of digitalization and sustainability. Furthermore, by promoting the selected tools to relevant stakeholders and prioritizing them in capacity-building actions, this research has the potential to drive significant advancements in the green and digital transition across Europe. Consequently, this study not only enriches the academic discourse on integrating multiple competence frameworks but also provides actionable insights grounded in empirical evidence. By empowering stakeholders to make data-informed decisions regarding the implementation of impactful tools and programs, the aim is to advance sustainable and inclusive growth in various educational and professional sectors.

The paper is organized as follows: Section 2 provides a literature overview of DGE competences. Section 3 explains the research data and methodology. Section 4 presents the research results, including a sensitivity analysis and discussion. Finally, Section 5 offers concluding remarks.

2. Literature review

A relatively scarce but emerging body of literature on sustainable entrepreneurship emphasises the potential of digital technologies to prevent further development of threats related to environmental and social challenges [17]. Sustainable and innovative growth depends, at least partly, on quality entrepreneurship education and training programs. Still, a slow pace in adaptation to circularity and sustainability can be seen in entrepreneurship program development due to a need for more experience in introducing this important topic into study programs [38]. Green topics and digital technologies will play a crucial role in this transition [16], while the complexity and importance of the subject raise the urge for a multi-dimensional approach and solutions.

In that sense, critical terms related to the further development of the discipline are digitalisation, sustainability and entrepreneurship, whereas their interconnections should be investigated further ([29]; [15]). [17] proposed a comprehensive framework based on a systematic literature review that contributes to the role of business models for sustainability with the reconceptualisation of digital technologies as business model actors. In this model, sustainability and digitalisation are integrated into business operations, while outcomes are oriented toward multiple stakeholders, venture viability and value for society and the environment.

Furthermore, [18] emphasise the vulnerability of SMEs due to sustainability and digitalisation issues with a high level of reactive response. It can be argued that a higher level of competences in digital, green, or entrepreneurship will lead to higher resilience of actors in today's fast-changing environment. Therefore, entrepreneurs and future entrepreneurs should be able to evaluate their competences to improve them in the future and become more resilient to crises.

According to [36], entrepreneurial competences mediate the relationship between digital entrepreneurial education and entrepreneurial intentions. [12] noted that the promotion of entrepreneurial skills among students might contribute to sustainable development due to the development of entrepreneurial skills of self-consciousness and its relationship with environmental commitment. In that sense, DigComp [39], GreenComp [8] and EntreComp [6] frameworks can be helpful tools to estimate and track the progress toward upskilling for sustainability.

There is a rise of interest in exploring these frameworks' usefulness and operationalisation. Still, most studies focus on a single competence framework. For example, [22] claim that EntreComp can be addressed as a unidimensional construct which explains start-up behaviour. [35] emphasise the urge for further support in EntreComp implementation in policy-driven frameworks, while [6] emphasises complexity in the development of entrepreneurial competences measurement scales, therefore, an unclear vision of what competences should be developed through education programs.

The body of knowledge on DigComp and GreenComp implementation and effectiveness is also limited. [14] analysed initiatives at the EU level related to digital readiness of higher education, emphasising challenges relatively similar to prior findings and determining the main obstacles in developing digital skills. Although DigComp provides a structured approach in various elements of digital proficiency [7], the willingness to adopt and implement this approach among key stakeholders is crucial for further development. [32] indicate overlap and cause-related relationship between GreenComp and EntreComp, suggesting that GreenComp should be seen as a part of EntreComp, while their combined usage is a precondition in the transition from system thinking to action. A combination of all frameworks in achieving necessary competences is mandatory for sustainable entrepreneurship education and practice (ibid) in the future.

In terms of upskilling, [3] explored the usability of DigCom in employability after finishing education. Their findings suggest that attitudes moderate the relationship between digital competences and employability, where diverse nature and requirements across different sectors should be considered. GreenComp, a relatively new concept directly related to sustainability, is least present in the current body of knowledge. [37] emphasise its flexibility to serve as a framework for different purposes, which is why it should be widely used for educational and business purposes.

There is a consensus that the European Commission's competence frameworks should become standard in estimating needed competences and creating appropriate training programs accordingly. If the unified approach is adopted, it might be possible to measure, track and compare effects, creating a foundation for equal development toward sustainability. To follow up on this ambition, this study seeks to find its contribution. Despite this agreement that a holistic approach to competence frameworks is needed, ready-to-use solutions still need to be developed. Reasons for this discrepancy can be derived from subject complexity. Namely, according to [11] , frameworks' competences address knowledge, skills, values and attitudes that should be estimated from critical and transformative perspectives. They conclude that this is only possible if competences are perceived from a learning perspective that leads to reductionism and a pragmatic approach to complex subjects. Therefore, other competences, such as emotional management and uncertainty, should be included, with a "learning to be" perspective. The combination of multiple competence frameworks as a tool for sustainable growth can be expected to be in the scope of researchers in the future. Furthermore, the joint effect of technological innovation and green orientation might be crucial elements of an organisation's success [33], where organisations are "forced" to pursue innovation.

In addition to the analysis related to competence frameworks and various aspects of DGE-related upskilling presented above, additional empirical studies employing MCDM in educational or sustainability contexts have been examined. Some, for example, use MCDM methods to benchmark e-Learning tools, helping users and developers select tools with optimal features for training and course development [21], to recommend optimal learning paths for students and teachers to enhance educational outcomes [1]), or to identify key enablers for green entrepreneurship and digitalization [30].

Overall, the literature reveals a scarcity of studies that comprehensively integrate multiple competence frameworks specifically targeting DGE-related upskilling needs across diverse contexts. This study addresses this gap by synthesizing multiple competence frameworks into a unified tool designed to meet the specific needs of DGE upskilling. This comprehensive approach goes beyond existing literature, which often focuses on isolated frameworks or aspects of competence development, by exploring and critically evaluating support tools and programs for enhancing DGE competences among a diverse set of project partners. Through this synthesis, the study not only enriches the academic discourse on integrating multiple competence frameworks but also establishes a ranking system that applies internationally validated selection criteria to prioritize the most effective tools for upskilling. By offering actionable insights grounded in empirical evidence, this research empowers stakeholders in their efforts to build DGE capacities and make data-informed decisions on implementing the most impactful tools and programs, ultimately advancing sustainable and inclusive growth in various educational and professional sectors.

3. Data and Methodology

3.1. Data

Data collection of DGE tools/programs was based on an online questionnaire developed and collated using Qualtrics. The questionnaire was distributed among 11 project partner institutions (Interreg Central Europe Capacity2Transform), all of whom possess extensive expertise in supporting businesses, upskilling, and delivering training programs. There were in total 30 features that merited elaboration per tool/program (including name, category, description, length, delivery flexibility, level of the program, audience, upskilling topic addressed, learning outcome alignment, etc.). Considering 11 professionals entered data (one per institution), two workshops were organised by the authors of this paper, for the participants to reach a common understanding of the purpose of the questionnaire, and the questions and to outline the overall process. After the workshops, data collection started in October 2023 and finished in late December 2023.

At the data collection stage, the targeted number of tools and/or programs/modules was

88 (as each partner had to collect at least 11 tools/programs) and the total collected number was 90 tools/programs addressing various aspects of upskilling support to enhance DGE competences. All collected tools/programs formed a comprehensive project catalogue published at https://capacitytotransform.eu/tool/program-identification-tool/. Each tool was described in an extensive level of quantitative and qualitative detail to capture different features for each item (i.e. determining upskilling topics, delivery mode, learning pace, competence areas, enrolment criteria, intellectual property rights issues etc.).

3.2. Methodology

As discussed earlier, due to the complexities of uncovering the best portfolio of tools/programs that would meet the (international) training needs of diverse target groups, MCDM method was identified as the most effective. Namely, MCDM methods provide an overall ranking of alternatives, from the most preferred to the least preferred, based on a stated set of alternatives (in this particular case 90 DGE tools/programs) and a definite number of decision-making criteria. As the aim was to select the 15 best support tools/programs available from the newly founded catalogue, the tool/program selection process incorporated project partners' collective knowledge and experience. The research model is shown in Figure 1 and each phase is described in the following text.

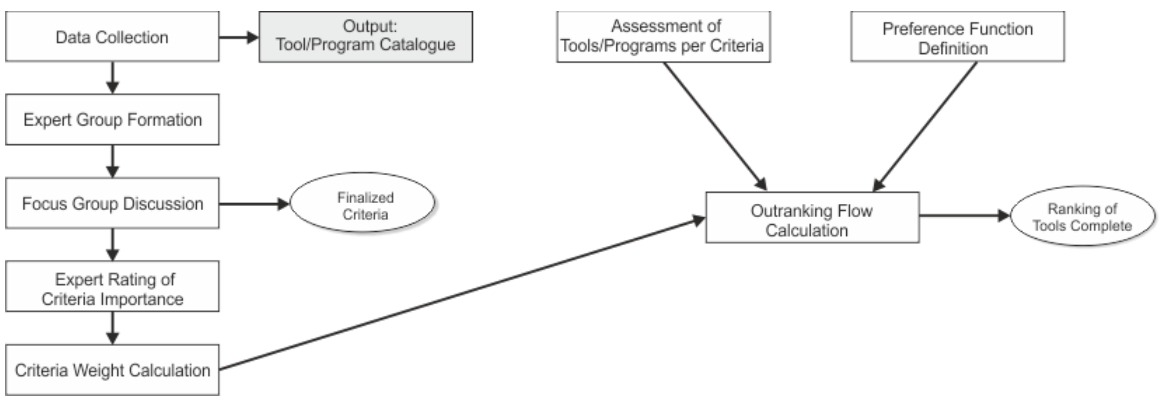


Figure 1: *Study's research model (made by authors)*

First, the criteria formulation process involved creating an expert group comprised of project partners who possess extensive knowledge and experience. All project partner members' competences were evaluated using the DGE competences self-assessment tool European Commission's Digital Competence Framework for Citizen [39], GreenComp: The European sustainability competence framework [8] and EntreComp: The Entrepreneurship Competence Framework [5]. Subsequently, a focus group consisting of seven experts with the highest scores of DGE competences was organised to establish the framework and criteria for tool/program selection. Out of 7 experts, 5 were female, 2 male, 5 had PhD diploma, 2 master diploma and in average they had more than 13 years of experience. The 14 criteria were initially drafted by the focus group moderator (one of the authors of the paper) based on the literature review and were afterwards refined through focus group discussion in a workshop in late January 2024, resulting in a finalised list of 9 criteria (Table 1).

The importance of criteria was assessed by experts as well, using group weight assessment grading. Specifically, after the focus group, the conclusions were collated and 10 experts (3 additional were included who scored the highest in DGE competences) in total were provided with the set criteria (C) and asked to evaluate the importance of the criteria from 1 (not

Criteria (criteria type)	References	Weights
C1 Availability of recognition of learning outcomes - reliability (max)	[2]	0.1089
C2 Development of multiple DGE competences areas (max)	[5]; [8]; [39]	0.1080
C3 Adaptability - individual learning paces (max)	[1]	0.1043
C4 Total number of competences' development (max)	[5]; [8], [39]	0.0999
C5 Level of learner support (max)	[41]	0.0986
C6 Timeframe flexibility of the program (max)	[31]	0.0916
C7 Program delivery mode diversity (offline/online) (max)	[9]	0.0814
C8 Availability of previous users' feedback or rating (max)	[10]	0.0760
C9 Price (min)	[40]	0.0524

Table 1: *Consolidated list of criteria for DGE upskilling tool/program selection (ranked by weight from highest to lowest) (Source:authors)*

important) to 10 (very important). Out of 10 experts, 6 were female, 4 male, 5 had PhD degree, 4 master degree, one bachelors degree and in average they had more than 15 years of experience. The weights assigned to each criterion in Table 1 are derived from expert evaluations using a scale from 1 to 10. It is important to note that these weights do not sum up to 1 or 100% because they represent raw scores (importance of raw score discussed by e.g. [19] reflecting each criterion's importance rather than normalized proportions. Given that there are nine criteria, the total score reflects this count (9) rather than requiring normalization. This method allows for a more nuanced understanding of each criterion's significance in assessing DGE upskilling tools and programs. For each criterion, a criterion type is noted in brackets (min or max) showing if the criteria are to be maximised (benefit criterion) or minimised (cost criterion). The most important criteria for DGE tools per experts opinions are tools' reliability and development of multiple DGE competences areas. The least important one is tool price. The table also outlines relevant and recent studies that were conducted in a similar context (education/training tools) and have focused on a particular criterion; this is listed in the References column.

Out of the prominent MCDM methods, PROMETHEE (Preference Ranking Organization Method for Enrichment of Evaluations) was pinpointed as the optimal in the given scenario (capability to provide clear rankings and allows strict preference thresholds to be set). As a decision-making technique, PROMETHEE helps to determine the relative rankings of the alternatives based on the overall preferences of experts [4]. It takes into account the criteria weights assigned by the experts to prioritise the tools/programs accordingly and assists in converting qualitative data into an ordinal scale. Alternative methods such as AHP were considered; however, they require normalization of scores and may not accommodate the complexity of our criteria as effectively as PROMETHEE II. TOPSIS and VIKOR were also evaluated but were found lacking in flexibility regarding preference functions and ranking capabilities (e.g. [23]). PROMETHEE has been used in different fields including finance (e.g. share investments by [24], management (e.g. supplier selection in blockchain by [26], regional tourism competitiveness (e.g. [27] and city governance (e.g. selection of IoT platforms by [28]. In the field of education, PROMETHEE has been applied for various purposes, for example, evaluation of school performance [20], remote education software evaluation [41], and more recently for ChatGPT in higher education when evaluating collaborative work alternatives [31].

4. Results and Discussion

4.1. Results

By using the PROMETHEE II method, the 15 best DGE tools/programs were then selected based on their performance against the established criteria and their respective weights. In

		C1	C2	C3	C4	C5	C6	C7	C8	C9
	Toola name/Preference Function	Usual	Usual	Usual	V-shape	Usual	Usual	Usual	Usual	Usual
A1	Go Startup Master	1	2	2	20	5	1	2	4	1
A2	DoToho!	1	3	3	11	5	4	5	5	4
A3	EDIH Soc. impact ... and ESG	5	2	2	5	5	3	2	2	3
A4	Startup Ecosystem Canvas	1	3	3	30	3	4	5	1	1
A5	Google Grow tools	5	3	5	21	3	5	4	5	3
A6	Circular bus. models and knowl. base	1	2	4	8	1	3	2	1	1
A7	Co-creation Lab Method./Program	4	2	3	10	5	4	5	4	1
A8	EIC Summer School	5	2	3	10	5	2	3	3	4
A9	EDIH Digital bus. development	5	2	2	12	5	3	2	2	3
A10	ACCESS-3D printing e-learning	4	3	4	12	2	3	5	2	2
...	...									
A90	Upcycling training modules	1	2	5	8	3	5	4	3	1

Table 2: *Decision matrix (Source: authors)*

addition to its benefits regarding the use of qualitative data, this method was chosen as each criterion can be set with particular preference functions. In this study, authors have set, along with the usual criterion, the V-shape criterion or criterion with linear preference for C4 along with the p parameter – i.e. strict preference threshold. If the difference between the total number of competences is higher than 15, there is a strict preference threshold of one tool over the other, while if the difference between the total number of competences is lower that 15, the preference increases linearly. There are maximum 48 competences. The aforementioned has been done to allow us to emphasize significant differences in tool effectiveness when addressing people's competences. According to [34], the V-shape function is characterized by a strict preference threshold (p-value). Along with defining the preference function, the assessment of DGE tools was conducted, the first 10 tools and the 90th have been shown in Table 2.

C2 and C4 were measured numerically where the maximum DGE competence areas are 3 (C2) and there is a maximum of 48 competences that DGE tools/programs can address (consisting of 21 Digital, 12 Green and 15 Entrepreneurial and forming C4). Assessment of the DGE tools/programs per defined criteria was conducted by the authors in two phases: first individually and then in a group decision workshop. This approach was designed to reduce subjectivity and ensure consistency in evaluations. Authors assigned values from 1 to 5 values for descriptive criteria, where 1 signifies "very low", 2 "low", 3 "medium", 4 "high" and 5 "very high". For example, the timeframe flexibility of Go Startup (A1) is assessed as 1 indicating that it is fixed term program and not flexible and therefore assessed as very low.

Utilising the PROMETHEE II method (using Decision Lab 2000 software, version 1.01.0388), the 15 best tools were then selected based on their performance against the established criteria and their respective weights. The positive, negative and net flow along with complete ranking of the alternatives from the best to worst is presented in Table 3 and a short description of an optimal case for using the selected tools/programs is given on link.

4.2. Sensitivity Analysis

To assess the robustness of our findings, a sensitivity analysis was conducted focusing on both the weights assigned to each criterion and the V-shape criterion (C4). The weights for each criterion were changed to weights calculated from first 7 experts ($w1=0.1417$, $w2=0.1080$, $w3=0.1283$, $w4=0.1289$, $w5=0.1248$, $w6=0.1124$, $w7=0.0955$, $w8=0.0902$, $w9=0.0703$) to observe potential shifts in rankings of DGE tools/programs. Additionally, we varied the strict preference threshold associated with the V-shape criterion (C4) to evaluate its impact on overall rankings (by changing preference function $\pm 20\%$). The PROMETHEE II method was reapplied after each adjustment, allowing us to document if there are any significant changes in ranking positions.

The results indicate that while several tools maintained their original rankings across different scenarios, notable changes were observed for certain programs. For example, SmartCulTour Toolkit dropped from rank 3 to rank 4 when new weights were applied, indicating a relative decrease in effectiveness compared to other tools. However, when the weight of the V-shape criterion was decreased, it regained its rank of 3, demonstrating that its performance is more favorable under certain weighting conditions. Conversely, the University Green Digital Hub e-learning program experienced a significant shift, moving from rank 5 to 7 with new weights but rising back to 5 when the V-shape criterion was increased from 15 to 18.

Tool/Program	Orig. model pos. flow	Orig. model neg. flow	Orig. model net flow	Orig. Rank	Rank-new weight	V-shape decr. rank	V-shape incr. rank
Google Grow tools	0.55	0.14	0.41	1	1	1	1
InnoSchool	0.56	0.16	0.40	2	2	2	2
SmartCulTour Toolkit	0.51	0.13	0.38	3	4	3	3
PoMP	0.52	0.15	0.37	4	3	4	4
University Green Digital Hub	0.53	0.20	0.33	5	7	7	5
Start Cup Veneto 2023	0.48	0.15	0.33	6	6	5	6
DIH Tourism 4.0	0.56	0.23	0.33	7	8	6	7
GoStartup Academy – Program Company	0.52	0.19	0.33	8	5	8	8
Sustainability management in companies	0.51	0.21	0.30	9	9	9	9
Life Cycle Assessment basic course	0.49	0.21	0.28	10	11	11	10
GoStartup Academy – Program Startup	0.47	0.20	0.27	11	10	10	11
SDG Identification	0.44	0.18	0.25	12	12	12	12
GostartUp - Video Masterclass	0.45	0.22	0.23	13	13	13	13
Popri Digital Video Course	0.42	0.19	0.23	14	14	14	14
Design Sprint 3.0 / Problem Framing	0.52	0.19	0.21	15	15	15	15

Table 3: *Original preference flows and sensitivity analysis alternative rankings (Source: authors)*

These findings highlight the sensitivity of certain tools to changes in ranking criteria and emphasize the importance of considering both weight adjustments and preference functions in evaluating DGE tools and programs. Future analyses may benefit from further exploration of these dynamics to enhance decision-making processes.

4.3. Discussion

Following the analysis of the results, the selection of upskilling tools/programs aimed at enhancing DGE competences was presented to stakeholders and, upon review and approval, published online as its usefulness was projected to go beyond the international project partnership. Specifically, within this exercise, each tool/program has been evaluated for its effectiveness and relevance, showcasing a diverse range of resources designed to support different aspects of DGE development.

As was expected, the list contained tools that were developed for a wide number of audiences, such as Google Grow Tools. This tool, ranked as number 1, stands out as a versatile learning material with a significant focus on improving business operations. With a 45% DGE competence rating (denoting the percentage of DGE competence areas it covers), this tool is broadly applicable and particularly valuable for startups and small to medium-sized enterprises (SMEs) as an accessible and cost-effective choice for those looking to boost their digital capabilities sustainably. The InnoSchool Tool, ranked as number 2, with a 55% DGE competence rating, is another tool of broad application, designed to raise awareness about social needs and improve the abilities and entrepreneurial skills of its learners, helping to build the next generation of social entrepreneurs and innovators. Based on the unique combination of a Serious Game, lectures, mentoring and reflection, with very positive user feedback, it succeeds in accumulating social entrepreneurship knowledge and skills. Ranked as number 3, SmartCulTour Toolkit (53% DGE competence) is an innovative resource intended for use in Living Labs, local communities, and policy-making contexts where stakeholder engagement is needed. The toolkit supports inclusive and sustainable development, making it a key resource for those working to drive sustainable projects. The analysis uncovered a portfolio of tools that are relevant to a diverse audience, including educational institutions, businesses, policymakers, and individuals (all relevant stakeholders within the project), seeking to enhance their DGE competences. Varied formats – ranging from digital learning programs and practical toolkits to entrepreneurial competitions – cater to project-relevant needs and international contexts. Being accessible, effective, and innovative, this tailored portfolio supports the development of crucial competences needed to advance in the digital and green transition, one of the main aims of the actions foreseen by the project.

