

Peculiarities of gender disambiguation and ordering of non-English authors' names for Economic papers beyond core databases^①

Olesya Mryglod^{1†}, Serhii Nazarovets², Serhiy Kozmenko³

¹Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine, 1 Svientsitskii St., 79011 Lviv, Ukraine

²Borys Grinchenko Kyiv University, 18/2 Bulvarno-Kudriavska Str., 04053 Kyiv, Ukraine

³University of Social Sciences Spoleczna Akademia Nauk, 9 Sienkiewicza St., 90–113 Lodz, Poland

Abstract

Purpose: To supplement the quantitative portrait of Ukrainian Economics discipline with the results of gender and author ordering analysis at the level of individual authors, special methods of working with bibliographic data with a predominant share of non-English authors are used. The properties of gender mixing, the likelihood of male and female authors occupying the first position in the authorship list, as well as the arrangements of names are studied.

Design/methodology/approach: A data set containing bibliographic records related to Ukrainian journal publications in the field of Economics is constructed using Crossref meta-data. Partial semi-automatic disambiguation of authors' names is performed. First names, along with gender-specific ethnic surnames, are used for gender disambiguation required for further comparative gender analysis. Random reshuffling of data is used to determine the impact of gender correlations. To assess the level of alphabetization for our data set, both Latin and Cyrillic versions of names are taken into account.

Findings: The lack of well-structured metadata and the poor use of digital identifiers lead to numerous problems with automatization of bibliographic data pre-processing, especially in the case of publications by non-Western authors. The described stages for working with such specific data help to work at the level of authors and analyse, in particular, gender issues. Despite the larger number of female authors, gender equality is more likely to be reported at the individual level for the discipline of Ukrainian Economics. The tendencies towards collaborative or solo-publications and gender mixing patterns are found to be dependent on the

Citation: Mryglod, O., Nazarovets, S., & Kozmenko, S. (2023). Peculiarities of gender disambiguation and ordering of non-English authors' names for Economic papers beyond core databases. *Journal of Data and Information Science*, 8(1), 1–18. <https://doi.org/10.2478/jdis-2023-0001>

Received: Jul. 7, 2022

Revised: Oct. 6, 2022;

Nov. 16, 2022

Accepted: Nov. 16, 2022

[†] Corresponding author: Olesya Mryglod (Email: olesya@icmp.lviv.ua, ORCID: 0000-0003-4415-7061).

^① This paper presents the results of further exploration of Crossref data related to Ukrainian Economics research (the first part can be found in [Mryglod, O., Nazarovets, S. & Kozmenko, S. Universal and specific features of Ukrainian economic research: publication analysis based on Crossref data. *Scientometrics* 126, 8187–8203 (2021). <https://doi.org/10.1007/s11192-021-04079-7>]).



journal: the differences for publications indexed in Scopus and/or Web of Science databases are found. It has also been found that Ukrainian Economics research is characterized by rather a non-alphabetical order of authors.

Research limitations: Only partial authors' name disambiguation is performed in a semi-automatic way. Gender labels can be derived only for authors declared by full First names or gender-specific Last names.

Practical implications: The typical features of Ukrainian Economic discipline can be used to perform a comparison with other countries and disciplines, to develop an informed-based assessment procedure at the national level. The proposed way of processing publication data can be borrowed to enrich metadata about other research disciplines, especially for non-English speaking countries.

Originality/value: To our knowledge, this is the first large-scale quantitative study of Ukrainian Economic discipline. The results obtained are valuable not only at the national level, but also contribute to general knowledge about Economic research, gender issues, and authors' names ordering. An example of the use of Crossref data is provided, while this data source is still less used due to a number of drawbacks. Here, for the first time, attention is drawn to the explicit use of the features of the Slavic authors' names.