Another important contribution of this study is the integrated viewpoint on the DGE competences, corresponding to the noted importance of the potential of digital technologies in addressing environmental and social challenges [17]. Namely, the top-ranked tools are the ones that demonstrate how technologies can be leveraged to enhance sustainability and entrepreneurship education, addressing more competence areas and supporting the holistic approach advocated by [29]; this integration is crucial as it reflects the interconnectedness of digitalisation, sustainability, and entrepreneurship, which are essential for achieving the SDGs, aligning well with recommendations of [38] in a call for speedier integration of sustainability into entrepreneurship programs and [32], who suggest that combining GreenComp and EntreComp is essential for the transition from system thinking to action.

With regards to the use of the PROMETHEE II in ranking and evaluating the upskilling tools and programs, it proved effective for building a custom-built and relevant portfolio by providing a robust basis for decision-making based on both quantitative and qualitative criteria, ensuring a balanced and comprehensive assessment. This paper thus complements well studies (given systematically in [40] that evaluate similar educational tools and programs (using MCDM) that enhance skills measured via well-established and complex competence frameworks. [6] in particular emphasised the complexity of developing entrepreneurial competence measurement scales. On that note, this study validated the usefulness of three well-known competence frameworks, both in the phase of self-assessment of experts and as categories/criteria listing competence areas that the tools/programs contribute to. The practical application of competence frameworks such as DigComp, GreenComp, and EntreComp was highlighted, demonstrating their effectiveness in real-world settings, supporting the work of [39], [8], and [5] who developed these frameworks to guide competence development in DGE areas. Additionally, the study provides practical insights into the operationalisation of these frameworks, as called for by [22]. The multi-purpose nature of these frameworks is evident, however, as [14] noted, the willingness to adopt and implement these frameworks among key stakeholders is crucial for further development.

The practical implications of this study are significant for various stakeholders, including educational institutions, policymakers, and business support organisations. The ranked tools and programs provide actionable insights for designing and implementing upskilling initiatives, aligning with the recommendations of [18], who emphasise the need for higher resilience through enhanced competences. Individuals or organisations can adopt this portfolio of tools to enhance their curricula, policymakers can support the dissemination and adoption of these programs, and business support organisations can develop tailored training sessions to meet the specific needs of their clients. Also, the process outlined here in the paper can be followed to capture, curate and reuse tools/programs for other target groups and in different competence areas.

5. Conclusion

This study addresses the need for a holistic approach to sustainable entrepreneurship upskilling support. By using multiple competence frameworks in two distinct ways – as a competence self-assessment tool for experts engaged in MCDM exercise and as a mechanism to identify and categorise DGE upskilling tools/programs, the study contributes to a more comprehensive understanding of how to effectively upskill individuals in DGE areas. By ranking the 15 most useful DGE tools/programs for an international partnership with a diverse group of stakeholders, the study provides actionable insights into which tools are most effective in enhancing DGE competences. This ranking aids stakeholders in making informed decisions about which tools to adopt and promote, ultimately supporting capacity-building initiatives. PROMETHEE II was used as an effective tool for evaluating and ranking upskilling programs, ensuring also a comprehensive assessment of each tool's effectiveness. The practical application of competence frameworks such as DigComp, GreenComp, and EntreComp demonstrated their utility in real-world settings. The study highlights the potential for these frameworks to guide the development of effective upskilling programs that can be tailored to various educational and professional contexts.

There are several challenges and limitations to this study; in particular, the reliance on self-reported data from project partners may introduce bias, a concern echoed by [11] . Even if workshops and focus groups were organised to reach a common understanding and consensus, there is a need for objective and more critical perspectives in evaluating competence(s) areas. The subjective nature of qualitative assessments could affect the reliability of the rankings. The rapidly changing nature of digital and green technologies also means that the relevance of the identified tools and programs evolves over time, and while the GreenComp framework is rather novel and very flexible, as noted by [37], the DigComp is updated regularly, meaning the selected portfolio would need review accordingly.

While this study ranks tools and programs across various criteria, it does not explicitly assess their actual impact on learners' competence development, highlighting a need for future research to incorporate effectiveness data and post-implementation evaluations to provide a more comprehensive perspective.

Future research could focus on expanding the database of upskilling tools and programs to include more diverse geographical regions and industries, such as regions in Africa, Asia, and Latin America as foreseen by [3]. Additionally, there is a need to explore the impact of cultural differences on the adoption and effectiveness of these tools, whereby a study that would evaluate the performance of these tools with different global target groups can provide deeper insights into the relevance of this portfolio for upskilling in DGE competences.

Acknowledgements

This research has been supported by Interreg Central Europe project CE0100048 "Capacity2Transform".

References

[1] Alshamsi, A.M., El-Kassabi, H., Serhani, M.A. et al. (2023). A multi-criteria decision-making (MCDM) approach for data-driven distance learning recommendations. Educ Inf Technol, 28, 10421–10458 doi: 10.1007/s10639-023-11589-9

[2] Ayouni, S., Menzli, L.J., Hajjej, F., Maddeh, M,. and Al-Otaibi, S. (2021). Fuzzy vikor application for learning management systems evaluation in higher education. IJICTE, 17, pp. 17-35. doi: 10.4018/IJICTE.2021040102

[3] Baah-Acheamfuor, K., Qutieshat, A., and Yangailo, T. (2023). Critical Review: Measures of Digital Competency related to Employability. Economia Aziendale Online-, 14(2), 379-391. doi: 10.13132/2038-5498/14.2.379-391

[4] Babić, Z. (2017). *Models and methods of decision making in business.* Split: Faculty of Economics, Business and Tourism.

[5] Bacigalupo M., Kampylis P, Punie Y. and Van Den Brande L. (2016). EntreComp: The Entrepreneurship Competence Framework. EUR 27939 EN. Luxembourg (Luxembourg): Publications Office of the European Union. JRC101581 doi: 10.2791/593884

[6] Bacigalupo, M. (2022). The European Entrepreneurship Competence Framework (EntreComp). A conceptual model built and tested by the European Commission's joint research centre. Journal of Creative Industries and Cultural Studies-JOCIS, (4), 38-53. doi: 10.56140/jocis-v4-2

[7] Berniak-Woźny, J., Plebańska, M. and Wójcik-Jurkiewicz, M. (2023). University students' perception of employability and workability skills for the workplace in the digital era. Scientific Journal of Bielsko-Biala School of Finance and Law, 27(4), 39-45. doi: 10.19192/wsfip.sj4.2023.5

[8] Bianchi, G., Pisiotis, U. and Cabrera Giraldez, M. (2022). GreenComp The European sustainability competence framework, Punie, Y. and Bacigalupo, M. editor(s), EUR 30955 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-46485-3. doi: 10.2760/13286, JRC128040.

[9] Brinthaupt, T. M., Clayton, M. A., Calahan, P. T., Draude, B. J. (2014). How should I offer this course? The course delivery decision model (CDDM). Journal of Online Learning and Teaching, 10(2), 326.

[10] Can, Y., Aksoy, M. A., Aksoy, E., and Narli, S. (2022). Investigation of educational mathematics mobile applications (EMMAs) with multi-criteria decision-making methods: A TOPSIS algorithm implementation. Journal of Educational Technology and Online LearningVol. 5(4), pp. 1203–1218. doi: 10.31681/jetol.1156464

[11] Corres, A., Rieckmann, M., Espasa, A., and Ruiz-Mallén, I. (2020). Educator competences in sustainability education: A systematic review of frameworks. Sustainability, 12(23), 9858. doi: 10.3390/su12239858

[12] Fabregá, M. B., Masferrer, N., Patau, J., and Miró Pérez, A. P. (2020). Self-counciousness competence as driver of innovation and environmental commitment in higher education students. International Journal of Sustainability in Higher Education, 21(7), 1507-1523. doi: 10.1108/IJSHE-03-2020-0083

[13] Fitsilis, P. (2024). Navigating the Skills Revolution: The Essential Role of Competence Frameworks. Qeios. doi: 10.32388/V28REV

[14] Fleaca, B., Fleaca, E., and Maiduc, S. (2022). Digital Transformation and Current Challenges of Higher Education. TEM Journal, 11(3). doi: 10.18421/TEM113-32

[15] George, G., and Schillebeeckx, S. J. (2022). Digital transformation, sustainability, and purpose in the multinational enterprise. Journal of World Business, 57(3), 101326. doi: 10.1016/j.jwb.2022.101326

[16] Hamdan, M., and Anshari, M. (2024). Green Education to Promote Green Technological Skills Asfand Yar Universiti Brunei Darussalam, Brunei. Harnessing Green and Circular Skills for Digital Transformation, 72. doi: 10.4018/979-8-3693-2865-1.ch005

[17] Holzmann, P., and Gregori, P. (2023). The promise of digital technologies for sustainable entrepreneurship: A systematic literature review and research agenda. International Journal of Information Management, 68, 102593. doi: 10.1016/j.ijinfomgt.2022.102593

[18] Isensee, C., Teuteberg, F., and Griese, K. M. (2023). Success factors of organizational resilience: a qualitative investigation of four types of sustainable digital entrepreneurs. Management Decision, 61(5), 1244-1273. doi: 10.1108/MD-03-2022-0326

[19] Ishizaka, A., and Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software.* Wiley. doi: 10.1002/9781118644898

[20] Ishizaka, A., and Resce, G. (2021). Best-Worst PROMETHEE method for evaluating school performance in the OECD's PISA project. Socio-Economic Planning Sciences, 73, 100799. doi: 10.1016/j.seps.2020.100799

[21] Islas-Pérez, E., Pérez, Y. H., Pérez-Ramírez, M., García-Hernández, C. F., and Zayas-Pérez, B. (2015). Multicriteria Decision Making for Evaluation of e-Learning Tools. Res. Comput. Sci., 106, 27-37. doi: 10.13053/RCS-106-1-3

[22] Joensuu-Salo, S., Viljamaa, A., and Varamäki, E. (2022). Testing the EntreComp framework and its relation to start-up behaviour in seven European countries. Journal of Small Business and Enterprise Development, 29(6), 920-939. doi: 10.1108/JSBED-04-2021-0156

[23] Kostelic, K. (2025). TOPSIS-based framework for evaluating employee cybersecurity risk. Croatian Operational Research Review, 16(1), 31-44. doi: 10.17535/crorr.2025.0003

[24] Kovač, D., and Podrug, D. (2023). Improving portfolio liquidity: MCDM approach to share selection on the Zagreb Stock Exchange. Croatian Operational Research Review, 14(1), 29-39. doi: 10.17535/crorr.2023.0003

[25] Leal Filho, W., Raath, S., Lazzarini, B., Vargas, V. R., de Souza, L., Anholon, R., Quelhas, O.L.G., Klavins, V.L., and Orlovic, V. L. (2018). The role of transformation in learning and education for sustainability. Journal of cleaner production, 199, 286-295. doi: 10.1016/j.jclepro.2018.07.017

[26] Liang, D., Fu, Y., and Garg, H. (2024). A novel robustness PROMETHEE method by learning interactive criteria and historical information for blockchain technology-enhanced supplier selection. Expert Systems with Applications, 235. doi: 10.1016/j.eswa.2023.121107

[27] Lopes, A. P. F. Muñoz, M. M., and Alarcón-Urbistondo, P. (2018). Regional tourism competitiveness using the PROMETHEE approach. Annals of Tourism Research, 73, 2018, 1-13. doi: 10.1016/j.annals.2018.07.003

[28] Mijač, T., Ninčević Pašalić, I., and Tomat, L. (2021). Selection of IoT platforms in smart cities: multicriteria decision making. In: *Proceedings of the 16th International Symposium on Operational Research in Slovenia*, SOR 2021, pp. 35–40. url: http://fgg-web.fgg.uni-lj.si/ /sdrobne/sor/SOR'21%20-%20Proceedings.pdf [Accessed 23/11/2024]

[29] Modgil, S., Dwivedi, Y. K., Rana, N. P., Gupta, S., and Kamble, S. (2022). Has Covid-19 accelerated opportunities for digital entrepreneurship? An Indian perspective. Technological Forecasting and Social Change, 175, 121415. doi: 10.1016/j.techfore.2021.121415

[30] Mondal, S., Singh, S., and Gupta, H. (2023). Green entrepreneurship and digitalization enabling the circular economy through sustainable waste management-An exploratory study of emerging economy. Journal of Cleaner Production, 422, 138433. doi: 10.1016/j.jclepro.2023.138433

[31] Pinochet, L., H. C., Lellis Moreira, M. A., Fávero, L. P., dos Santos, M., and Itacaramby Pardim, V. (2023). Collaborative Work Alternatives with ChatGPT Based on Evaluation Criteria for its Use in Higher Education: Application of the PROMETHEE-SAPEVO-M1 Method. Procedia Computer Science, 221, 177-184. doi: 10.1016/j.procs.2023.07.025

[32] Planck, S., Wilhelm, S., Kobilke, J., and Sailer, K. (2024). Greater than the Sum of Its Parts: Combining Entrepreneurial and Sustainable Competencies in Entrepreneurship Education. Sustainability, 16(9), 3725. doi: 10.3390/su16093725

[33] Rehman, S. U., Giordino, D., Zhang, Q., and Alam, G. M. (2023). Twin transitions and industry 4.0: Unpacking the relationship between digital and green factors to determine green competitive advantage. Technology in Society, 73, 102227. doi: 10.1016/j.techsoc.2023.102227

[34] Schneider, M., Strzalka D., Kahlen, J., and Lichte, D. (2023). Optimizing Observation Strategy in Emergency Response by Combining Bayesian Network and Multi-Criteria Decision Analysis. 33rd European Safety and Reliability Conference. doi: 10.3850/978-981-18-8071-1_P351-cd

[35] Seikkula-Leino, J., Salomaa, M., Jónsdóttir, S. R., McCallum, E., and Israel, H. (2021). EU policies

driving entrepreneurial competences—Reflections from the case of EntreComp. Sustainability, 13(15), 8178. doi: 10.3390/su13158178

[36] Singh, R., Kumar, V., Singh, S., Dwivedi, A., and Kumar, S. (2024). Measuring the impact of digital entrepreneurship training on entrepreneurial intention: the mediating role of entrepreneurial competencies. Journal of Work-Applied Management, 16(1), 142-163. doi: 10.1108/JWAM-11-2022-0076

[37] Sourgiadaki, M., and Karkalakos, S. (2023). "GreenComp" as a tool for examining motivation of vocational teachers to create learning opportunities for the green transition. SN Social Sciences, 3(7), 114. doi: 10.1007/s43545-023-00699-3

[38] Uvarova, I., Mavlutova, I., and Atstaja, D. (2021). Development of the green entrepreneurial mindset through modern entrepreneurship education. In *IOP Conference Series: Earth and Environmental Science*, 628(1), p. 012034. IOP Publishing. doi: 10.1088/1755-1315/628/1/012034

[39] Vuorikari, R., Kluzer, S. and Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes, EUR 31006 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-48883-5, JRC128415. doi: 10.2760/490274

[40] Yüksel, F. Ş., Kayadelen, A. N., and Antmen, F. (2023). A systematic literature review on multi-criteria decision making in higher education. In International Journal of Assessment Tools in Education (Vol. 10, Issue 1, pp. 12–28). International Journal of Assessment Tools in Education. doi: 10.21449/ijate.1104005

[41] Ziemba, P., Piwowarski, M., and Nermend, K. (2023). Software systems supporting remote education – Fuzzy assessment using a multi-criteria group decision-making method. Applied Soft Computing, 149, Part A, 110971. doi: 10.1016/j.asoc.2023.110971

Going concern prediction – A horse race between traditional and regularization machine learning models

Tina Vuko[1,*], **Slavko Šodan**[2] and **Ivana Perica**[2]

[1] *Faculty of Economics, Business and Tourism, University of Split, Cvite Fiskovića 5, 21000 Split, Croatia*
E-mail: ⟨tina.vuko@efst.hr⟩

[2] *Faculty of Economics, Business and Tourism, University of Split, Cvite Fiskovića 5, 21000 Split, Croatia*
E-mail: ⟨{ssodan, iperica}@efst.hr⟩

Abstract. Regularization machine learning (ML) methods have been increasingly applied in accounting research, offering new possibilities in predictive modeling. Their forte lies in the effective regularization methods for resolving the biggest concern of generalization, which is the risk of overfitting the training data. While these sophisticated methods are known to outperform traditional regression approaches in large and balanced datasets, this may not be the case when facing imbalanced and small datasets. Moreover, model validation is also challenging in such settings because traditional performance measures, such as prediction accuracy, may be misleading. We address this problem by comparing two traditional and five regularization-based methods in predicting going concern uncertainty (GCU) on the sample of listed companies in Croatia. We take caution when evaluating the models due to class-imbalanced problems and include different classification performance measures, as well as calibration of the models to account for their uncertainty. As expected, no model performs best across all evaluation criteria, but regularization methods are better calibrated. Given our results, we suggest that model selection should consider the results of the model calibration, a combination of different performance metrics, and the economic impact of the statistical performance of the model, if feasible.

Keywords: elastic net, lasso, ridge regression, class imbalance, prediction

Received: October 31, 2024; accepted: December 17, 2024; available online: April 2, 2025

DOI: 10.17535/crorr.2025.0014

Original scientific paper.

1. Introduction

Going concern (GC) is the key assumption in financial reporting, which assumes that management has no intention nor will be forced to terminate operations and liquidate a company's assets at least within the next 12 months. However, when there is substantial uncertainty about a company's ability to continue as a GC, an auditor should emphasize this in the audit report. Therefore, GC emphasis in the auditor report is the result of an auditor's prediction of a GCU for a given company, taking into account all available information for the near future. The auditor's GC emphasis should be understood as an early warning sign of the potential inability of the company to continue as a GC and not as an absolute and ongoing assurance of

*Corresponding author.

the company's ability to withstand future shocks. Since GCU is a complex and highly judgmental decision with widespread consequences [8], it is interesting from the aspect of predictive modeling.

One of the major problems when predicting GCUs (and many other interesting accounting events, such as fraud, accounting restatement, or bankruptcy) is that they occur infrequently. Consequently, researchers face rather imbalanced datasets and relatively small sample sizes compared to the number of potential predictors. Building a predictive model in such circumstances can be a significant challenge both from the aspect of selecting important variables and from the aspect of model evaluation and selection (i.e. choosing the final model among a set of models). Over the years, logistic regression (LR) has become the most widely used method to predict discrete outcomes in accounting research [20]. Likewise, LR is the most frequently used statistical method to predict auditors' GCU decisions [5], [21], [8]. While LR has many attractive properties, a potential limitation from the aspect of predictive modeling is related to variable selection. Namely, traditional variable selection methods like stepwise selection may lead to overfitting. Also, due to pre-testing bias, they may overlook some important but insignificant predictors, leading to the loss of predictive information [16]. This limitation is strongly emphasized when dealing with small and imbalanced datasets and when it is not apparent whether a variable should be included in the model or not.

Recent advancements in statistical and ML algorithms offer new opportunities for variable or feature selection and prediction problems. Shrinkage methods such as *lasso* (the least absolute shrinkage and selection operator) are known to address some of these challenges through regularization techniques. The attractiveness of shrinkage methods in predictive learning models is majorly based on the bias-variance tradeoff, following the general idea that prediction accuracy can be improved by shrinking the values of the regression coefficients or setting some coefficients to zero [16], [17]. However, the shrinkage does not necessarily lead to better predictive performance, particularly with small sample sizes. Moreover, when the sample size is small, it may be even advisable not to build a prediction model [28]. Given these concerns, we question if sophisticated statistical and ML algorithms that have been increasingly applied in accounting research [6], [14], [20] can outperform traditional statistical methods in predictive modeling of rare classes.

To answer this question, we illustratively use auditors' GCU predictions from a sample of companies listed in Croatia over ten years. Our dataset can be described as small and moderately imbalanced. In order to build competing models, we define 25 plausible covariates. We focus on predictors that can be easily calculated from publicly available financial statements and are commonly used to evaluate a GCU (see for example Cyhe Koh et al., 2004; Martens et al., 2008; Yeh et al., 2014; Goo et al., 2016 studies [9], [21], [31], [15]). In addition to financial indicators, we use three audit-related indicators: type of auditor (Big Four auditors or others), type of audit opinion (unmodified audit opinion or modified audit opinion), and change of auditor (a company has changed auditor or not). Since they do not apply any regularization, we use conventional LR with stepwise variable selection (*LRsw*) and a full set of predictors (*LRfull*) as the reference methods. Finally, we compare the predictive performance of standard LR models with five regularization-based models (*lasso, adaptive lasso, plugin lasso, ridge, and elastic net*) to assess the predictive performance of the models. Given that our sample size is relatively small and that some covariates are highly correlated, there is a substantial chance of overfitting.

An additional problem arises when evaluating models with unequally distributed dichotomous outcomes. Commonly used model evaluation metrics such as prediction accuracy, an overall error rate, or a proportion of correctly (miss)classified samples might be misleading as they favor classifiers that accurately predict the majority class [7]. On the other hand, class balancing by applying different sampling techniques to overcome imbalanced data structure is problematic because it makes the number of minority and majority classes predetermined and not random [7]. To overcome this problem, we use a wide range of classification metrics and calibration belts to evaluate the competing models' predictive power comprehensively.

In a way, our research extends Bertomeu's (2020) and Krupa and Minutti-Mezza's (2022) studies ([6], [20]), which provide guidance on conducting accounting-related research involving models that predict discrete outcomes using traditional and ML algorithms. However, we focus on methodological challenges related to prediction modeling and valuation when facing imbalanced and small datasets. Our results indicate that while nowadays it is fashionable to construct different models using sophisticated statistical and ML methods to predict important accounting outcomes, more effort should be put into discussing how and why such models work or what underlying issues they address [29]. Different research design choices (variables selection, splitting methods, sampling strategies, model evaluation metrics, etc.) may obscure the comparison between the alternative prediction models. For example, when evaluating GCU predictive models, researchers usually report only a few classification measures, mainly prediction accuracy, ROC area, Type I and Type II errors, precision and recall rates [9], [21], [31], [15], [18], [23], [24]. Krupa and Minutti-Mezza [20] report similar findings for other mainstream accounting research on predictive modeling. In addition, studies that predict dichotomous accounting events rarely (if any) assess the calibration of the outcome models. We find this important as our results indicate that when predicting infrequent outcomes, no model performs best across all evaluation criteria, or moreover, several models perform about equally well using different algorithms and selecting different variables. Consequently, we may easily select the poorly calibrated model even when considering class-imbalance-adjusted performance measures.

The paper proceeds as follows. Section 2 compares traditional classification methods vs. regularization-based ML methods. Section 3 describes research procedures and the sample characteristics, while section 4 presents research findings. Section 5 concludes with a discussion of the results, limitations of our approach, and potential implications for future research.

2. Traditional vs. regularization methods

2.1. Variable selection

Over the years, LR has become a popular prediction model in accounting because it produces a linear combination of the variables with weights and confidence intervals for the weights that clearly show how the predictors affect the outcome variable. However, when the sample size is small or imbalanced, traditional classification methods could produce overfitted and unstable models, making model selection problematic. Namely, a stepwise regression approach to variable selection suffers from overfitting and pretesting biases [1], [16].

Therefore, regularization has become the cornerstone of modern statistics. The major benefit of regularization or shrinkage methods is that they can accommodate large predictor models (usually under the assumption of sparsity) and rely on tuning parameters. Regularization methods solve an optimization problem using two terms: fit measure (e.g. deviance ratio as a value of loss function) and penalty term (regularization parameter), also called tuning parameter. The term is included to penalize the complexity of the model. The lasso algorithm penalizes the absolute size of coefficients (L1 or lasso regularization), the ridge penalizes the sum of squared coefficients (L1 or ridge regularization), while the elastic net applies a mix of lasso and ridge-type penalties [1]. Therefore, contrary to the lasso which yields sparse solutions by setting some coefficient estimates precisely equal to zero, the ridge method preserves all predictors in the model [2]. Lasso commonly selects penalty term through cross-validation (CV), the adaptive and plugin methods. Adaptive lasso uses adaptive penalty weights for the lasso penalty term (L1) to achieve the oracle properties of the estimators [32]. Plugin lasso iteratively calculates the smallest penalty term that is large enough to dominate the estimation error in the coefficients. While plugin includes covariates in the model that best approximates the data, it risks missing some covariates with large coefficients and finding only some with small coefficients [3], [4]. Since the estimates' magnitudes are somewhat restricted, the regularized regression estimator will tend to have a smaller variance than traditional regression methods. Consequently, they may deliver an overall more accurate prediction, particularly in the presence of small samples [27]. Regularization methods also include nonlinear terms and interactions among the predictors without explicitly specifying them [20]. Furthermore, due to their low variance, they are immune to multicollinearity and robust to high-dimensional settings (i.e. the course of dimensionality).

On the other hand, regularization methods trade-off between estimator bias and variance so that they may underestimate, to some extent, the population parameters and fail to capture essential regularities in the training data. Another problem could arise if regularization is applied to domains in which sparsity is not a plausible assumption, and this is particularly questionable when the sample size is small relative to the number of the model's parameters [13]. While ridge regression shrinks the coefficients of correlated predictors towards each other, lasso tends to select one of them somewhat arbitrarily, so one may erroneously conclude that the selected covariate is essential, even though other more important covariate may be left out.

The recent success of some overparametrized models (e.g. neural networks) also indicated that in some cases, very complex models predict better [11] [13] [30]. However, such complex models are hard to interpret and tend to be unstable, so simple, easily interpretable models like LR might often perform just as efficiently as complex ML models [13]. Most importantly, when using regularization methods to develop prediction models on small samples or when having a larger number of potential predictors relative to the number of minority classes, researchers advise more caution and call for more research investigating the impact of specific combinations of shrinkage and tuning methods [26], [28].

2.2. Evaluation of the prediction models' performance

Successful ML training approaches often rely on sufficiently large and balanced data. However, many important real-world events generate only imbalanced, small data sets and pose problems for both traditional and sophisticated statistical or ML algorithms. Commonly used model evaluation criteria such as accuracy, ROC, or ROC AUC metrics can be misleading in this case since statistical and ML models are generally designed around the assumption of balanced class distribution. The high accuracy of such models might reveal more about the underlying distribution of classes than about actual model performance. In a situation like GCU prediction, where false negatives (FN) incur greater cost than false positives (FP), imbalance may lead to adverse consequences. Nevertheless, adding a specific cost-sensitive model to the training data may induce bias in the model if the true error cost of the minority class differs from that of the training data [29]. Also, frequently used sampling techniques (matched sample designs, oversampling the minority class or undersampling the majority class) on training samples are not a preferable solution to the class imbalance problem as they alter the relationship between majority and minority classes, which may affect the incidence [7], [21]. Therefore, the critical issue when evaluating alternative prediction models is the choice of an appropriate performance measure. However, finding the most appropriate and meaningful evaluation metrics for imbalanced data is not achievable without having accurate cost information which could utilize cost-sensitive learning to produce an accurate classifier [29].

To overcome this problem, we use different discrimination and classification metrics to select the model that yields good results over a wide variety of assumptions. In addition to the ROC curve that is commonly used to evaluate the performance of prediction models, we use the precision-recall curve that is better suited for imbalanced datasets [10], as well as different metrics derived from the confusion matrix. Using the confusion matrix, we calculate typical performance measures (accuracy, recall or sensitivity, specificity, precision or positive predicted values and negative predicted value) and several imbalance-adjusted measures (Balanced accuracy, F-measure, Geometric mean and Matthew's Correlation Coefficient). However, even when the model has good discrimination, the estimated risks can be unreliable if the model is poorly calibrated [28]. Therefore, we use calibration belts to examine the relationship between estimated probabilities and observed outcome rates.

3. Procedure and sample description

To address our research question, we use hand-collected data from financial statements and related audit reports of Croatian listed companies, excluding companies in the financial sector. The sample period covers ten years, from 2009 to 2018. After controlling for the missing data, the initial sample of 929 company-year observations is reduced to 891 company-year observations. For this analysis, we use all years affected by the GCU as our positive classes, so the proportion of positive classes (GCUs) is 18.52%. We define 22 financial and three audit-related variables as possible and commonly used GCU predictors (see Appendix A for variable description and descriptive statistics, available at: Appendix.pdf). We use Stata 18 software for our analyses.

Since our main objective is to compare the performance of traditional LR and regularization prediction models when classes are not equally represented in the dataset, we do not use a pre-specified model nor focus on any specific variable as a potential determinant of GCU by trying to answer whether it reflects an independent mechanism of the outcome. Rather, we focus on assessing the performance of example models when variable selection is used to specify a model using a variety of classification measures and calibrations. Namely, we apply different types of regularization (*lasso*, *adaptive lasso*, *plugin lasso*, *elastic net*, and *ridge regression*) to develop prediction models. We use penalized coefficients estimated on the training dataset to make predictions for regularization methods and unpenalized coefficients estimated on the training dataset to make predictions for traditional LR models. Traditional LR models are developed using stepwise selection (*LRsw*) and a full set of predictors (*LRfull*) to understand better the risk of overfitting on model performance. Because there is a certain trade-off between using fewer data to train (i.e. parameters will have greater variance) and using fewer data to test models (i.e. performance statistics will have greater variance), researchers can use different sample split strategies (for example see Kalinic Milicevic and Marasovic [19]). Therefore, we first assign 2/3 of the sample to the training and 1/3 to the test sample (67:33% sample split). Alternatively, we use an 80:20% sample split as a robustness check. We also set the random-number seed option so that we can reproduce our results.

In the case of ML algorithms, the training sample also uses sample split referred to as the "k-fold cross-validation" where the dataset is divided into k equally sized subsets. One subset is chosen to be the validation set and the remaining k-1 folds are used to train the model. The validation set is then evaluated with a performance metric such as the deviance ratio. This method is repeated k times so that each subset acts as a validation set exactly once. The performance metric is then averaged across all k iterations to give an estimated performance for the model [12]. Table 1 presents the list of predictors selected by shrinkage methods.

As expected, the results from Table 1 show that the plugin produces the most parsimonious model (only four predictors). The most important variables selected by the *plugin* are profit/loss indicator (*x17*), working capital ratio (*x04*), return on equity (*x15*), and leverage ratio (*x05*). Lasso selected ten predictors; *adaptive* lasso selected nine, and *elastic net* selected 13 predictors, which is the same number of predictors selected by the stepwise method (*LRsw*) (untabulated).

The results of an alternative sample split can be found in Appendix B (available at: Appendix.pdf).

Variable	lasso	adaptive	plugin	elastic net	ridge
cons	x	x	x	x	x
x17	x	x	x	x	x
x04	x	x	x	x	x
x24	x	x		x	x
x15	x	x	x	x	x
x12	x	x		x	x
x05	x	x	x	x	x
x21	x	x		x	x
x18	x	x		x	x
x16	x	x		x	x
x10	x			x	x
x14				x	x
x13				x	x
x19				x	x
x20					x
x07					x
x23					x
x02					x
x11					x
x01					x
x25					x
x03					x
x09					x
x06					x
x22					x
x08					x

Table 1: Standardized coefficients sorted

4. Results

We evaluate the performance of the two traditional LR and five regularization-based GCU prediction models on the hold-out sample using several metrics. We start our analysis by comparing deviance (D) and deviance ratio (D^2) goodness-of-fit statistics, principally used for generalized linear models (GLM). Smaller D and larger $D2$ indicate a better model (Table 2).

When looking at the test sample results, D^2 shows that ridge regression performs as the best model and as the most robust model relative to the alternative sample split strategy. We can notice that a simpler model with stepwise selection (*LRsw*) performs better on the test sample data than the model without variable selection (*LRfull*).We can also notice that traditional LR models generally perform better than regularization models on the training sample. In contrast, regularization models perform as effectively or slightly better than traditional LR models on the test samples. This is expected since traditional estimation methods are based on optimizing the estimated model's in-sample fit without any regularization to optimize the out-of-sample fit. However, plugin lasso performs as the worst model. As already explained, the plugin is a rigorous lasso that tends to favor very parsimonious models by selecting the covariates that best approximate the data but also runs the risk of missing some important covariates.

Given that in the test sample deviance ratio is the same or similar for alternative models (e.g. *LRsw* and *adaptive* lasso or *lasso* and *elastic net*), we can conclude that the importance of selected predictors can vary between the models having about the same D^2.
While we can opt for a simpler prediction model in this case, the question remains whether

Model	Sample	Deviance	Deviance ratio
LRSW	1	0.639	0.340
LRFULL	1	0.614	0.365
lasso	1	0.693	0.284
adaptive	1	0.656	0.322
plugin	1	0.826	0.147
elastic net	1	0.696	0.281
ridge	1	0.715	0.261
LRSW	2	0.712	0.241
LRFULL	2	0.747	0.204
lasso	2	0.707	0.247
adaptive	2	0.712	0.241
plugin	2	0.795	0.153
elastic net	2	0.706	0.247
ridge	2	0.702	**0.251**

Table 2: Goodness of fit (Deviance ratio)

Notes: *Sample 1 is a training sample n=594 and Sample 2 is a test sample n=297. The highest D^2 values in the test sample are in bold.*

those models are equally good in prediction accuracy and stability.

In order to evaluate the classification performance of alternative models, we use results obtained from a confusion matrix. Confusion matrix commonly uses a 50% decision threshold so that all values equal or greater than the threshold are assigned to one class and all other values to another. Using the information provided by the confusion matrix, we calculate some commonly used and several imbalance-adjusted classification metrics (Appendix C provides a detailed description of each measure used, available at: Appendix.pdf).

Evaluation metrics	LRSW	LRFULL	Lasso	Adaptive	Plugin	Elastic Net	Ridge
Accuracy (%)	**87.2**	86.5	86.2	**87.2**	82.5	86.2	86.2
Recall (%)	43.4	**47.2**	30.2	45.3	3.8	30.2	28.3
Specificity (%)	96.7	95.1	98.4	96.3	**99.6**	98.4	98.8
Precision (%)	74.2	67.6	80.0	72.7	66.7	80.0	**83.3**
NPV (%)	88.7	**89.2**	86.6	89.0	82.7	86.6	86.4
F-measure	0.55	**0.56**	0.44	**0.56**	0.07	0.44	0.42
F-adjusted	**0.65**	0.62	0.60	**0.65**	0.15	0.60	0.60
BA	0.70	**0.71**	0.64	**0.71**	0.52	0.64	0.64
MCC	0.50	0.49	0.44	**0.51**	0.13	0.44	0.43
GM	0.65	**0.67**	0.54	**0.66**	0.19	0.54	0.53

Table 3: Confusion matrix (test sample results)

Notes: *The highest values of each evaluation metric are in bold.*

Table 3 shows that all models achieve overall accuracy above 82.5 % (the bottom value is for the *plugin*) but not higher than 87.2% (*LRsw* and *adaptive*). Although no model outperforms other models across all metrics when looking at the overall number of the highest values of specified evaluation metrics, *LRfull*, and *adaptive* lasso seem to perform as the best models. All classifiers perform quite similarly concerning accuracy, specificity, and negative predicted value (NPV).

The difference between classifiers is more pronounced when observing precision or positive predicted value (PPV) and recall or true predicted value (TPR). Those metrics are commonly used to improve the process of model evaluation when dealing with imbalanced data. However, there is an inevitable trade-off between those two measures. As the ability of a model to find all GCU classes increases, the ability of a model to identify only GCUs decreases. We can notice that TPRs are relatively small for all models, never greater than 50%, and extremely low for plugin lasso (3.8%). This also means that false negative rates (FNRs) or type II errors are high (FNR=1-Recall). While traditional *LRfull* (47.2 %) and *adaptive* lasso (45.3%) have the highest recall, *ridge regression* has the greatest precision (83.7%).

However, when looking at selected imbalance-adjusted metrics (F-measure, Balanced accuracy, Geometric mean and Matthew's Correlation Coefficient), adaptive lasso performs slightly better than *LRfull*. Therefore, based on calculated classification metrics, regularization models have no clear advantage over traditional LR models.

Beside classification metrics based on nominal class prediction using a given threshold (in our case, standard 50%) threshold), we use two probabilistic classification metrics (ROC and precision-recall curve) that plot values for all possible thresholds. ROC analysis and AUC ROC are most commonly used to evaluate predictive performance [10]. ROC visualizes a nonlinear trade-off between TPR (recall or sensitivity) and FPR (1-specificity) values, and AUC summarizes this information into a single number, which facilitates model comparison when there is not a dominating ROC [30]. Table 4 shows the results of the ROC/AUC analysis for the test sample.

Model	**AUC**	**SE**	**95%-CI Lower**	**95%-CI Upper**
LRSW	0.8370	0.0317	0.7749	0.8991
LRFULL	0.8378	0.0314	0.7763	0.8994
lasso	0.8337	0.0314	0.7722	0.8952
adaptive	**0.8422**	0.0309	0.7816	0.9028
plugin	0.8166	0.0326	0.7527	0.8804
elastic net	0.8349	0.0313	0.7735	0.8963
ridge	0.8398	0.0326	0.7760	0.9037
Ha: At least one classifier has a different AUC value $\chi^2 = 4.60$, $p = 0.5961$				

Table 4: Area under the ROC

Notes: *The highest values of AUC ROC are in bold.*

The results again favor the *adaptive* lasso. *Adaptive* lasso has the highest AUC value (84%), while *plugin* lasso, again, has the lowest value (82%). However, the difference between the AUC values of the models in the test sample is not statistically significant (Table 4).

While AUC ROC is a popular metric, it can be inflated in the presence of class skew. Therefore, we compare the models' area under the precision-recall curve. In addition, this metric may be preferable if we are more concerned about the number of false negatives [10], which is the case for GCU prediction. Table 5 shows the values of precision-recall (PR) AUC. The results show that PR AUC values in test samples for different classifiers and under different sample split strategies (see Appendix B, available at: Appendix.pdf) are approximately the same. Again, there is no straightforward evidence that simpler or penalized models perform noticeably better than traditional unpenalized and overparametrized models.

Finally, we use calibration belts to examine the relationship between test sample estimated probabilities and observed GCU rates. Calibration gives insight into model uncertainty by adjusting the probability distribution to better match the expected distribution observed in the

Model	PR AUC
LRsw	0.5988
LRFULL	0.5805
lasso	0.5771
adaptive	0.5932
plugin	0.5623
elastic net	0.5789
ridge	**0.6014**

Table 5: Area under the Precision-recall curve

Notes: *The highest values of AUC ROC are in bold.*

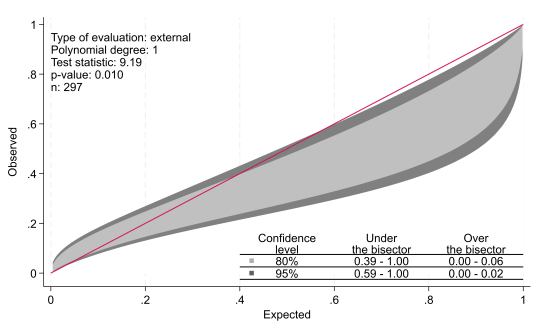


Figure 1: LRsw

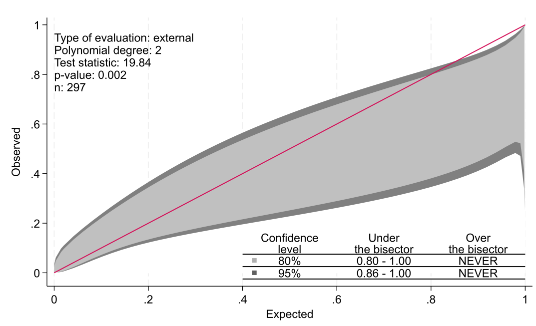


Figure 2: LRfull

data. Calibrated predictions imply that among those observations assigned a predicted rate of, for example, 0.10 for a GCU, the actual GCU does occur at a 10% rate. We present calibration belts using “calibrationbelt” command in Stata, which implements the calibration belt and its associated test (see: Nattino et al., 2017 [22]). Figures from 1 to 7 show calibration belts for estimated models on the test sample and corresponding test statistics of the deviations from the line of perfect calibration.

As can be noticed from Figures 1-7, only a few models (*lasso, elastic net and ridge regression*) showed satisfactory external calibration in the test sample. Because they are well-calibrated, the predictions of those models have greater economic significance.