Keywords Scholarly metadata; Economics; Crossref; Open Ukrainian Citation Index; Ukraine, Gender; non-Western authors

1 Introduction

This paper contains the results that are part of a more general study (Mryglod et al., 2021), the purpose of which is to perform a large-scale quantitative analysis of the Ukrainian Economics discipline using the publication data predominantly beyond Web of Science and Scopus databases, where currently more than 150 Ukrainian journals are indexed while the National List of recognized scientific journals² includes almost 1,500 titles. Motivated by the fact that Ukrainian research is still understudied because of its poor representation in core databases (Aksnes & Sivertsen, 2019) (and this is especially true for Social Sciences and Humanities – SSH), we have made an attempt to provide a quantitative portrait of one of Ukrainian SSH disciplines using the Crossref database as an alternative data source. Our interest in Ukraine is natural, as all three authors are Ukrainians and therefore motivated to contribute to a more transparent and evidence-based management of national research. Nevertheless, we also believe that this is an interesting case study that contributes to a better understanding of the research process in the developing



countries of Eastern Europe, countries with a special historical heritage. Ukraine is characterized by non-English speaking and Cyrillic writing; this is especially true for the analysis of SSH. Economics is chosen as one of the most “visible” SSH disciplines, which is often considered as a transitional science between the “hard” and “social” sciences (Cainelli et al., 2012; Mryglod, 2012).

The quantitative analysis at the level of publications based on the Crossref data is rather straightforward. For example, an estimation of the number of authors per paper can be done even without sophisticated data pre-processing. However, the consideration of individual publication histories at the level of authors or gender analysis requires name disambiguation. Obviously, this task is a big challenge, especially when dealing with non-Western names (Gomide et al., 2017; Kim et al., 2021; Treeratpituk & Giles, 2012). Since the majority of publications in the Ukrainian Economic discipline are related to local authors (Mryglod et al., 2021) it is natural to find mainly Ukrainian first and last names in our data set. Although such peculiarities as the use of a middle name or a prespecified order of parts in composite names (Gomide et al., 2017; Treeratpituk & Giles, 2012) are not typical for Ukraine, a huge problem of transliteration (e.g., see also Müller et al., 2017) still exists. But there is another side of the coin: the so-called ethnicity can be used to improve gender disambiguation. Therefore, along with the initial motivation to contribute to the quantitative description of Ukrainian Economics discipline, a special emphasis is made on the methods of processing such specific bibliographic data.

The research questions here are related both to the methods of data pre-processing and the results of the analysis of these data:

- RQ1: What peculiarities of Ukrainian authors' names have to be taken into account during the process of name and gender disambiguation?
- RQ2: What gender proportion is typical for the Ukrainian Economics discipline, and how can it be compared with similar results for other data sets (countries)?
- RQ3: What level of alphabetization characterizes the Ukrainian Economics discipline, and is it possible to identify any gender-related distinctions?

Answering the main research questions, this paper serves also as another evidence of the usefulness of Crossref data as a potential source for bibliometric analysis. Economic publications are considered in many other studies, where the data from Web of Science or Scopus databases are exploited, see (Schläpfer, 2010; Truc et al., 2021; Vaio & Weisdorf, 2009; Wei, 2018; Zhao et al., 2016). And this is reasonable in order to assess top-impact output and reveal the research front in Economics. However, if the rest of the entire picture is needed, the potential of other sources such



as Crossref can be efficiently used. In this context, Ukraine has an advantage – a special interface called Open Ukrainian Citation Index (OUCI) was developed a few years ago. It provides a possibility to extract structured Crossref metadata related to all journals published in Ukraine (Cheberkus & Nazarovets, 2019). Moreover, all these journals are labeled by subject category according to the Ukrainian national classification scheme. While this data source is not as comprehensive as a national current research information system could be, it provides a unique opportunity to supplement knowledge about the research output of Ukraine.