Overall, our results indicate that, without considering calibration results, we would not see a clear cut between the traditional and sophisticated regularization models' performance and might choose an unstable predictive model. For example, while classification metrics favor adaptive lasso and LRfull GCU models, even after taking into account imbalance-adjusted

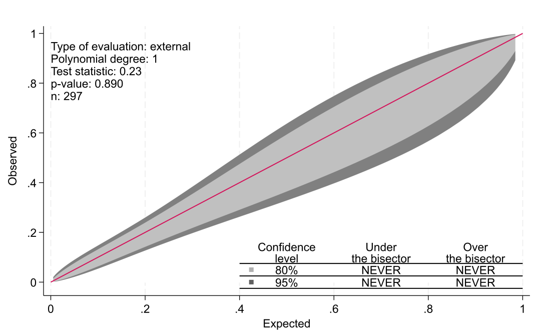


Figure 3: LASSO CV

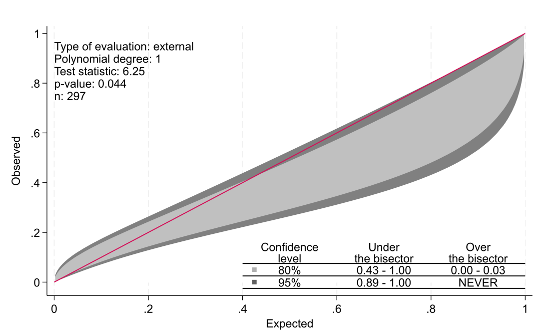


Figure 4: ADAPTIVE LASSO

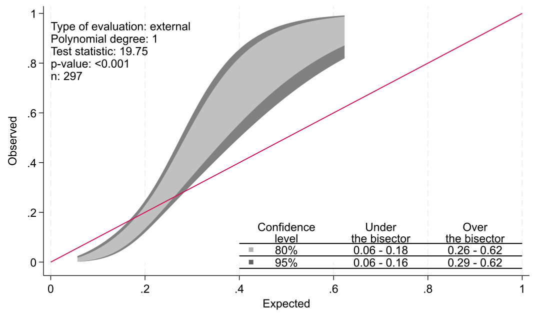


Figure 5: PLUGIN LASSO

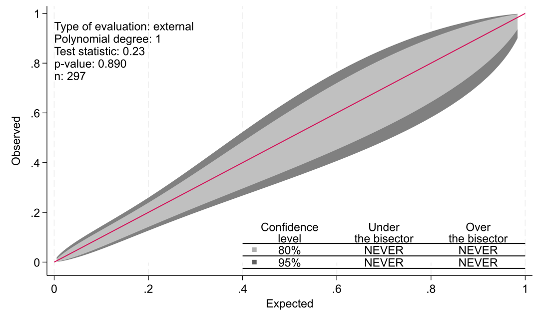


Figure 6: ELASTIC NET

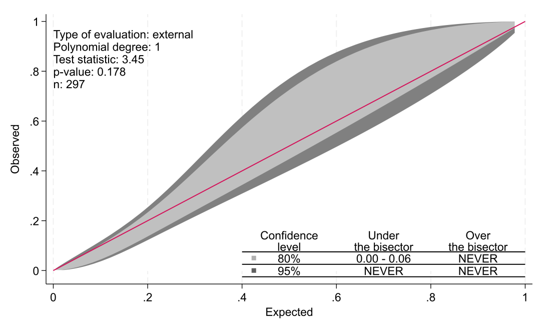


Figure 7: RIDGE REGRESSION

classification metrics, they may significantly underestimate small risks and/or overestimate high risks. In this regard, calibration aims to prevent predicted risks from being too extreme and makes the predictive model more relevant. Finally, we used an alternative sample split (80:20%) to train the models with more data. Our results are robust to alternative sample split decisions (see Appendix B, available at:Appendix.pdf).

5. Summary and Conclusion

Although research on different default prediction models has been around for quite a long time, new technologies and big data have opened the door for the implementation of more advanced and powerful prediction models. However, choosing the proper risk prediction model based on imbalanced datasets is rather challenging, even for sophisticated ML methods. Such datasets can cause traditional, as well as sophisticated classification algorithms to have a biased decision boundary. Our results show that the traditional, unregularized, and likely overfitted models (*LRsw* and *LRfull*) may perform as well as sophisticated regularization models (*lasso, adaptive lasso, plugin lasso, elastic net and ridge regression*) in terms of their classification properties when facing imbalanced datasets. However, we find that the models with the highest performance accuracy (*adaptive* and *LRfull*) are not well-calibrated by default, therefore lacking sufficient confidence in their GCU predictions. This implies that predictive model evaluation should be carried out carefully, preferably over the range of classification and calibration metrics and, if possible, taking into account the costs of prediction errors. While previous studies on GCU prediction have investigated a wide variety of financial and non-financial predictors using different approaches to variable selection, different sample splitting and selection methodologies, different model evaluation criteria, little or no attention, to the authors' best knowledge, has been given on assessing models' uncertainty. Our results show that the final decision in model selection should consider the model calibration results and a combination of different classification performance metrics. Further efforts could be made to explore potential heterogeneity and more profoundly investigate the generalizability of model performance. For example, instead of focusing on undersampling, oversampling or different sample matching techniques, researchers using sample sizes that are large enough (which is our limitation) could instead focus on potentially interesting subdomains (e.g. companies with indications of poor financial performance) without trying to remove imbalance via sampling methods artificially.

Acknowledgements

This work has been fully supported by Croatian Science Foundation under the project IP-2020-02-9372 "Disentangling Financial Reporting Quality".

References

[1] Ahrens, A., Hansen, C. B. and Schaffer, M. E. (2020). lassopack: Model selection and prediction with regularized regression in Stata. The Stata Journal, 20 (1), 176 - 235. doi: 10.1177/1536867X20909697

[2] Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11, 685 - 725. doi: 10.1146/annurev-economics-080217-053433

[3] Belloni, A., Chen, D., Chernozhukov, V. and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica, 80(6), 2369 - 2429. doi: 10.48550/arXiv.1010.4345

[4] Belloni, A., Chernozhukov, V. and Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. Journal of Business and Economic Statistics, 34 (4), 606 - 619. doi: 10.1080/07350015.2016.1166116

[5] Bellovary, J. L., Giacomino, D. E. and Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. Journal of Financial Education, 1 - 42.

[6] Bertomeu, J. (2020). Machine learning improves accounting: discussion, implementation and research opportunities. Review of Accounting Studies, 25(3), 1135 - 1155. doi: 10.1007/s11142-020-09554-9

[7] Blagus, R. and Goeman, J. J. (2018). What (not) to expect when classifying rare events. Briefings in Bioinformatics, 19 (2),. 341 - 349. doi: 10.1093/bib/bbw107

[8] Carson, E., Fargher, N. L., Geiger, M. A., Lennox, C. S., Raghunandan, K. and Willekens, M. (2013). Audit reporting for going-concern uncertainty: A research synthesis. Auditing: A Journal of Practice and Theory, 32(1), 353 - 384. doi: 10.2308/ajpt-50324

[9] Chye Koh, H. and Kee Low, C. (2004). Going concern prediction using data mining techniques. Managerial Auditing Journal, 19 (3), 462 - 476. doi: 10.1108/02686900410524436

[10] Cook, J. and Ramadas, V. (2020). When to consult precision-recall curves, The Stata Journal, 20 (1), 131 - 148. doi: 10.1177/1536867X2090969

[11] Curth, A., Jeffares, A. and van der Schaar, M. (2024). A U-turn on double descent: Rethinking parameter counting in statistical learning. Advances in Neural Information Processing Systems, 36. doi: 110.48550/arXiv.2310.18988

[12] Čorić, R., Matijević, D. and Marković, D. (2023). PollenNet-a deep learning approach to predicting airborne pollen concentrations. Croatian Operational Research Review, 14(1), 1-13. doi: 10.17535/crorr.2023.0001

[13] Del Giudice, M. (2024). The prediction-explanation fallacy: a pervasive problem in scientific applications of machine learning, Methodology, 20(1), 22 - 46. doi: 10.5964/meth.11235

[14] Gadžo, A., Suljić, M., Jusufović, A., Filipović, S. and Suljić, E. (2025). Data mining approach in detecting inaccurate financial statements in government-owned enterprises. Croatian Operational Research Review, 16(1), 1-15. doi: 10.17535/crorr.2025.0001

[15] Goo, Y. J. J., Chi, D. J. and Shen, Z. D. (2016). Improving the prediction of going concern of Taiwanese listed companies using a hybrid of LASSO with data mining techniques. SpringerPlus, 5(1), 1 - 18. doi: 10.1186/s40064-016-2186-5

[16] Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction, Springer, New York.
[17] Hastie, T., Tibshirani, R., Wainwright, M. (2015). Statistical learning with sparsity, Monographs on statistics and applied probability 143, CRC Press, Boca Raton.
[18] Hsu, Y. F. and Lee, W. P. (2020). Evaluation of the going-concern status for companies: An ensemble framework-based model. Journal of Forecasting, 39(4), 687-706. doi: org/10.1002/for.2653
[19] Kalinić Milićević, T., and Marasović, B. (2023). What factors influence Bitcoin's daily price direction from the perspective of machine learning classifiers? Croatian Operational Research Review, 14(2), 163-177. doi: 10.17535/crorr.2023.0014
[20] Krupa, J. and Minutti-Meza, M. (2022). Regression and Machine Learning Methods to Predict Discrete Outcomes in Accounting Research. Journal of Financial Reporting, 7(2), 131-178. doi: 10.2308/JFR-2021-010
[21] Martens, D., Bruynseels, L., Baesens, B., Willekens, M. and Vanthienen, J. (2008). Predicting going concern opinion with data mining. Decision Support Systems, 45(4), 765-777. doi: 10.1016/j.dss.2008.01.0
[22] Nattino, G., Lemeshow, S., Phillips, G., Finazzi, S. and Bertolini, G. (2017). Assessing the Calibration of Dichotomous Outcome Models with the Calibration Belt, The Stata Journal, 17(4), 1003-1014. doi: 10.1177/1536867X1801700414
[23] Saeedi, A. (2021). Audit opinion prediction: A comparison of data mining techniques. Journal of Emerging Technologies in Accounting, 18(2), 125 - 147. doi: 10.2308/JETA-19-10-02-40
[24] Saeedi, A. (2023). A High-Dimensional Approach to Predicting Audit Opinions. Applied Economics, 55(33), 3807 - 3832. doi: 10.1177/1536867X1801700414
[25] Shmueli, G. (2010). To explain or to predict?, Statistical Science, 25(3), 289-310. doi: 10.1214/10-STS330
[26] Šinkovec, H., Heinze, G., Blagus, R. and Geroldinger, A. (2021). To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. BMC medical research methodology, 21, 1 - 15. doi: 10.1186/s12874-021-01374-y
[27] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267 - 288. doi: 10.1111/j.2517-6161.1996.tb02080.x
[28] Van Calster, B., McLernon, D. J., Van Smeden, M., Wynants, L. and Steyerberg, E. W. (2019). Calibration: the Achilles heel of predictive analytics. BMC medicine, 17(1), 1- 7. doi: 10.1186/s12916-019-1466-7
[29] Weiss, G. M. (2013). Foundations of imbalanced learning. In: He, H., and Ma, Y. (Eds.). Imbalanced learning: Foundations, Algorithms, and Applications, Hoboken, NJ, USA: Wiley, 13 - 41. doi: 10.1002/9781118646106.ch2
[30] Yang, Z., Yu, Y., You, C., Steinhardt, J. and Ma, Y. (2020). Rethinking bias-variance trade-off for generalization of neural networks. In H. Daumé, III and A. Singh (Eds.), Proceedings of the 7th International Conference on Machine Learning, pp.10767–10777. url: https://proceedings.mlr.press/v119/yang20j.htm
[31] Yeh, C. C., Chi, D. J. and Lin, Y. R. (2014). Going-concern prediction using hybrid random forests and rough set approach. Information Sciences, 254, 98 - 110. doi: 10.1007/978-3-319-09333-8_24
[32] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418 - 1429. doi: 10.1198/016214506000000735

Croatian Operational Research Review
CRORR **17:2**(2025), 177–190

Inverse cost and revenue efficiency in network processes with uncontrollable inputs

Fatemeh Gholami Golsefid[1], Monireh Jahani Sayyad Noveiri[1] and Sohrab Kordrostami[1,*]

[1] *Department of Mathematics, Lahijan Branch, Islamic Azad University, Lahijan, Iran*
E-mail: ⟨*gholami_ so@yahoo.com, monirehjahani@yahoo.com, Sohrabkordrostami@gmail.com*⟩

Abstract. Inverse data envelopment analysis (IDEA) estimates the inputs/outputs of each decision-making unit (DMU) based on the perturbations in the outputs/inputs while maintaining relative efficiency. When the cost of inputs or the price of outputs is available, it is possible to calculate cost, revenue, and profit efficiencies. This study develops an inverse network DEA that includes uncontrollable measures. For this purpose, models are presented for calculating relative, cost, and revenue efficiencies. Then, an algorithm is proposed to estimate the inputs of the first stage, considering unchanged technical and cost efficiencies. Also, an algorithm is presented to estimate the outputs of the second stage, considering unchanged output-oriented technical efficiency and revenue efficiency. Then, the introduced algorithms are applied to a numerical example and a dataset related to salmon farming, obtaining logical results. Finally, the proposed method is compared with one of the existing methods, and their differences are discussed.

Keywords: Cost efficiency, Inverse data envelopment analysis, revenue efficiency, uncontrollable inputs

Received: July 14, 2024; accepted: February 11, 2025; available online: April 2, 2025

DOI: 10.17535/crorr.2025.0015

Original scientific paper.

1. Introduction

Data envelopment analysis (DEA), first presented by Charnes et al. [5], is a technique for the relative efficiency calculation of decision-making units (DMUs). However, the inverse DEA (IDEA) is an approach to estimate the performance indicators (inputs and outputs) of individual units by perturbing other measures (outputs and inputs) while maintaining the unit's efficiency [21]. Yan et al. [23] further explored IDEA models by incorporating conic constraints. Subsequently, Jahanshahloo et al. [13] extended the technique provided by Yan et al. [23], utilizing IDEA to assess the output levels of a DMU under increased inputs or improved efficiency. Hadi-Vencheh and Foroughi [11] rendered a technique that involved simultaneous increases in certain inputs (outputs) while decreasing others.

Researchers can analyze the performance of individual units in terms of cost, revenue, or profit when input and output prices are known. The main models for assessing cost efficiency (CE) and revenue efficiency (RE) focus on minimizing costs and maximizing revenue. Additional models have been introduced to calculate maximum profit and determine profit efficiency. Fare et al. [7] introduced a linear programming model to estimate the cost, revenue, and profit efficiencies of DMUs. Tone [20] proposed an alternative approach for the cost, revenue, and profit efficiency evaluation considering different prices for inputs and outputs. Sahoo et al.

*Corresponding author.

[17] suggested a directional distance function method to evaluate cost, revenue, and profit efficiencies. Ghiyasi [10] developed an IDEA model for situations where price information was available, proposing models to estimate inputs and outputs while maintaining technical and cost (or revenue) efficiencies. Asadi et al. [2] introduced an IDEA approach based on non-convex cost efficiency. Soleimani Chamkhorami et al. [19] determined the minimum required changes in inputs for perturbing outputs so that the cost efficiency of the evaluated unit remained unchanged. In their research, Fathi and Izadikhah [9] and Khoshfetrat and Ghiyasi [15] examined the issue of resource allocation when non-discretionary inputs were present using radial IDEA models. Jahani Sayyad Noveiri et al. [12] introduced non-radial DEA-based models to evaluate the efficiency and output variations of Iranian restaurants, incorporating non-discretionary measures.

Neglecting the internal structure of processes and treating systems as black boxes can lead to inaccurate performance evaluations. Internal processes can significantly affect performance, which is why researchers such as Fare and Grosskopf [6] introduced the concept of network structure for DMUs in DEA. Subsequent methods, such as those developed by Kao and Hwang [14] and Wu et al. [22], have sought to assess the efficiency of two-stage networks. Recent studies, including Lozano's approach [16], have focused on analyzing the cost efficiency of two-stage processes considering intermediate factors. Banihashemi and Tohidi [4] used a slacks-based model to determine the cost, revenue, and profit efficiencies of network structures based on input and output price vectors, also extending their analysis to supply chain networks [3]. Shiri et al. [18] applied the IDEA concept to calculate cost and revenue efficiencies in two-stage structures while expanding Ghiyasi's method for two-stage networks. Despite the efforts made to estimate performance indicators in the presence of uncontrollable measures, changes in performance measures related to network processes with uncontrollable indexes and price data have not been investigated.

Therefore, this study considers each DMU as a two-stage network, which includes uncontrollable inputs in the second stage. Given that the price of inputs of the first stage (initial input) and outputs of the second stage (final output) are available, models are presented to calculate the relative and cost efficiencies. Then, the inputs of the first stage are estimated using the generalized IDEA by perturbing the final outputs so that the technical and cost efficiencies remain unchanged. Using another presented model by perturbing the inputs of the first stage, final outputs are estimated while maintaining technical and revenue efficiencies. The proposed algorithms are then applied to a numerical example and a dataset related to 17 salmon farms in five western provinces of Iran. Moreover, the proposed method is compared with one of the existing methods.

Section 2 presents methods to estimate the performance measures of network systems with uncontrollable inputs and price information. Datasets are presented in Section 3 to demonstrate the introduced approach. Finally, Section 4 is allocated to the results and suggestions.

2. Inverse cost and revenue efficiencies for networks with uncontrollable inputs

Suppose there are n DMUs, where each unit has a two-stage network structure with uncontrollable inputs in the second stage. That is, DMU_o is in the form shown in Figure 1: Here, $x_{io},\ (i = 1, 2, ..., m)$ and $z_{ko},\ (k = 1, 2, ..., t)$ are the inputs of the first stage and the intermediate measures, respectively. The index z_{ko} is considered the output of the first stage and the input of the second stage. In the second stage, $d_{lo},\ (l = 1, 2, ..., g)$ and $y_{ro},\ (r = 1, 2, ..., s)$ show uncontrollable inputs and the final outputs, respectively.

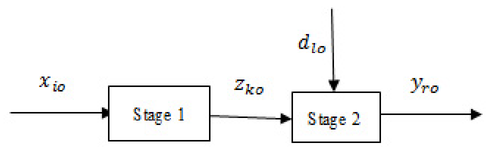


Figure 1: *The two-stage network structure.*

Accordingly, the technical efficiency can be obtained from the following model:

$$\begin{array}{lll} & \theta = \min\ \theta_o & \\ \text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_o x_{io}, & i = 1, \dots, m, \\ & \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \geq 0, & k = 1, \dots, t, \\ & \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1, \dots, g, \\ & \sum_{j=1}^{n} \mu_j y_{rj} \geq y_{ro}, & r = 1, \dots, s, \\ & \lambda_j, \mu_j \geq 0, & j = 1, \dots, n. \end{array} \tag{1}$$

Suppose $(\theta^*, \lambda^*, \mu^*)$ is the optimal solution for model (1), with θ^* indicating the technical efficiency of DMU_o. If $\theta^* = 1$, then DMU_o is efficient; otherwise, it is inefficient. If the cost or price of the inputs of the first stage is available, the cost efficiency for DMU_o can be obtained. Let $c \in R^m$ be the cost or price of the inputs of the first stage. The minimum cost related to DMU_o is calculated using the following model:

$$\begin{array}{lll} \min & \sum_{i=1}^{m} c_i x_i & \\ \text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \leq x_i, & i = 1, \dots, m, \\ & \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \geq 0, & k = 1, \dots, t, \\ & \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1, \dots, g, \\ & \sum_{j=1}^{n} \mu_j y_{rj} \geq y_{ro}, & r = 1, \dots, s, \\ & \lambda_j, \mu_j \geq 0, & j = 1, \dots, n, \\ & x_i \geq 0, & i = 1, \dots, m. \end{array} \tag{2}$$

Definition 1. Suppose (x^*, λ^*, μ^*) is the optimal solution of model (2). The cost efficiency is defined as the ratio of the minimum cost $(c^t x^*)$ to the actual cost of DMU_o $(c^t x_o)$, i.e.,

$$CE_o = \frac{c^t x^*}{c^t x_o} = \frac{\sum_{i=1}^{m} c_i x_i^*}{\sum_{i=1}^{m} c_i x_{io}} \tag{3}$$

2.1. Minimum inverse cost for two-stage networks with uncontrollable inputs

If the prices of the inputs related to the first stage are available, they can also be used in IDEA models. Now suppose that in two-stage networks, where uncontrollable inputs are presented in the second stage, the final outputs of the second stage are perturbed from the level y_o to $y_o + h_o$. The following multi-objective model is proposed to estimate the inputs of the first stage for this perturbation, considering unchanged technical and cost efficiencies:

$$
\begin{array}{lllr}
\min & \displaystyle\sum_{i=1}^{m} c_i\bar{x}_i & & \\
\min & (\alpha_1, \alpha_2, \ldots, \alpha_m) & & \\
\text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \le \bar{x}_i, & i = 1, \ldots, m \quad (4.1), & \\
& \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta_o^* \alpha_i, & i = 1, \ldots, m \quad (4.2), & \\
& \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \ge 0, & k = 1, \ldots, t \quad (4.3), & (4) \\
& \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1, \ldots, g \quad (4.4), & \\
& \sum_{j=1}^{n} \mu_j y_{rj} \ge y_{ro} + h_{ro}, & r = 1, \ldots, s \quad (4.5), & \\
& \sum_{i=1}^{m} c_i\bar{x}_i = \mathrm{CE}_o \sum_{i=1}^{m} c_i\alpha_i \quad (4.6), & & \\
& \lambda_j, \mu_j \ge 0, & j = 1, \ldots, n, & \\
& \bar{x}_i, \alpha_i \ge 0, & i = 1, \ldots, m. &
\end{array}
$$

According to the statement (4.6), we have:

$$
\min \sum_{i=1}^{m} c_i\overline{x}_i = \min CE_o \sum_{i=1}^{m} c_i\alpha_i
$$

Because the cost efficiency CE_o calculated using equation (3) is a numerical value, it is sufficient to calculate $min \sum_{i=1}^{m} c_i\alpha_i$. In this way, the multi-objective model (4) becomes as the following linear programming (LP) model:

$$
\begin{array}{lllr}
\min & \displaystyle\sum_{i=1}^{m} c_i\alpha_i & & \\
\text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \le \bar{x}_i, & i = 1, \ldots, m \quad (5.1) & \\
& \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta_o^* \alpha_i, & i = 1, \ldots, m \quad (5.2) & \\
& \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \ge 0, & k = 1, \ldots, t \quad (5.3) & \\
& \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1, \ldots, g \quad (5.4) & (5) \\
& \sum_{j=1}^{n} \mu_j y_{rj} \ge y_{ro} + h_{ro}, & r = 1, \ldots, s \quad (5.5) & \\
& \sum_{i=1}^{m} c_i\bar{x}_i = CE_o \sum_{i=1}^{m} c_i\alpha_i \quad (5.6) & & \\
& \lambda_j, \mu_j \ge 0, & j = 1, \ldots, n & \\
& \bar{x}_i, \alpha_i \ge 0, & i = 1, \ldots, m &
\end{array}
$$

Here, CE_o is the cost efficiency calculated using Eq. (3).

Definition 2. Let $(\lambda, \mu, \alpha,\ \overline{x})$ is a feasible solution for model (5). $(\lambda, \mu, \alpha,\ \overline{x})$ is a weak efficient solution of model (5) if there is no feasible solution $(\lambda', \mu', \alpha',\ \overline{x}')$ such that $\alpha_i' \le \alpha_i\ (\forall i)$.

Theorem 1. *Suppose DMU_o is a two-stage network with uncontrollable inputs in the second stage. Let θ_o^* be the optimal objective of model (1). If $(\lambda, \mu, \alpha,\ \overline{x})$ is the weak efficiency solution for model (5), then the technical and cost efficiencies of DMU_o remain unchanged.*

Proof. Consider the following problem:

$$
\begin{array}{lll}
\min & \bar{\theta} & \\
\text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \leq \bar{\theta}\alpha_i, & i=1,\dots,m \\
& \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \geq 0, & k=1,\dots,t \\
& \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l=1,\dots,g \\
& \sum_{j=1}^{n} \mu_j y_{rj} \geq y_{ro} + h_{ro}, & r=1,\dots,s \\
& \lambda_j, \mu_j \geq 0, & j=1,\dots,n
\end{array}
\tag{6}
$$

Suppose $\overline{\theta}^*$ is the optimal objective value of model (6), It is necessary to show that $\overline{\theta}^* = \theta_o^*$. Because $(\lambda, \mu, \alpha,\ \overline{x})$ is a weak efficiency solution for model (5), it applies to all constraints of model (5), including the following conditions:

$$
\begin{array}{ll}
\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_o^* \alpha_i, & \forall i, \\
\sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \geq 0, & \forall k, \\
\sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & \forall l, \\
\sum_{j=1}^{n} \mu_j y_{rj} \geq y_{ro} + h_{ro}, & \forall r.
\end{array}
$$

As a result, $(\lambda\ ,\ \mu\ ,\ \theta_o^*)$ is a feasible solution to problem (6), so $\overline{\theta}^* \leq \theta_o^*$ (Eqs. (5.1) and (5.6) of model (5) are also established). Now it is enough to show that $\overline{\theta}^* \not< \theta_o^*$. Suppose $\overline{\theta}^* < \theta_o^*$, then it can be assumed $\overline{\theta}^* = k\theta_o^*$ that $0 < k < 1$. Thus, according to the first constraint of problem (6), we have:

$$
\begin{array}{ll}
\sum_{j=1}^{n} \lambda_j x_{ij} \leq k\theta_o^* \alpha_i, & \forall i, \\
\sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \geq 0, & \forall k, \\
\sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & \forall l, \\
\sum_{j=1}^{n} \mu_j y_{rj} \geq y_{ro} + h_{ro}, & \forall r.
\end{array}
$$

As a result, $(\lambda, \mu, k\alpha,\ \overline{x})$ is a feasible solution for model (5) that $0 < k < 1$, which contradicts the weak efficient solution of $(\lambda, \mu, \alpha,\ \overline{x})$ for model (5). Hence, $\overline{\theta}^* \not< \theta_o^*$ leads to $\overline{\theta}^* = \theta_o^*$. That is, technical efficiency does not change.

For cost efficiency, according to the constraint (5.6) of model (5), we have $\sum_{i=1}^{m} c_i \overline{x}_i = CE_o \sum_{i=1}^{m} c_i \alpha_i$ that leads to $\frac{\sum_{i=1}^{m} c_i \overline{x}_i}{\sum_{i=1}^{m} c_i \alpha_i} = CE_o$, where CE_o is the cost efficiency related to DMU_o. By perturbing the final output from the level of y_{ro} to $y_{ro} + h_{ro}$, inputs can also change from the level of x_{io} to $\alpha_i = x_{io} + \Delta x_{io}$. Considering $\sum_{i=1}^{m} c_i \overline{x}_i$ as the minimum cost after perturbing the final outputs, $\sum_{i=1}^{m} c_i \alpha_i$ is the actual cost of the unit after perturbation. In other words, $\frac{\sum_{i=1}^{m} c_i \overline{x}_i}{\sum_{i=1}^{m} c_i \alpha_i}$ is the cost efficiency of the new DMU_o with the changed inputs and outputs of $(x_o + \Delta x_o,\ y_o + h_o)$, which is equal to the cost efficiency of the initial DMU_o (equal to CE_o). As a consequence, cost efficiency will not change.

According to IDEA models, when inputs/outputs are perturbed, the outputs/inputs may also change. In two-stage networks with uncontrollable inputs in the second stage, the minimum value of the network's initial inputs by perturbing the outputs is estimated while keeping technical and cost efficiencies unchanged. For this purpose, the following algorithm (cost algorithm) is presented using models (2), (3), and (5). The steps of the proposed algorithm are shown in Figure 2.

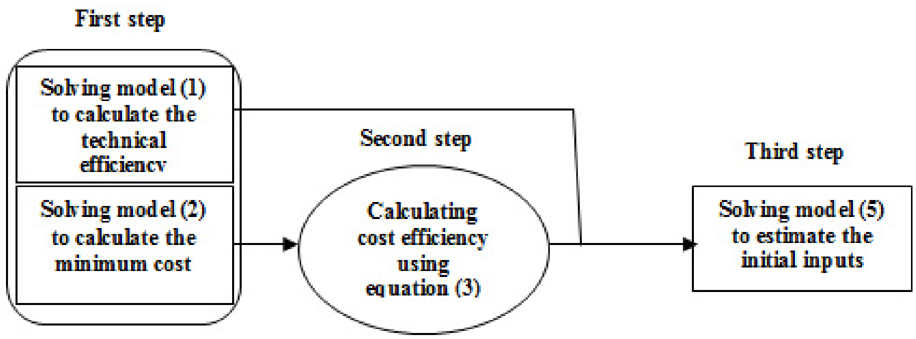


Figure 2: *The proposed algorithm for the input estimation of the initial stage with uncontrollable inputs.*

2.2. Output-oriented technical and revenue efficiencies for two-stage networks with uncontrollable inputs

Suppose there are n DMUs in the frame of two-stage networks with uncontrollable inputs as shown in Figure (1). The output-oriented efficiency of DMU_o is calculated using the following model:

$$\begin{array}{lll} & \varphi = \max\ \varphi_o & \\ \text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io}, & i = 1, \dots, m, \\ & \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \ge 0, & k = 1, \dots, t, \\ & \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1, \dots, g, \\ & \sum_{j=1}^{n} \mu_j y_{rj} \ge \varphi_o y_{ro}, & r = 1, \dots, s, \\ & \lambda_j, \mu_j \ge 0, & j = 1, \dots, n. \end{array} \tag{7}$$

Let $(\varphi^*, \lambda^*, \mu^*)$ be the optimal solution for model (7), with φ^* indicating the output-oriented technical efficiency of DMU_o. If $\varphi^* = 1$, then DMU_o is efficient; otherwise, it is inefficient.

If the output prices of the second stage are available, the revenue efficiency for DMU_o can be obtained. Let $p \in R^s$ be the output price of the second stage. The maximum revenue of DMU_o is calculated using the following model:

$$\begin{array}{lll} \max & \sum_{r=1}^{s} p_r y_r & \\ \text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io}, & i = 1, \dots, m \\ & \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \ge 0, & k = 1, \dots, t \\ & \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1, \dots, g \\ & \sum_{j=1}^{n} \mu_j y_{rj} \ge y_r, & r = 1, \dots, s \\ & \lambda_j, \mu_j \ge 0, & j = 1, \dots, n \\ & y_r \ge 0, & r = 1, \dots, s. \end{array} \tag{8}$$

Definition 3. Let (y^*, λ^*, μ^*) are the optimal solution for model (8). The ratio of real revenue $(p^t y_o)$ to the maximum revenue of DMU_o $(p^t y^*)$ shows the revenue efficiency.

$$RE_o = \frac{p^t y_o}{p^t y^*} = \frac{\sum_{r=1}^{s} p_r y_{ro}}{\sum_{r=1}^{s} p_r y_r^*} \tag{9}$$

2.3. Maximum inverse revenue for two-stage networks with uncontrollable inputs

Given the availability of the output prices of the second stage, this information can also be used in IDEA models. Suppose that in a two-stage network with uncontrollable inputs in the second stage, the inputs of the first stage are perturbed from the level x_o to $x_o + b_o$. The following multi-objective model is proposed to estimate the outputs of the second stage while keeping the output-oriented technical and revenue efficiencies unchanged:

$$
\begin{array}{lll}
\max & \sum_{r=1}^{s} p_r \bar{y}_r & \\
\max & (\beta_1,\ \beta_2,\ \ldots,\ \beta_s) & \\
\text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} + b_{io}, & i = 1,\ldots,m \quad (10.1) \\
& \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \ge 0, & k = 1,\ldots,t \quad (10.2) \\
& \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1,\ldots,g \quad (10.3) \\
& \sum_{j=1}^{n} \mu_j y_{rj} \ge \bar{y}_r, & r = 1,\ldots,s \quad (10.4) \\
& \sum_{j=1}^{n} \mu_j y_{rj} \ge \varphi_o^* \beta_r, & r = 1,\ldots,s \quad (10.5) \\
& \sum_{r=1}^{s} p_r \bar{y}_r = \frac{\sum_{r=1}^{s} p_r \beta_r}{RE_o} \quad (10.6) & \\
& \lambda_j, \mu_j \ge 0, & j = 1,\ldots,n \\
& \bar{y}_r, \beta_r \ge 0, & r = 1,\ldots,s
\end{array}
\tag{10}
$$

According to the statement (10.6), we have:

$$
\max \sum_{r=1}^{s} p_r \bar{y}_r = \max \frac{\sum_{r=1}^{s} p_r \beta_r}{RE_o}
$$

Here, $\frac{1}{RE_o}$ is the inverse of the revenue efficiency calculated by equation (9), representing a numerical value. Therefore, it is enough to calculate $max \ \sum_{r=1}^{s} p_r \beta_r$. In this way, the multi-objective model (10) becomes the following LP model:

$$
\begin{array}{lll}
\max & \sum_{r=1}^{s} p_r \beta_r & \\
\text{s.t.} & \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} + b_{io}, & i = 1,\ldots,m, \\
& \sum_{j=1}^{n} (\lambda_j - \mu_j) z_{kj} \ge 0, & k = 1,\ldots,t, \\
& \sum_{j=1}^{n} \mu_j d_{lj} = d_{lo}, & l = 1,\ldots,g, \\
& \sum_{j=1}^{n} \mu_j y_{rj} \ge \bar{y}_r, & r = 1,\ldots,s, \\
& \sum_{j=1}^{n} \mu_j y_{rj} \ge \varphi_o^* \beta_r, & r = 1,\ldots,s, \\
& \sum_{r=1}^{s} p_r \bar{y}_r = \frac{\sum_{r=1}^{s} p_r \beta_r}{RE_o}, & \\
& \lambda_j, \mu_j \ge 0, & j = 1,\ldots,n, \\
& \bar{y}_r, \beta_r \ge 0, & r = 1,\ldots,s.
\end{array}
\tag{11}
$$

Here, RE_o is the revenue efficiency calculated using equation (9).

Definition 4. Let $(\lambda, \mu, \beta,\ \bar{y})$ is a feasible solution for model (11). Then, $(\lambda, \mu, \beta,\ \bar{y})$ is a weak efficient solution for model (11) provided that there is no feasible solution $(\lambda', \mu', \beta',\ \bar{y}')$ such that $\beta'_r \ge \beta_r \ \ (\forall r)$.

Theorem 2. *Suppose DMU_o is a two-stage network with uncontrollable inputs in the second stage. Let φ_o^* is the optimal objective of model* (7). *If* $(\lambda, \mu, \beta, \ \overline{y})$ *is the weak efficiency solution for model* (11), *the output-oriented technical and revenue efficiencies of DMU_o remain unchanged.*

Proof. The proof is similar to Theorem 1.

In network structures with uncontrollable inputs in the second stage, perturbing the inputs of the first stage can change the final outputs. The following proposed (revenue) algorithm is used to estimate the maximum output of the second stage of the network while keeping both output-oriented and the revenue efficiencies unchanged.

Algorithm 2: Inverse revenue algorithm for networks with uncontrollable inputs

Step 1) Calculate the output-oriented efficiency of the units using model (7).

Step 2) Calculate the maximum revenue using model (8) and then calculate the revenue efficiency using equation (9).

Step 3) Apply model (11) to estimate the maximum value of the final outputs by perturbing the inputs of the first. □

2.4. Numerical Example

Consider 10 DMUs with a two-stage network structure shown below:

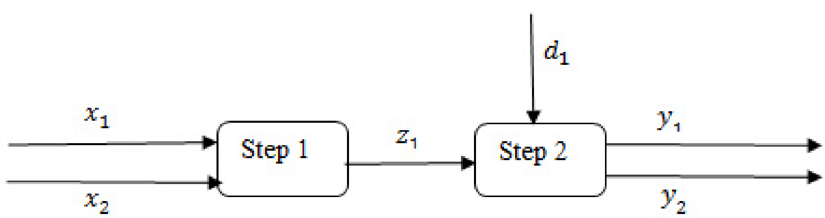


Figure 3: *The two-stage network structure.*

The inputs and outputs of these 10 units are given in Table 1.

	x_1	x_2	z_1	d_1	y_1	y_2
DMU_1	6.2	62.4	58	8.1	51.1	1.64
DMU_2	4.1	47.4	88	8.1	33.3	1.74
DMU_3	10.0	100.0	99	6.3	100.0	9.99
DMU_4	2.8	44.9	25	10.0	31.4	2.69
DMU_5	5.5	51.9	13	8.1	37.5	2.27
DMU_6	5.0	60.6	48	8.1	45.8	4.18
DMU_7	3.1	39.6	73	9.1	34.9	0.15
DMU_8	3.0	35.4	67	10.0	26.3	1.25
DMU_9	2.9	38.0	86	8.1	29.1	1.43
DMU_{10}	6.1	70.8	48	6.3	72.4	6.02

Table 1: *Inputs and outputs of 10 networks*

According to the cost algorithm, we first obtain the technical efficiency of the units using model (1) and then calculate the cost efficiency considering the input prices ($c_1 = 2$ and $c_2 = 3$), model (2), and equation (3). Then, we estimate the minimum amount of inputs using model (5) by perturbing the first output by 10% and the second output by 20%. The estimation results of the inputs related to the first stage and the percentage of changes are given in Table 2.

	New x_1	Percentage of changes x_1	New x_2	Percentage of changes x_2
DMU_1	5.62	-8.8%	73.61	17.9%
DMU_2	3.62	-11.1%	47.4	-0.1%
DMU_3	9.61	-4.1%	125.93	25.9%
DMU_4	3.46	25.1%	45.35	1%
DMU_5	5.19	-6%	68.04	31.2%
DMU_6	5.77	14.6%	75.63	24.8%
DMU_7	3.02	-3.3%	39.6	-0.1%
DMU_8	2.7	-11.5%	35.4	0.1%
DMU_9	2.9	1.2%	38	0.1%
DMU_{10}	7.14	17.8%	93.61	32.2%

Table 2: *Estimated inputs of the first stage and their percentage of changes*

According to Table 2, the first input decreases for units 1, 2, 3, 5, 7, and 8 and increases for other units. The largest decrease in the first input is related to DMU 8, with an 11.5% decrease, and the largest increase in this input is observed in DMU 4 by about 25%. Also, the second input of DMUs 2 and 7 decreases by approximately 0.1% but increases for other units. The largest increase in this input is related to unit 10 by about 32%.

We use the method presented by Shiri et al. [18] to compare the proposed method with existing studies, highlighting the following differences between the proposed technique and Shiri et al.'s model:

- Shiri et al. [18] use four models to solve such problems. They first calculate technical efficiency and cost efficiency using two models, then use two other models (the first has three and the second has four types of constraints) to estimate the minimum input while maintaining technical efficiency and cost efficiency. After calculating technical and cost efficiencies, the proposed model would reach the same results using a model with only one additional constraint. Accordingly, results confirm that the proposed method would reduce the number of computational operations and the time required for calculations.

- Another difference is that there are uncontrollable inputs in the proposed network, which have not been taken into account by Shiri et al.'s method.

For comparisons, we consider the uncontrollable input index in Shiri et al.'s method, adding the third constraint of model (11) to the models of Shiri et al. [18]. The results are presented in Table 3.

According to Table 3, the first input decreases for units 1, 2, 3, 5, and 8 and increases for other units. The largest decrease is 8.8%, associated with DMU 1, and the largest increase is 30.5%, related to DMU 4. Also, the second input values of all units increase, with the smallest and largest increase reported for units 4 and 10, respectively. A comparison of Tables 2 and 3 reveals that the inputs of some units decrease in the proposed method but increase in the method of Shiri et al. (such as the first input of DMU 7 and the second input of units 2 and 7). The changes in some units are the same in both methods.

2.5. Case Study

The data are related to the performance of 17 salmon farms in five western provinces of Iran in 2018 (Abdali [1]). The framework under examination of fish farms is shown in Fig. 4, considering each salmon farm as a two-stage network system. The first stage is the baby fish's production pool, and the second stage is the adult fish's breeding pool. First, all the fish are

	New x_1	Percentage of changes x_1	New x_2	Percentage of changes x_2
DMU_1	5.62	-8.8%	73.61	17.9%
DMU_2	3.98	-2.3%	52.14	9.9%
DMU_3	9.61	-4.1%	125.93	25.9%
DMU_4	3.61	30.5%	47.25	5.2%
DMU_5	5.19	-6%	68.04	31.2%
DMU_6	5.77	14.6%	75.63	24.8%
DMU_7	3.32	6.3%	43.56	9.9%
DMU_8	2.97	-2.6%	38.94	10.1%
DMU_9	3.19	11.4%	41.8	10.1%
DMU_{10}	7.14	17.8%	93.61	32.2%

Table 3: *Results of Shiri et al.'s method considering the uncontrollable input*

kept in the first pool for spawning, and then the baby fish are sent to the second pool to grow up. The final output is the adult fish for human consumption. The cost of ecosystem damage is the uncontrollable input in the second pool, where the young fish are placed to grow up. The terms used are defined as follows:

x_i: Value of fish brought into the first pool and kept in that pool for spawning (unit of 1000 kg).

d_l: Costs of ecosystem damage in the second pool of the salmon farm (unit of one million Tomans)

z_k: The number of baby fish sent to the second pool to grow up (unit of 1000 pieces).

y_r: Value of mature fish for people's food consumption (unit of 1000 kg).

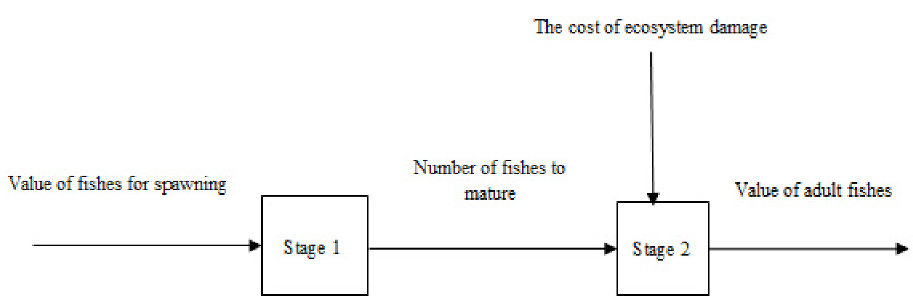


Figure 4: *The two-stage process related to fish farms.*

Table 4 shows the inputs and outputs of 17 salmon farms in five western provinces of Iran. Given the cost algorithm and the price of the inputs of the first stage (c=5), the cost efficiency of the units is determined using model (2) and Expression (3) after determining the technical efficiency (using model (1)). Then, the output of the second stage (amount of mature fish for consumption) is perturbed by 10% to estimate the minimum amount of fish for spawning after this perturbation using model (5) so that technical and cost efficiencies remain unchanged. The results of the amount of fish for spawning, the percentage of changes, the technical efficiency, and the cost efficiency of the units are given in Table 5.

Stage 1			Stage 2	
DMU_j	x_i	z_k	d_l	y_r
DMU_1	656	98.541	56	3564
DMU_2	5356	32.654	29	2145
DMU_3	354	82.365	67	3698
DMU_4	542	24.563	45	5246
DMU_5	652	98.543	19	7145
DMU_6	752	54.263	45	3369
DMU_7	326	65.478	36	5214
DMU_8	542	36.985	68	5369
DMU_9	632	45.638	45	5412
DMU_{10}	623	96.353	36	2148
DMU_{11}	354	65.878	45	4532
DMU_{12}	654	36.546	42	4521
DMU_{13}	423	54.692	25	3678
DMU_{14}	362	24.578	42	4213
DMU_{15}	452	36.925	36	2846
DMU_{16}	325	57.892	24	6532
DMU_{17}	542	65.245	45	4215

Table 4: *Inputs and outputs of 17 salmon farms*

DMU_j	*New* x_i	Percentage of changes	Technical efficiency	Cost efficiency
DMU_1	656	0%	0.1996	0.1996
DMU_2	5357.23	0.02%	0.0127	0.0127
DMU_3	354	0%	0.4424	0.4424
DMU_4	680.85	25.62%	0.1948	0.1948
DMU_5	778.25	19.36%	0.6496	0.6496
DMU_6	752.24	0.03%	0.1399	0.1399
DMU_7	390.01	19.63%	0.4199	0.4199
DMU_8	542.41	0.08%	0.2933	0.2933
DMU_9	786.51	24.45%	0.1806	0.1806
DMU_{10}	623	0%	0.1351	0.1351
DMU_{11}	354.34	0.1%	0.2978	0.2978
DMU_{12}	680.87	4.11%	0.1505	0.1505
DMU_{13}	504.83	19.35%	0.2317	0.2317
DMU_{14}	362.34	0.09%	0.2718	0.2718
DMU_{15}	452.34	0.08%	0.1862	0.1862
DMU_{16}	422.93	30.13%	0.7656	0.7656
DMU_{17}	542.48	0.09%	0.1944	0.1944

Table 5: *Findings of performance*

According to the results of Table 5, the amount of fish for spawning has not changed in units 1, 3, and 10 (0% change), while the amount of fish for spawning has a slight change in units 2, 6, 8, 11, 14, 15, and 17 (increased by less than one-tenth percent). Units 16, 4, and 9 have undergone the greatest changes, by about 30%, 26%, and 24% increase, respectively.

The fourth and fifth columns of Table 5 show the technical and cost efficiencies of the units, respectively. Notice that each unit has only one input and one output, leading to the equivalence of technical and cost efficiencies [20].

Using the revenue algorithm, we first obtain the output-oriented technical efficiency and the revenue efficiency of the units using models (7) and (8) and the relation (9) (we consider the output price p=10). Then, the amount of fish for spawning is perturbed by 10%, and the maximum amount of final output is estimated for this perturbation using model (11) so that the output-oriented technical and revenue efficiencies remain unchanged. The new output results, the percentage of their changes, the output-oriented technical efficiency, and the revenue efficiency of the units are given in Table 6.

DMU_j	New y_r	Percentage of changes	Output-oriented efficiency	Revenue efficiency
DMU_1	3692.12	3.59%	4.4609	0.2242
DMU_2	2145	0%	5.0842	0.1967
DMU_3	3916.37	5.91%	3.1445	0.3180
DMU_4	5437.96	3.66%	2.4600	0.4065
DMU_5	7145	0%	1.0000	1.0000
DMU_6	3518.8	4.45%	4.3736	0.2286
DMU_7	5585.2	7.12%	1.7035	0.5870
DMU_8	5736.81	6.85%	2.8583	0.3499
DMU_9	5629.69	4.02%	2.5294	0.3954
DMU_{10}	2245.37	4.53%	5.5746	0.1794
DMU_{11}	4841.19	6.82%	2.2208	0.4503
DMU_{12}	4714.18	4.27%	2.9496	0.3390
DMU_{13}	3842.67	4.48%	2.2380	0.4468
DMU_{14}	4508.62	7.02%	2.3752	0.4210
DMU_{15}	2952.9	3.76%	3.6839	0.2715
DMU_{16}	6789.04	3.94%	1.1016	0.9078
DMU_{17}	4369.23	3.66%	3.0617	0.3266

Table 6: *Efficiencies, new outputs, and changes*

According to Table 6, the amount of adult fish has not changed in units 2 and 5 (0% changes), and the amount of adult fish has changed by 7.12%, 7.02%, and 6.85% in units 7, 14, and 8, respectively. Unit 11 shows the highest increase by 6.82%. Output-oriented technical and revenue efficiencies are provided in the fourth and fifth columns, revealing that unit 5 has both output-oriented technical and revenue efficiencies.

Uncontrollable indicators in network structures are important and effective in finding accurate results. Using the proposed models, it is possible to examine the changes of indicators in network processes with uncontrollable inputs while the prices of some indicators are known. For more explanation, the methodology employed allows for the efficiency evaluation of DMUs and performance measures estimation within a two-stage network framework. By distinguishing between the controllable inputs in the first stage and the uncontrollable inputs in the second stage, we can gain deeper insights into overall performance. Businesses can utilize the inverse network models to estimate the necessary adjustments to inputs and outputs without altering cost/revenue and technical efficiencies. This allows managers to develop targeted strategies. Understanding the efficiency metrics can guide decision-makers in resource reallocation within the network with uncontrollable inputs. Regularly process assessment and recalibration based on efficiency metrics enables businesses to adapt to changing conditions and maintain competitive advantage.

In summary, the results of our study provide a robust framework for organizations to diagnose inefficiencies, implement targeted improvements, and ultimately enhance their business processes.

3. Conclusions

DEA evaluates the relative efficiency of a set of DMUs using mathematical programming models. In IDEA, the required inputs (outputs) are estimated by perturbing outputs (inputs) while maintaining the efficiency value. On the other hand, if input or output prices are available, the cost or revenue efficiencies of DMUs can be obtained using DEA. This paper has considered each DMU a two-stage network with uncontrollable inputs in the second stage. First, according to the availability of the input cost of the first and the output price of the second stage, models have been proposed to calculate the relative, cost, and revenue efficiencies of the desired network structures. Then, using the introduced inverse network models, inputs/outputs are estimated, while the cost/revenue efficiency remains unchanged in addition to the technical efficiency. Finally, the proposed algorithms are applied to a numerical example and data related to the performance information of 17 salmon farms in five western provinces of Iran. Also, the proposed method is compared with one of the existing methods, confirming that it is useful and practical for estimating performance indicators in network systems with uncontrollable inputs.

While the presented approach effectively addresses two-stage systems with precise data, it is important to acknowledge several key limitations. Our method assumes that data are precise, which may not always be the case in real-world applications. Future research could explore the robustness of our approach when dealing with imprecise or uncertain data, potentially incorporating techniques from fuzzy logic or stochastic modeling. Furthermore, the current method is tailored to specific network structures. Developing adaptive techniques that can cater to a wider variety of network configurations would enhance the versatility of our approach. Future work could focus on generalizing our method to accommodate different types of networks, including dynamic structures.

References

[1] Abdali, E., and Fallahnejad, R. (2020). A bargaining game model for measuring efficiency of two-stage network DEA with non-discretionary inputs. International Journal of Computer Mathematics: Computer Systems Theory, 5(1), 48-59. doi: 10.1080/23799927.2020.1723708

[2] Asadi, F., Kordrostami, S., Amirteimoori, A., & Bazrafshan, M. (2022). Inverse data envelopment analysis on the base of non-convex cost efficiency. Iranian Journal of Operations Research,13(1), 135-148.

[3] Banihashem, S., Sanei, M., and Manesh, Z. M. (2013). Cost, revenue and profit efficiency in supply chain. African Journal of Business Management, 7(41), 4280-4287. doi: 10.5897/AJBM12.499

[4] Banihashemi, S., & Tohidi, G. (2013). Allocation efficiency in network DEA. International journal of data envelopment analysis, 1(2), 85-96.

[5] Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429-444. doi: 10.1016/0377-2217(78)90138-8

[6] Färe, R. and Grosskopf, S., (2000). Network DEA, Socio-Economic Planning Sciences, 34, 35-49. doi: 10.1007/978-0-387-71607-7_12

[7] Fare, R., Grosskopf, S., Lovell, C.A.K. (1985). The Measurement of Efficiency of Production. Kluwer-Academic Publishers, Boston.

[8] Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the royal statistical society: series A (General), 120(3), 253-281. doi: 10.2307/2343100

[9] Fathi, N. and Izadikhah, M. (2013). Inverse data envelopment analysis model in the present of non-discretionary and discretionary data to preserve relative efficiency values: The case of variable returns to scale. Journal of Basic and Applied Scientific Research, 3(4), 872-880.

[10] Ghiyasi, M. (2017). Inverse DEA based on cost and revenue efficiency. Computers & Industrial Engineering, 114, 258-263. doi: 10.1016/j.cie.2017.10.024

[11] Hadi-Vencheh, A., and Foroughi, A. A. (2006). A generalized DEA model for inputs/outputs estimation. Mathematical and Computer Modelling, 43(5-6), 447-457. doi: 10.1016/j.mcm.2005.08.005

[12] Jahani Sayyad Noveiri, M., Kordrostami, S., and Anarestani, R. R. (2023). Estimating outputs using an inverse non-radial model with non-discretionary measures: An application for restaurants. Croatian Operational Research Review, 14(1), 41-52. doi: 10.17535/crorr.2023.0004

[13] Jahanshahloo, G. R., Lotfi, F. H., Shoja, N., Tohidi, G., and Razavyan, S. (2004). The outputs estimation of a DMU according to improvement of its efficiency. Applied Mathematics and Computation, 147(2), 409-413. doi: 10.1016/s0096-3003(02)00734-8

[14] Kao, C., and Hwang, S. N. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. European journal of operational research, 185(1), 418-429. doi: 10.1016/j.ejor.2006.11.041

[15] Khoshfetrat, S. and Ghiyasi, M. (2020). Using Multi-Objective Linear Programming (MOLP) and Data Envelopment Analysis (DEA) models in Non-discretionary Performance Measurement. International Journal of Data Envelopment Analysis, 8(4), 17-38.

[16] Lozano, S. (2011). Scale and cost efficiency analysis of networks of processes. Expert Systems with Applications, 38(6), 6612-6617.doi: 10.1016/j.eswa.2010.11.077

[17] Sahoo, B. K., Mehdiloozad, M., & Tone, K. (2014). Cost, revenue and profit efficiency measurement in DEA: A directional distance function approach. European Journal of Operational Research, 237(3), 921-931. doi: 10.1016/j.ejor.2014.02.017

[18] Shiri Daryani, Z., Tohidi, G., Daneshian, B., Razavyan, S., & Hosseinzadeh Lotfi, F. (2021). Inverse DEA in two-stage systems based on allocative efficiency. Journal of Intelligent & Fuzzy Systems, 40(1), 591-603. doi: 10.3233/JIFS-200386

[19] Soleimani Chamkhorami, K., Hosseinzadeh Lotfi, F., Jahanshahloo, G. R., & Rostamy-Malkhalifeh, M. (2020). Preserving cost and revenue efficiency through inverse data envelopment analysis models. INFOR: Information Systems and Operational Research, 58(4), 561-578. doi: 10.1080/03155986.2019.1627780

[20] Tone, K. (2002). A strange case of the cost and allocative efficiencies in DEA. Journal of the operational research society, 53, 1225-1231. doi: 10.1057/palgrave.jors.2601438

[21] Wei, Q., Zhang, J., & Zhang, X. (2000). An inverse DEA model for inputs/outputs estimate. European Journal of Operational Research, 121(1), 151-163. doi: 10.1016/S0377-2217(99)00007-7

[22] Wu, J., Zhu, Q., Ji, X., Chu, J., & Liang, L. (2016). Two-stage network processes with shared resources and resources recovered from undesirable outputs. European Journal of Operational Research, 251(1), 182-197. doi: 10.1016/j.ejor.2015.10.049

[23] Yan, H., Wei, Q., & Hao, G. (2002). DEA models for resource reallocation and production input/output estimation. European Journal of Operational Research, 136(1), 19-31. doi: 10.1016/S0377-2217(01)00046-7

Aspects of blocking on time-varying tandem queueing network

Anjale Ramesh[1,*] **and M. Manoharan**[1]

[1] *Department of Statistics, University of Calicut, Kerala-673 635, India*
E-mail: ⟨{*anjaleramesh, manumavila*}*@gmail.com*⟩

Abstract. Most real-world service systems are susceptible to queue capacity constraints, leading to a blockage of entities. Several blocking mechanisms can be implemented when capacity constraints influence the flow of entities through the system. This study investigates the comparative performance of blocking-after-service (BAS) and blocking-before-service (BBS) mechanisms by modelling a hospital emergency department with a finite capacity queue. The model is based on a two-station tandem network with finite capacity on the intermediate queue, and we developed transient performance measures for the system in both mechanisms. Using a numerical approach, we highlight how these mechanisms influence the time-varying number of patients and the virtual workload in the system. Our results demonstrate that the BAS mechanism slightly outperforms the BBS mechanism in reducing unwanted congestion.

Keywords: blocking-after-service (BAS), blocking-before-service (BBS), tandem queues, transient analysis, virtual workload

Received: August 7, 2024; accepted: January 22, 2025; available online: April 2, 2025

DOI: 10.17535/crorr.2025.0016

Original scientific paper.

1. Introduction

In queueing theory, capacity restriction on waiting line is a crucial aspect to study and it is common for real-world service systems to have queues of finite capacity. In such systems, the flow of customers from the source node will be blocked if the waiting room at the destination node is full. There are mainly two blocking mechanisms, i.e. blocking-after-service (BAS) and blocking-before-service (BBS) that describe different scenarios when there are restrictions on waiting rooms. BAS occurs when an entity after service from a node finds that waiting room of the next station is full (saturated). So, they are being blocked before entering the waiting room of the next node. They must have to wait until the space becomes available. In BBS system, before starting service at current node, entities are blocked if there is no available space in the waiting room of next station. Once the space is available at the destination node, the blocked entity resumes service at the source node.

Tandem queueing networks with finite capacity are useful for modelling healthcare, communication, and manufacturing systems [6, 13, 14, 10]. A BAS mechanism is also known as manufacturing blocking. In manufacturing and production lines, items move through the workstations which can only process a limited number of items at a time. If the next workstation is full, it is not possible to move the items that have been processed from the current workstation, leading to a blockage. In [2], a steady-state analysis was performed under the BAS mechanism with two single-server queues connected in tandem. In [1], the same model is extended to a k-station tandem network with general arrival times, deterministic service times and finite

*Corresponding author.

waiting room between stations. Transient behaviour of a two station tandem network with no restriction on first station and no queue allowed for second station was investigated in [11]. Zychlinski et. al [18] developed time-varying fluid models for tandem network with a general time-varying arrival rate, a finite waiting room before first station and no intermediate waiting room. The BBS mechanism, which is also known as communication blocking, is commonly used in telecommunication networks [15, 8]. A detailed description of the different types of BBS mechanisms is presented in [5]. In communication networks, data packets move through a series of nodes/router, where each node processes the packets and forward it to the next node. If the buffer of the next node is full, the current node cannot forward the packet causing the packet to be blocked before the current node. Healthcare systems can also use the BBS mechanism in short medical procedures, such as cataract surgery, laparoscopic surgery and cardiac catheterization. These procedures can only begin when the room is available in the recovery area. Avi-Itzhak and Levy[4] introduced a k-stage blocking scheme as a generalisation of the results presented in [2, 1]. In [3], they analyzed the steady-state performance measures of a single-server network of k stations with no intermediate queue and an unlimited buffer prior to the first station under both BAS and BBS mechanisms. Fluid limits for tandem model of time-varying multi-server queues with finite buffers before the first station and between stations under BBS mechanism is considered in [19]. To facilitate comparison, they also developed steady-state closed-form expressions for system performance measures under the BAS and BBS mechanisms.

There has been extensive research conducted on tandem networks with finite capacity queues. However, there is limited research on time-varying tandem queues with blocking. In this study, we provide an analytical comparison between BBS and BAS in time-varying tandem queues, with special reference to a case of healthcare system. We develop a stochastic model for a two station finite capacity tandem network under different blocking mechanisms. In the second section, explicit expressions for transient performance measures such as number of patients and average virtual workload at time t under BAS and BBS mechanisms are discussed. We also conducted a numerical study with the blocking mechanisms under different traffic intensity and queue capacity.

2. Two station Tandem Model with finite queue capacity

We consider a health care system, such as a hospital emergency department with a triage system where patients arrive according to a non-homogeneous Poisson process. The model considered here is a tandem network with two stations. Patients are first assessed by a triage nurse to determine their level of need for medical assistance. Subsequently, the patients are sent to the consultation room, where medical professionals provide the necessary treatment. There is an unlimited waiting room for triage and finite waiting room for treatment. Suppose that only a limited number of patients can wait in the treatment area due to space and resource constraints. In the above situation, there are two ways to manage heavy traffic of patients. In BAS, if the waiting space of treatment room is full (saturated), patients cannot join the queue for treatment after the triage process, causing a blockage. In BBS, even though the triage nurse is present, the treatment waiting room is full (saturated), preventing patients from being triaged and causing them to be blocked. In this section, we establish transient performance measures for the two-station tandem queueing network model, under the two blocking mechanisms. Ramesh and Manoharan [12] derived explicit expressions for time-varying measures, such as queue length and virtual workload, for a time-varying tandem queueing network of k stations. Building on that work, this paper extends these measures by incorporating the effects of blocking within the context of an outpatient clinic to facilitate meaningful comparisons.

2.1. Blocking After Service (BAS)

Initially, we model a healthcare system with a two-station tandem queueing network with finite queue capacity in second station. In the first-come-first-served (FCFS) model illustrated in Figure 1, patients arrive according to non-homogeneous Poisson process and the time-dependent service time following an exponential distribution, i.e. triage node is $M(t)/M(t)/1/\infty$ and treatment node is $M(t)/M(t)/1/K$. Following are the parameters that characterise the model.

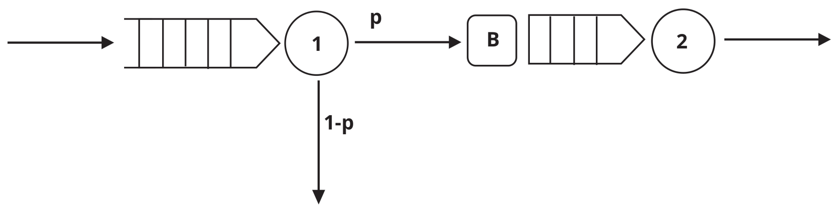


Figure 1: *A BAS two station tandem queueing network model*

1. $\{A_1(t), t \geq 0\}$ is the external arrival process to the triage node with arrival rate $\lambda_1(t)$.

2. $\{V(t), t \geq 0\}$ is the service requirement of the patient, i.e., the total amount of service (in terms of time) that a patient requires. $V_1(t)$ is the service requirement of a patient arriving at the triage at time t. Similarly, $V_2(t)$ is the service requirement of a patient arriving at the treatment area at time t. $\mu_i(t),\ i = 1, 2$ is the service processing rate at time t.

3. After triage process, patients move to the treatment area with probability p and leave the system with probability $1 - p$ and at a rate of $(1 - p)\mu_1(t)$.

4. $\{A_2(t), t \geq 0\}$ denote the arrival of patients to the treatment area after triage process. Therefore, the arrival rate is $\lambda_2(t) = p\mu_1(t)$.

   The instantaneous traffic intensity at the triage node, $\rho_1(t)$ is defined as,

   $$\rho_1(t) = \lambda_1(t)/\mu_1(t).$$

   Thus, the two arrival rate functions are related as $\lambda_2(t) = p\ \lambda_1(t)/\rho_1(t)$.

5. In this study, the instantaneous traffic intensity, $\rho(t) = \lambda(t)/\mu(t)$, which measures the utilisation of a service system is chosen to be invariant of time. This adaptation is made to choose service rate function properly in order to adjust with the arrival rate and traffic intensity.

   We use the principle of rate-matching control, as discussed in [16], to determine the service rate function. In rate-matching control, the service rate is set to be proportional to the arrival rate for a fixed traffic intensity ρ. Thus, for a constant traffic intensity ρ_i, the time-dependent service rate function, $\mu_i(t)$ can be written as,

   $$\mu_i(t) \equiv \lambda_i(t)/\rho_i, \quad i = 1, 2 \quad t \geq 0. \tag{1}$$

6. There is an infinite waiting room at the triage node and a finite waiting room at the treatment node with a maximum capacity of K. If the finite buffer at treatment node is full, patients will be blocked. When the capacity of the waiting room of treatment node is $K - 1$, blocked patients will join the queue for consultation.

The following are the formulations of some transient performance measures associated with the model under consideration.

1. $\{W_1(t), t \geq 0\}$ is the waiting time of a patient who arrives at triage node at time t. An explicit expression for the probability distribution of waiting time $W(t)$ for $M(t)/M(t)/1/\infty$ is derived by Whitt[16]. If a patient who arrives at time s is still waiting for service in the queue at time t, then we can express the probability that the waiting time of the patient who arrives at time s, is larger than $t-s$, for $0 \leq s \leq t$ as,

$$P(W_1(s) > t-s) = \rho_1 e^{-((1-\rho_2)\Lambda_{t,1}(s))/\rho_1}, \tag{2}$$

where $\Lambda_{t,1}(s) = \Lambda_1(t) - \Lambda_1(s)$, $\Lambda(.)$ is the cumulative arrival rate function defined as,

$$\Lambda_1(u) = \int_0^u \lambda_1(r)\ dr, \quad r \geq 0 \tag{3}$$

and $\Lambda_{t,1}(u)$ need to be strictly increasing and continuous, see [16].

2. $\{W_2(t), t \geq 0\}$ is the waiting time of a patient who joins the queue of treatment area at time t. Since the queue capacity is finite, i.e. $M(t)/M(t)/1/K$, here we derive a closed form expression for probability distribution of waiting time.

Let P_n be the probability that there are n patients in the queue in front of treatment area, see [9].

$$P_n = \frac{(1-\rho_2)\ \rho_2^n}{1-\rho_2^{K+1}},$$

where ρ_2 and K are the traffic intensity and queue capacity of the treatment area. Let Q_n be the probability of the arrival point, that is, the probability that there are n patients in the queue at the time of arrival, for $n < K$. This is derived in Gross and Harris[9] using Baye's theorem.

$$Q_n = \frac{P_n}{1-P_{K+1}}. \tag{4}$$

Then the probability distribution of waiting time for the stationary $M/M/1/K$ system can be obtained by reducing the expression for multi-server system in [9] to single server, i.e.,

$$\begin{aligned} P\{W > t\} &= \sum_{n=1}^{K-1} Q_n \sum_{i=0}^{n-1} \frac{(\mu t)^i\ e^{-\mu t}}{i!} \\ &= \sum_{n=1}^{K-1} Q_n \sum_{i=0}^{n-1} \frac{(\lambda t/\rho)^i\ e^{-(\lambda t/\rho)}}{i!}. \end{aligned} \tag{5}$$

The parameter μ is replaced by λ/ρ. "From this, the waiting time distribution for a non-stationary system can be derived using Corollary 5.1 from Whitt [16]. Specifically, under the assumptions of a non-stationary system, λt in equation (5) becomes the cumulative arrival rate function $\Lambda(t)$.

For the non-stationary system $M(t)/M(t)/1/K$, let $P(W_2(s) > t-s)$ denote the probability that the waiting time of a patient joining the queue of the treatment node at time s exceeds $t-s$, for $0 \leq s \leq t$,

$$P(W_2(s) > t-s) = \sum_{n=1}^{K-1} Q_n \sum_{i=0}^{n-1} \frac{(\frac{\Lambda_{t,2}(s)}{\rho_2})^i\ e^{\left(\frac{\Lambda_{t,2}(s)}{\rho_2}\right)}}{i!}. \tag{6}$$

The cumulative rate function $\Lambda_{t,2}(s) = \Lambda_2(t) - \Lambda_2(s)$ with

$$\Lambda_2(u) = \int_0^u \lambda_2(r)\ dr, \quad r \geq 0. \tag{7}$$

This gives the waiting time distribution for an $M(t)/M(t)/1/K$ system under constant traffic intensity.

3. To account for the effects of blocking, we incorporate a blocking probability into the formulation of transient performance measures, i.e. the probability that a patient arriving at time s is blocked after triage at time t. In other words, probability that the blocking time of a patient who arrive at time s is greater than t-s, $P\{B(s) > t - s\}$.

4. $L_1(t)$ represents the number of patients present at the triage node. These patients arrived during the interval $[0, t]$ and have not yet completed their service. Mathematically, this corresponds to the arrivals $\{A_1(s), \quad 0 \leq s \leq t\}$, i.e.,

$$L_1(t) = \int_0^t (I_{\{W_1(s)>t-s\}})\ dA_1(s),$$

where $I_{\{W_1(s)>t-s)\}}$ denotes the number of patients who entered the queue in front of the triage node at time s and are still waiting for service at time t, $0 \leq s \leq t$.

By using Campbell–Mecke formula in [7] for taking expectations of stochastic integrals, we get the average number of patients present at the triage node at time t, i.e.,

$$E(L_1(t)) = \int_0^t (P\{W_1(s) > t - s\})\ \lambda_1(s)\ ds, \tag{8}$$

where $E(I_{\{W_1(s)>t-s\}}) = P\{W_1(s) > t - s\}$ and $E(dA_1(s)) = \lambda_1(s)$ ds.

5. $L_2(t)$ denotes the number of patients at treatment area, including those in the blocking space. These patients completed their service at triage node and moved to treatment area during the interval $[0, t]$. They are either waiting to join the queue in front of the treatment area or already in the queue. This corresponds to the arrivals $\{A_2(s), \quad 0 \leq s \leq t\}$, i.e.,

$$L_2(t) = \int_0^t (I_{\{W_2(s)>t-s\}} + I_{\{B(s)>t-s\}})\ dA_2(s),$$

where $I_{\{W_2(s)>t-s)\}}$ represents the number of patients who entered the queue of the treatment area at time s and are still waiting for service at time t, $0 \leq s \leq t$. Similarly, $I_{\{B(s)>t-s\}}$ denotes the number of patients who entered the blocking space in front of the treatment area at time s and are still waiting to join the queue of treatment area at time t.

While applying expectations on both sides, Campbell–Mecke formula together with additive property of expectation we get $E(I_{\{W_2(s)>t-s\}} + I_{\{B(s)>t-s\}}) = P\{W_2(s) > t - s\} + P\{B(s) > t - s\}$. Therefore,

$$E(L_2(t)) = \int_0^t \big(P\{W_2(s) > t - s\} + P\{B(s) > t - s\}\big)\ p\ \mu_1(s)\ ds. \tag{9}$$

This represents the average number of patients present in both the blocking space and the queue of the treatment area at time t.

6. $Z_1(t)$ represents the time required to triage all patients who arrived at first node up to time t, i.e.

$$Z_1(t) = \int_0^t I_{\{W_1(s)>t-s\}} \; V_1(s) \; dA_1(s) + \int_0^t \frac{V_1(s)^2}{2} \; dA_1(s).$$

While applying expectations on both sides, Campbell–Mecke formula, we get,

$$E(Z_1(t)) = \int_0^t P\{W_1(s) > t-s\} \; E(V_1(s)) \; E(A_1(s)) \; ds + \int_0^t \frac{E(V_1(s)^2)}{2} \; E(A_1(s))ds.$$

Introducing another term, squared coefficient of variation $c^2 = Var(V(s))/E(V(s))^2$. Using the formula of variance, this can be rewritten as,

$$c^2 = E(V(s)^2) \; (E(V(s)))^2 - 1 = \frac{E(V(s)^2)}{\mu(s)^2} - 1. \tag{10}$$

Therefore the average workload at the triage node at time t can be represented as,

$$E(Z_1(t)) = \int_0^t P\{W_1(s) > t-s\} \; \frac{\lambda_1(s)}{\mu_1(s)} \; ds + \int_0^t \frac{c_1^2+1}{2} \; \frac{\lambda_1(s)}{\mu_1(s)^2} ds, \tag{11}$$

where c_1^2 is the squared coefficient of variation or relative variability in service times at triage node.

7. $Z_2(t)$ denotes the time required to complete the consultation of patients who arrived up to time t after triage, taking into account the queue and blocking space, i.e.

$$Z_2(t) = \int_0^t \left(I_{\{W_2(s)>t-s\}} + I_{\{B(s)>t-s\}} \right) V_2(s) \; dA_2(s) + \int_0^t \frac{V_2(s)^2}{2} \; dA_2(s).$$

By using campell-Mecke formula and additive property of expectation, $E(Z_2(t))$ can be written as,

$$E(Z_2(t)) = \int_0^t \left(P\{W_2(s) > t-s\} + P\{B(s) > t-s\} \right) E(V_2(s)) E(A_2(s)) ds + \int_0^t \frac{E(V_2(s)^2)}{2} \; E(A_2(s))ds.$$

Let c_2^2 be the squared coefficient of variation of service times in treatment node and applying equation (10),

$$E(Z_2(t)) = \int_0^t \left(P\{W_2(s) > t-s\} + P\{B(s) > t-s\} \right) \frac{\lambda_2(s)}{\mu_2(s)} \; ds + \int_0^t \frac{c_2^2+1}{2} \; \frac{\lambda_2(s)}{\mu_2(s)^2} ds \tag{12}$$

This represents the average workload at the treatment node, taking into account the patients in the blocking space and the queue at time t.

2.2. Blocking Before Service (BBS)

Here, we examine the application of the BBS mechanism in a hospital emergency department by modelling it with a two-station tandem queueing network. This model, illustrated in Figure 2, is built under the non-stationary Markovian assumption, and patients are served according to the FCFS discipline.

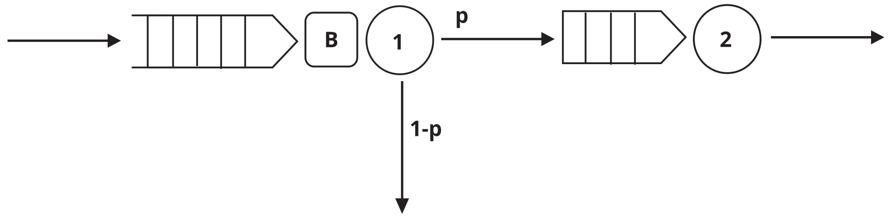


Figure 2: *A BBS two station tandem queueing network model*

Similar to the BAS tandem model, $\{A_i(t), t \geq 0\}, i = 1, 2$ is the arrival process with arrival rates $\lambda_1(t)$ and $\lambda_2(t) = p\,\mu_1(t)$. p represents the probability of moving to treatment after triage, while patients leave the system with probability $1 - p$. $\{V_i(t), t \geq 0\}, i = 1, 2$ is the service requirement of a patient arriving at the triage node and treatment node at time t with service processing rate at time t, $\mu_i(t),\ i = 1, 2$. An infinite waiting room is available for patients at the triage node, whereas the treatment node has a finite waiting room with a maximum capacity of K.

Before providing service from triage node, nurse checks whether the waiting room in front of treatment area is saturated or not. If it contains less than K patients, triage continues. If the waiting room has attained maximum K patients (saturated), stops service at triage node until the next waiting room can accommodate a new patient.

The following are the formulations of some transient performance measures related to the BBS tandem model considered here.

1. $\{W_1(t), t \geq 0\}$ is the waiting time of a patient who arrives at triage node at time t. Since the queue capacity is infinite, the probability distribution of waiting time is defined similar to BAS system, i.e.,

$$P(W_1(s) > t - s) = \rho_1 e^{-((1-\rho_1)\Lambda_{t,1}(s))/\rho_1},$$

where $\Lambda_{t,1}(s) = \Lambda_1(t) - \Lambda_1(s)$ and $\Lambda_1(.)$ is defined in (3).

2. $\{W_2(t), t \geq 0\}$ is the waiting time of a patient who arrives at the queue of treatment area at time t. Since the queue capacity is finite, i.e., $M(t)/M(t)/1/K$, the probability that the waiting time of the patient who arrives at time s, is larger than $t - s$, for $0 \leq s \leq t$ is,

$$P(W_2(s) > t - s) = \sum_{n=1}^{K-1} Q_n \sum_{1=0}^{n-1} \frac{(\frac{\Lambda_{t,2}(s)}{\rho_2})^i\ e^{\left(\frac{\Lambda_{t,2}(s)}{\rho_2}\right)}}{i!},$$

where $\Lambda_{t,2}(s) = \Lambda_2(t) - \Lambda_2(s)$ and $\Lambda_1(.)$ is defined in (7).

3. $L_1(t)$ represents the number of patients present at the triage node, including those blocked before service. These patients arrived during the interval $[0, t]$, i.e., $\{A_1(s),\quad 0 \leq s \leq t\}$ and have not yet completed their service. They have not completed their service at the triage node, either because they are in the queue or are blocked due to capacity constraints at the treatment node. Therefore,

$$L_1(t) = \int_0^t (I_{\{W_1(s)>t-s\}} + I_{\{B(s)>t-s\}})\ dA_1(s),$$

where $I_{\{W_1(s)>t-s)\}}$ represents the number of patients who entered the queue of the triage node at time s and are still waiting for service at time t, $0 \leq s \leq t$. Similarly, $I_{\{B(s)>t-s\}}$

denotes the number of patients who entered the blocking space in front of the triage node at time s and are still waiting for triage at time t. When taking expectations,

$$E(L_1(t)) = \int_0^t \big(P\{W_1(s) > t - s\} + P\{B(s) > t - s\}\big)\ \lambda_1(s)\ ds. \tag{13}$$

This represents the average number of patients present at the triage node at time t, including those blocked before service.

4. $L_2(t)$ represents the number of patients present at treatment node at time t. These patients moved after triage to treatment node during the interval $[0, t]$ and have not yet completed their service. Mathematically, this corresponds to the arrivals $\{A_1(s),\ \ 0 \leq s \leq t\}$, i.e.,

$$L_2(t) = \int_0^t (I_{\{W_2(s)>t-s\}})\ dA_2(s),$$

where $I_{\{W_2(s)>t-s)\}}$ represents the number of patients who entered the queue of the treatment area at time s and are still waiting for service at time t, $0 \leq s \leq t$. Then the average number of patients present in the queue of the treatment area at time t is,

$$E(L_2(t)) = \int_0^t \big(P\{W_2(s) > t - s\}\big)\ p\ \mu_1(s)\ ds. \tag{14}$$

5. $Z_1(t)$ represents the time required to triage all patients who arrived at first node up to time t, including those blocked patients, i.e.,

$$Z_1(t) = \int_0^t \big(I_{\{W_1(s)>t-s\}} + I_{\{B(s)>t-s\}}\big)\ V_1(s)\ dA_1(s) + \int_0^t \frac{V_1(s)^2}{2}\ dA_1(s).$$

Then the average workload at the triage node, taking into account the patients in the blocking space and the queue at time t can be represented as,

$$E(Z_1(t)) = \int_0^t \big(P\{W_1(s) > t - s\} + P\{B(s) > t - s\}\big)\ \frac{\lambda_1(s)}{\mu_1(s)}\ ds + \int_0^t \frac{c_1^2+1}{2}\ \frac{\lambda_1(s)}{\mu_1(s)^2} ds. \tag{15}$$

6. $Z_2(t)$ denotes the time required to complete the consultation of patients who arrived up to time t after triage, i.e.,

$$Z_2(t) = \int_0^t (I_{\{W_2(s)>t-s\}})\ V_2(s)\ dA_2(s) + \int_0^t \frac{V_2(s)^2}{2}\ dA_2(s).$$

Then the average workload at the treatment node at time t can be represented as,

$$E(Z_2(t)) = \int_0^t P\{W_2(s) > t - s\}\ \frac{\lambda_2(s)}{\mu_2(s)}\ ds + \int_0^t \frac{c_2^2+1}{2}\ \frac{\lambda_2(s)}{\mu_2(s)^2} ds, \tag{16}$$

where c_i^2, $i = 1, 2$ is the coefficient of variation of service process in station i.

3. Numerical Study

We consider a two-station tandem network with non-stationary Markovian queues, in which first station has infinite queue capacity and second station has finite queue capacity. We compare BAS system and BBS system by computing the transient performance measures. First, we outline the prerequisites for conducting the numerical study.

1. A more realistic choice for the arrival rate function would be a sinusoidal or periodic function, which is useful for modelling daily or weekly fluctuations in patient arrivals. However, for simplicity, we choose the identity function as the external arrival rate.

   Let the time-dependent arrival rate of patients to the triage node (external arrival rate), $\lambda_1(t)$ be the identity function, t, $t \geq 0$.

2. Transition rate or arrival rate of patients from triage station to treatment station, $\lambda_2(t)$ is,

$$\lambda_2(t) = p\ \mu_1(t) = p\ \lambda_1(t)/\rho_1 \tag{17}$$

   where p is the transition probability from station 1 to 2. Here we take $p = 0.75$.

3. The squared coefficient of variation of service time (c_i^2), $i = 1, 2$ appears in expressions of virtual workload. Since we are considering a Markovian queueing model, c_i^2 is assumed to be 1.

4. In the BAS system, blocking occurs only if a customer completes service at the triage node and attempts to move towards treatment node, but finds the queue is full. Therefore, the probability depends on the queue capacity and traffic intensity of the treatment area.

   For a BAS system, the blocking probability is defined by,

$$P(B_{BAS}) = \frac{(1-\rho_2)\rho_2^K}{1-\rho_2^{K+1}}, \tag{18}$$

   where K is the queue capacity and ρ_2 is the traffic intensity of the treatment area. Gross and Haris [9] and Ziya [17] developed formulations for blocking probability.

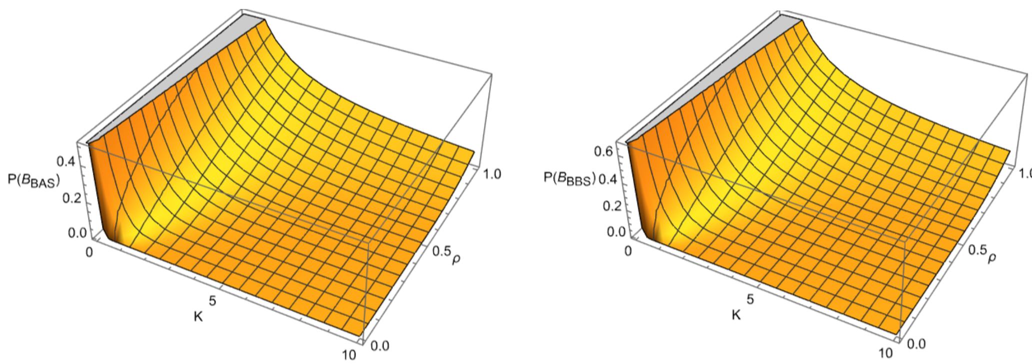
   Now we present an approximation for the blocking probability in the BBS system. In the BBS system, blocking occurs before the triage starts based on the availability of space in queue of treatment area. This means all patients might be blocked regardless of whether they would have moved for treatment or left the system after triage. As a result blocking probability is inflated.

   Since $P(B_{BAS})$ is applying only to transitioning customers, we can approximate $P(B_{BBS})$ by scaling $P(B_{BAS})$ with the proportion of transition, p. Here we assume that patient's decision to leave the system after triage or transition is independent of the congestion in the treatment area, so it only depends on the service at triage. $P(B_{BAS})$ is already calculated on transitioned patients, while formulating $P(B_{BBS})$, we need to undo this effect by dividing the proportion of transition, i.e.,

$$P(B_{BBS}) \approx \frac{P(B_{BAS})}{p} = \frac{(1-\rho_2)\rho_2^K}{(1-\rho_2^{K+1})\ p}. \tag{19}$$

   In this study, we have chosen a constant queue capacity and a constant traffic intensity using the rate-matching control principle. Consequently, the parameters in the above expressions are time-invariant, making them applicable to a non-stationary model.

   The Figure 3 illustrates the relationship between the queue capacity and traffic intensity and how their combined variation influences the blocking probabilities.



(a) BAS (b) BBS

Figure 3: *Blocking probability for queue capacity $K = 0$ to 10 and traffic intensity $\rho = 0$ to 1.*

5. In this study, we consider four cases, A, B, C, and D, by taking arbitrary values for traffic intensities and queue capacity of the second station, as shown in the Table 1. The blocking probabilities corresponding to each case are calculated and included in the table. The stations with traffic intensity close to 1 are considered bottleneck stations.

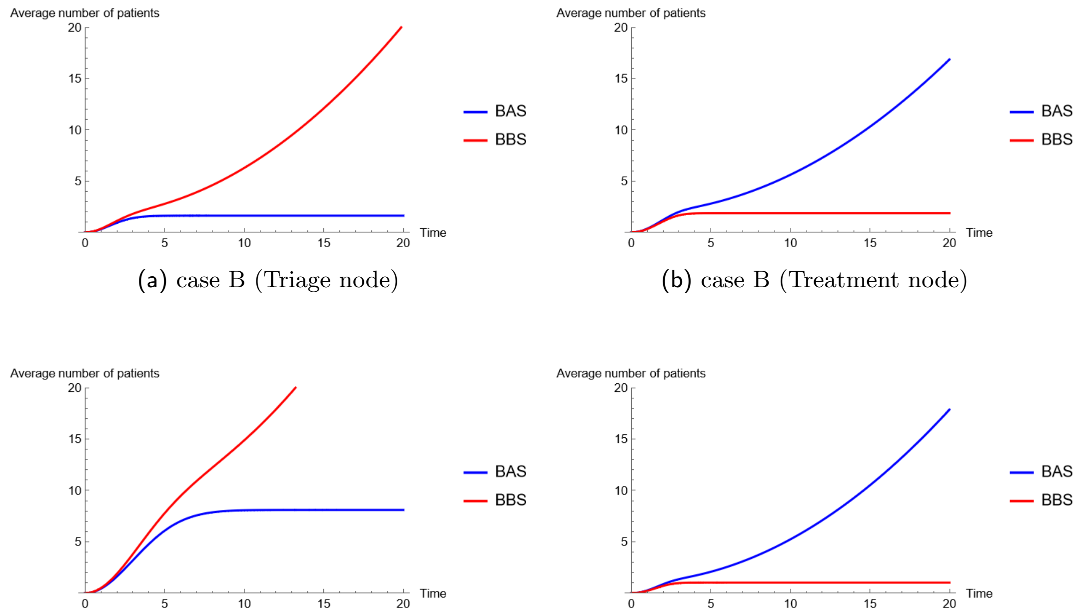
cases	ρ_1	ρ_2	K	$P(B_{BAS})$	$P(B_{BBS})$
A	0.80	0.60	4	0.056	0.074
B	0.70	0.90	8	0.070	0.093
C	0.90	0.90	6	0.101	0.136
D	0.90	0.90	10	0.050	0.067

Table 1: *Four cases of traffic intensities, queue capacity and corresponding blocking probabilities.*

The probability of blocking is observed to be relatively high in cases where the traffic intensity at the second station increases significantly and the queue capacity decreases. Among the cases considered, cases A and D exhibit the lowest blocking probabilities. In contrast, Cases B and C show comparatively higher probabilities of blocking due to the combination of high traffic intensity and low queue capacity. Therefore, we have focused on presenting numerical illustrations for these two cases to obtain critical observations.

Figure 4 illustrates the number of patients at both nodes under the two mechanisms. Blocking after service occurs at the treatment node, while blocking before service occurs at the triage node. As a result, the number of patients increases at the corresponding nodes with time. When the traffic intensity at a node is high, it leads to a significant increase in the number of patients. A similar trend is observed in the average workload, as shown in Figure 5. Bottleneck nodes exhibit a relatively higher workload compared to others. In both cases, the treatment node experiences a higher workload under the BAS and BBS mechanisms, as it serves as a bottleneck.

Figure 6 illustrates the average number of patients in the system over time. This analysis considers the four cases discussed in Table 1. In all cases, the total number of patients is consistently higher in the BBS system.



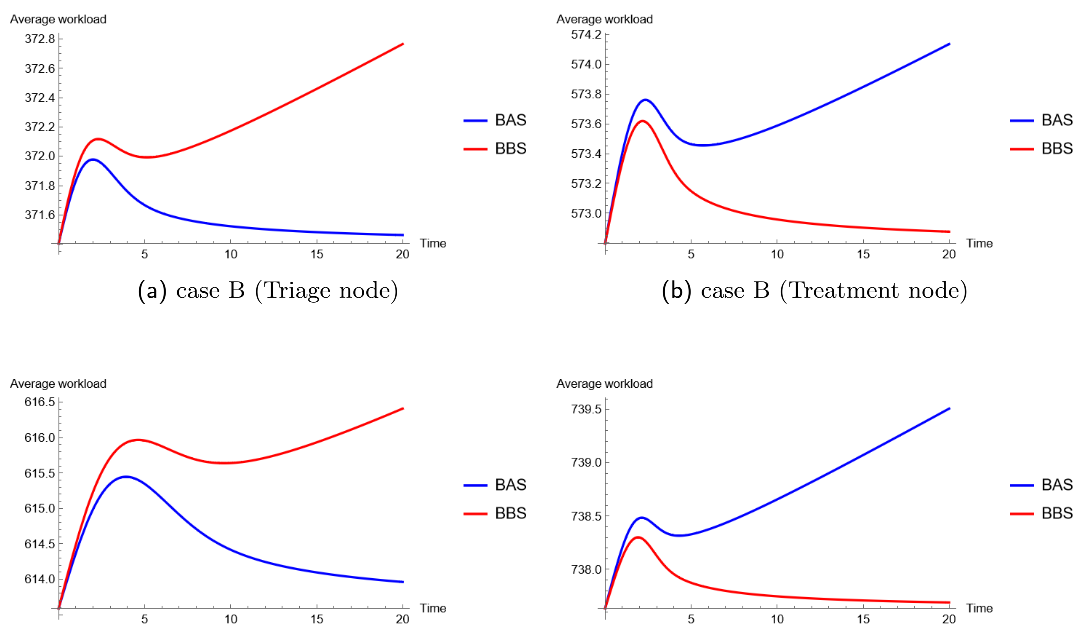
(a) case B (Triage node)

(b) case B (Treatment node)

(c) case C (Triage node)

(d) case C (Treatment node)

Figure 4: *Average number of patients under BAS and BBS mechanisms for cases B and C.*



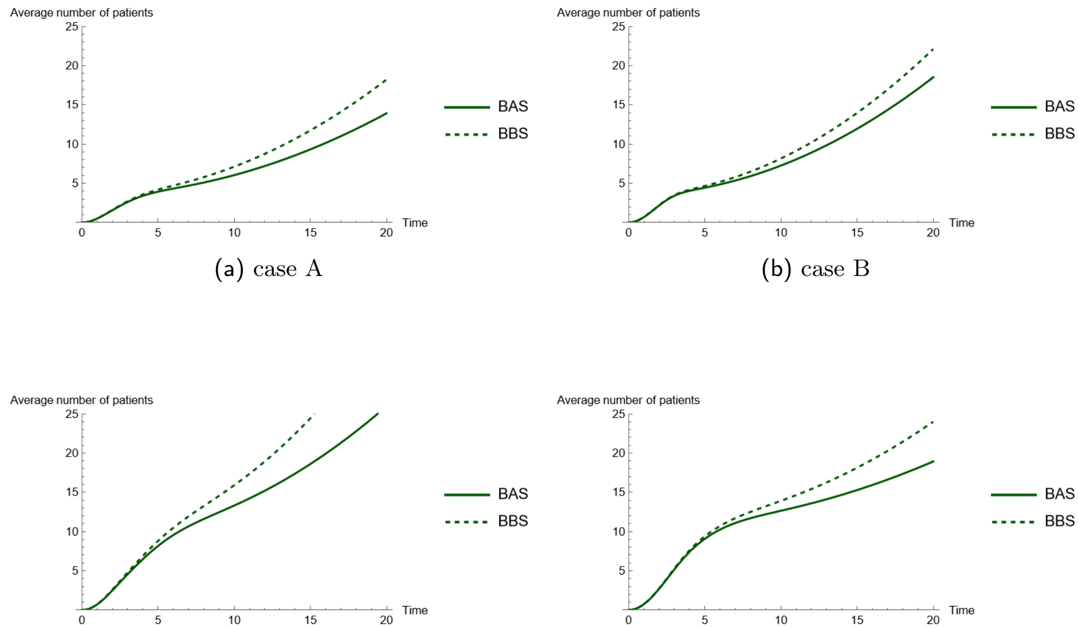
(a) case B (Triage node)

(b) case B (Treatment node)

(c) case C (Triage node)

(d) case C (Treatment node)

Figure 5: *Average workload under BAS and BBS mechanisms for cases B and C.*



(a) case A (b) case B

(c) case C (d) case D

Figure 6: *Average number of patients in the system (including both stations) for the four cases considered in this study.*

4. Conclusion

Finite capacity queues are a realistic and common feature of many real-world service systems, such as hospitals, call centres, and manufacturing units, due to physical and budgetary constraints. However, in most of the theoretical studies of queues, infinite capacity is often assumed to simplify the analysis and results. In this study, we examined various blocking mechanisms that can be applied in queueing systems with capacity restrictions. Most commonly, BAS and BBS blocking mechanisms are used in such systems. Here, we have modelled a time-dependent hospital emergency department using a tandem network of two stations and derived transient performance measures based on these mechanisms. These measures enable an effective comparison of the BAS and BBS mechanisms through numerical study. The results highlight the impact of capacity restrictions on both mechanisms, demonstrating how they influence the time-varying number of patients and the virtual workload in the system. In our model, patients who do not need immediate emergency treatment can leave after triage. But in the BBS system, these patients also get blocked, causing unwanted congestion in the system. So, BAS is slightly better than BBS in the above scenario. Generally, BAS is preferred when intermediate congestion can be managed, whereas BBS is preferred when smooth flow of entities through the system is more important. This study examined a two-station tandem network of single-server queues with capacity restriction on the intermediate queue. This framework can be generalised to a k-station tandem network, and the single-server queues can be extended to multi-server systems. Deriving more rigorous mathematical expressions for performance measures under both blocking mechanisms would enable a more extensive study of these systems. These are some directions for future research.

References

[1] Avi-Itzhak, B. (1965). A sequence of service stations with arbitrary input and regular service times. Management Science, 11(5), 565-571. doi: 10.1287/mnsc.11.5.565

[2] Avi-Itzhak, B. and Yadin, M. (1965). A sequence of two servers with no intermediate queue. Management Science, 11(5), 553-564. doi: 10.1287/mnsc.11.5.553

[3] Avi-Itzhak, B. and Halfin, S. (1993). Servers in tandem with communication and manufacturing blocking. Journal of Applied Probability, 30(2), 429-437. doi: 10.2307/3214851

[4] Avi-Itzhak, B. and Levy, H. (1995). A sequence of servers with arbitrary input and regular service times revisited: in memory of Micha Yadin, Management Science, 41(6), 1039-1047. doi: 10.1287/mnsc.41.6.1039

[5] Balsamo, S., de Nitto Personé, V., and Onvural, R. (2001). Analysis of queueing networks with blocking. Springer Science & Business Media. doi: 10.1007/978-1-4757-3345-7

[6] de Bruin, A. M., van Rossum, A. C. and Visser, M. C. et al. (2007). Modeling the emergency cardiac in-patient flow: an application of queuing theory. Health Care Management Science, 10(2), 125-137. doi: 10.1007/s10729-007-9009-8

[7] Fralix, B.H. and Riaño, G. (2010). A new look at transient versions of little's law, and m/g/1 preemptive last-come-first-served queues. Journal of Applied Probability, 47(2), 459-473. doi: 10.1239/jap/1276784903

[8] Frein, Y. and Dallery, Y. (1989). Analysis of cyclic queueing networks with finite buffers and blocking before service. Performance Evaluation, 10(3), 197–210. doi: 10.1016/0166-5316(89)90010-2

[9] Gross, D. and Harris, C. (1998) Fundamentals of Queuing Theory. John Wiley, Chichester. doi: 10.1002/9781119453765

[10] Meerkov, S. and Yan, C. B. (2016). Production lead time in serial lines: Evaluation, analysis, and control. IEEE Transactions on Automation Science and Engineering, 13(2), 663–675. doi: 10.1109/TASE.2014.2365108

[11] Prabhu, N. (1967). Transient behaviour of a tandem queue. Management Science, 13(9), 631–639. doi: 10.1287/mnsc.13.9.631

[12] Ramesh, A. and Manoharan, M. (2024). Transient Behaviour of Time-Varying Tandem Queueing Networks. OPSEARCH. doi: 10.1007/s12597-024-00790-0

[13] Seo, D. W., Lee, H. C. and Ko, S. S. (2008). Stationary waiting times in m-node tandem queues with communication blocking. Management Science and Financial Engineering, 14(1), 23-34.

[14] Seo, D. W. and Lee, H. C. (2011). Stationary waiting times in m-node tandem queues with production blocking. IEEE Transactions on Automatic Control, 56(4), 958-961. doi: 10.1109/TAC.2011.2105290

[15] Suri, R and Diehl, G. (1984). A new 'building block' for performance evaluation of queueing networks with finite buffers. ACM SIGMETRICS Performance Evaluation Review, 12, 134–142. doi: 10.1145/800264.809321

[16] Whitt, W. (2015). Stabilizing performance in a single-server queue with time-varying arrival rate. Queueing Systems, 81(4), 341–378. doi: 10.1007/s11134-015-9462-x

[17] Ziya, S. (2008). On the Relationships Among Traffic Load, Capacity, and Throughput for the M/M/1/m, M/G/1/m-PS, and M/G/c/c Queues. IEEE Transactions on Automatic Control, 53(11), 2696-2701. doi: 10.1109/TAC.2008.2007173.

[18] Zychlinski, N., Mandelbaum, A. and Momcilovic, P. (2018a). Time-varying tandem queues with blocking: modeling, analysis, and operational insights via fluid models with reflection. Queueing Systems, 89(1), 15–47. doi: 10.1007/s11134-018-9578-x

[19] Zychlinski, N., Momcilovic, P. and Mandelbaum, A. (2018b). Time-varying many-server finite-queues in tandem: Comparing blocking mechanisms via fluid models. Operations Research Letters 46(5), 492-499. doi: 10.1016/j.orl.2018.07.002

Croatian Operational Research Review

Information for authors

Journal homepage: http://www.hdoi.hr/crorr-journal

Aims and Scope: The aim of the Croatian Operational Research Review journal is to provide high quality scientific papers covering the theory and application of operational research and related areas, mainly quantitative methods and machine learning. The scope of the journal is focused, but not limited to the following areas: combinatorial and discrete optimization, integer programming, linear and nonlinear programming, multiobjective and multicriteria programming, statistics and econometrics, macroeconomics, economic theory, games control theory, stochastic models and optimization, banking, finance, insurance, simulations, information and decision support systems, data envelopment analysis, neural networks and fuzzy systems, and practical OR and application.

Papers blindly reviewed and accepted by two independent reviewers are published in this journal.

Author Guidelines: There is no submission fee for publishing papers in this journal. All manuscripts should be submitted via our online manuscript submission system at the journal homepage. A user ID and password can be obtained on the first visit. The initial version of the manuscript should be written using LaTex program or Word program, and submitted as a single PDF file. After the reviewing process, if the paper is accepted for publication in this journal, the corresponding author will be asked to submit proofreading confirmation along with signed copyright form. CRORR uses the Turnitin Similarity Check plagiarism software to detect instances of overlapping and similar text in submitted manuscripts. Any copying of text, figures, data or results of other authors without giving credit is defined as plagiarism and is a breach of professional ethics. Such papers will be rejected and other penalties may be accessed.

Electronic Presentations: The journal is also presented electronically on the journal homepage. All issues are also available at the Hrcak database – Portal of science journals od the Republic of Croatia, with full-text search of individual, at

http://hrcak.srce.hr/crorr

Offprints: The corresponding author of the published paper will receive a single print copy of the journal. All published papers are freely available online.

Publication Ethics and Malpractice Statement

Croatian Operational Research Review journal is committed to ensuring ethics in publication and quality of articles. Conformance to standards of ethical behaviour is therefore expected of all parties involved: Authors, Editors, Reviewers, and the Publisher. Our ethic statements are based on COPE's Best Practice Guidelines for Journals Editors.

Publication decisions. The editor of the journal is responsible for deciding which of the articles submitted to the journal should be published. The editor may be guided by the policies of the journal's editorial board and constrained by such legal requirements as shall then be in force regarding libel, copyright infringement and plagiarism. The editor may confer with other editors or reviewers in making this decision.

Fair play. An editor will at any time evaluate manuscripts for their intellectual content without regard to race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy of the authors.

Confidentiality. The editor and any editorial staff must not disclose any information about a submitted manuscript to anyone other than the corresponding author, reviewers, potential reviewers, other editorial advisers, and the publisher, as appropriate.

Disclosure and conflicts of interest. Unpublished materials disclosed in submitted manuscript must be used in an editor's own research without the express written consent of the author.

Duties of Reviewers

Contribution to Editorial Decisions. Peer review assists the editor in making editorial decisions and through the editorial communications with the author may also assist the author in improving the paper.

Promptness. Any selected referee who feels unqualified to review the research reported in a manuscript or knows that its prompt review will be impossible should notify the editor and excuse himself from the review process.

Confidentiality. Any manuscripts received for review must be treated as confidential documents. They must not be shown to or discussed with others except at authorized by the editor.

Standards of Objectivity. Reviews should be conducted objectively. Personal criticism of the author is inappropriate. Referees should express their views clearly with supporting arguments.

Acknowledgement of Sources. Reviewers should identify relevant published work that has not been cited by the authors. Any statement that an observation, derivation, or argument had been previously reported should be accompanied by the relevant citation. A reviewer should also call to the editor's attention any substantial similarity or overlap between the manuscript under consideration and any other published paper of which they have personal knowledge.

Disclosure and Conflict of Interest. Privileged information or ideas obtained through peer review must be kept confidential and not used for personal advantage. Reviewers should not consider manuscripts in which they have conflicts of interest resulting from competitive, collaborative, or other relationships or connections with any of the authors, companies, or institutions connected to the papers.

Duties of Authors

Reporting Standards. Authors' manuscripts of original research should present an accurate of the work performed as well as an objective discussion of its significance. Underlying data should be represented accurately in the paper. A paper should contain sufficient detail and references to permit others to replicate the work. Fraudulent or knowingly inaccurate statements constitute unethical behaviour are unacceptable.

Data Access and Retention. When submitting their article for publication in this journal, the authors confirm the statement that all data in article are real and authentic. Authors should be prepared to provide the raw data in connection with a paper for editorial review, and should be prepared to provide public access to such data (consistent with the ALPSP-STM Statement on Data and Databases), if practicable, and should in any event be prepared to retain such data for a reasonable time after publication. In addition, all authors are obliged to provide retractions or corrections of mistakes if any.

Originality and Plagiarism. The authors should ensure that they have written entirely original works, and if the authors have used the work and/or words of others that has been appropriately cited or quoted. Multiple, Redundant or Concurrent Publication An author should not in general publish manuscripts describing essentially the same research in more than one journal or primary publication. Submitting the same manuscript to more than one journal concurrently constitutes unethical publishing behaviour and is unacceptable.

Acknowledgement of Sources. Proper acknowledgement of the work of others must always be given. Authors should cite publications that have been influential in determining the nature of the reported work.

Authorship of the Paper. Authorship should be limited to those who have made a significant contribution to the concept, design, execution, or interpretation of the reported study. All those who have made significant contributions should be listed as co-authors. Where there are others who have participated in certain substantive aspects of the research project, they should be acknowledged or listed as contributors. The corresponding author should ensure that all appropriate co-authors and no inappropriate co-authors are included on the paper, and that all co-authors have seen and approved the final version of the paper and have agreed to its submission for publication.

Disclosure and Conflicts of Interest. All authors should disclose in their manuscript any financial or other substantive conflict of interest that might be construed to influence the results or interpretation of their manuscript. All sources of financial support for the project should disclosed. When an author discovers a significant error or inaccuracy in his/her own published work, it is the author's obligation to promptly notify the journal editor or publisher and cooperate with editor to retract or correct the paper.

Open Access Statement. Croatian Operational Research Review is an open access journal which means that all content is freely available without charge to the user or his/her institution. User are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles in this journal without asking prior permission form the publisher or the author. This is in accordance with the BOAI definition of open access. The journal has an open access to full text of all papers through our Open Journal System (http://hrcak.srce.hr/ojs/index.php/crorr) and the Hrcak database (http://hrcak.srce.hr/crorr).

Operational Research

Operational Research (OR) is an interdisciplinary scientific branch of applied mathematics that uses methods such as mathematical modelling, statistics, and algorithms to achieve optimal or near optimal solutions to complex problems. It is also known as Operations Research, and is related to Management Science, Industrial Engineering, Decision Making, Statistics, Simulations, Business Analytics, and other areas.

Operational Research is typically concerned with optimizing some objective function, or finding non-optimal but satisfactory solutions. It generally helps management to achieve its goals using scientific methods. Some of the primary tools used by operations researchers are optimization, statistics, probability theory, queuing theory, game theory, graph theory, decision analysis, and simulation.

Areas/Fields in which Operational Research may be applied can be summarized as follows:

- Assignment problem,
- Business analytics,
- Decision analytics,
- Econometrics,
- Linear and nonlinear programming,
- Inventory theory,
- Optimal maintenance,
- Scheduling,
- Stochastic process, and
- System analysis.

Operational Research has improved processes in business and all around us. It is used in everyday situations, such as design of waiting lines, supply chain optimization, scheduling of buses or airlines, telecommunications, or global resources planning decisions.

Operational Research is an important area of research because it enables the best use of available resources. New methods and models in operational research are to be developed continuously in order to provide adequate solutions to process enormous information growth resulted from rapid technology development in today's new economy. By improving processes it enables high-quality products and services, better customer satisfaction, and better decision making. There is no doubt that operational research contributes to the quality of life and economic prosperity on microeconomic, macroeconomic, and global level.

Croatian Operational Research Society

Croatian Operational Research Society (CRORS) was established in 1992. as the only scientific association in Croatia specialized in operational research. The Society has more than 150 members and its main mission is to promote operational research in Croatia and worldwide for the benefit of science and society. This mission is realized through several goals:

- to encourage collaboration of scientists and researchers in the area of operational research in Croatia and worldwide through seminars, conferences, lectures and similar ways of collaborations,
- to organize the International Conference on Operational Research (KOI),
- to publish a scientific journal Croatian Operational Research Review (CRORR).

Since 1994, CRORS is a member of The International Federation of Operational Research Societies (IFORS), an umbrella organization comprising the national Operations Research societies of over forty-five countries from four geographical regions: Asia, Europe, North America and South America. CRORS is also a member of the Association of European Operational Research Societies (EURO) and actively participates in international promotion of operational research.

If you are a researcher, academic teacher, student or practitioner interested in developing and applying operations research methods, you are welcome to become a member of the CRORS and join our OR community.

If you want to become a member of the CRORS, please fill in the application form available at

http://www.hdoi.hr