The paper is organized as follows: the description of our data set is provided in Section 2; the applied name disambiguation procedure is described in Section 3; Section 4 describes the peculiarities of the gender disambiguation procedure for our data and contains the results of the gender analysis. The ordering of author's name for the Ukrainian Economics discipline is studied in Section 5; the final discussion can be found in the last Section.

2 Our data

This paper is a continuation of the authors' previous work (Mryglod et al., 2021), which analyzes Crossref data for publications in Ukrainian journal papers in the field of Economics. The same principle of collecting data is used here, i.e., Crossref publication records related to Ukrainian Economics journals. A not-for-profit membership organization Crossref^③ collects metadata for publications with registered DOI (Digital Object Identifier) numbers. Each record contains basic bibliometric elements required for DOI registration (i.e., title, publication dates, authors, source title, volume, and issue number, etc.). In addition, Crossref encourages its depositors to enrich metadata with authors' affiliations, ORCID numbers, abstracts, lists of references, funding information, etc. These metadata are publically open, license-free and distributed through Crossref tools and APIs. Since 2018, DOI registration is required for any research paper to be officially recognized in Ukraine^④, therefore, Crossref metadata can be considered as a useful source of information about the published outputs related to Ukrainian research. Moreover, a special web-interface – Open Ukrainian Citation Index (OUCI)^⑤ – was developed to efficiently import these data. In particular, a number of search filters allowing the classification of journals by their specialty are implemented into OUCI. In addition, information about the current indexing of each journal in Scopus and Web of Science is provided.

^③ <https://www.crossref.org/>

^④ See the Order of the Ministry of Education and Science of Ukraine N32 (2018, January 15).
<http://zakon.rada.gov.ua/laws/show/z0148-18>

^⑤ <https://ouci.dntb.gov.ua/en/>

The topical relevance of each journal is defined using the Speciality search filter. The following specialties are considered to be related to the Economics field (a similar subject classification is used by the State Attestation Commission of Ukraine[®]: Economics; Tax and Accounting Policy; Finance, Banking, and Insurance; Management; Marketing; Business, Entrepreneurship, and Stock Markets; Public Administration; and International Economic Relations. Only journals in the National List of recognized scientific journals are considered. To exclude multidisciplinary editions, journal disciplines (upper classification level) are limited to the following list: Social and Behavioral Sciences; Management and Administration; Public Management and Administration; International Relations.

In addition to the data available from Scopus and Web of Science, Crossref provides an important piece of the puzzle required to build the full picture of Ukrainian Economics research. The results presented in this paper are based on the updated data set: data collection is performed at the end of February 2021. Altogether, 25,933 records for papers published in Ukrainian Economics journals between 2002 and 2020 were collected (the annual publication statistics is low before 2012 and rapidly increases afterward: 97% of records correspond to the period 2013–2020). The imported records contain the following fields: Publication year; Journal ISSN; DOI; Publisher; Title; Authors' names; Number of DOI-to-DOI citations (if the information is provided by Crossref depositors); Journal is indexed in Scopus Yes/No; Journal is indexed in Web of Sciences Yes/No (up-to-date information in the last two fields is added by OUCI).

3 Name disambiguation problem

While data analysis at the level of papers is performed (Mryglod et al., 2021), many interesting questions can be put at the level of authors. To give an example, typical individual productivity or authors' collaboration patterns have to be known to set benchmarks for comparing, assessing, or detecting examples of unusual publishing behavior. What is also important, the authors' gender is typically (and in this work) inferred from the given names. Therefore, the gender label cannot be assigned if only initials are specified instead of the full name. However, merging various records related to the same person, allows us to enlarge the statistics of papers with genderized authors. For example, gender can be defined for *KAFKA S.* (?) merged with *KAFKA SOFIYA* (Female). However, the widely-known problem of name disambiguation appears if unique digital identifiers are not commonly used. With only names, it is impossible to guarantee that two identically written names correspond to



[®] Official web-page of the State Attestation Commission of Ukraine: <https://mon.gov.ua/ua/tag/atestatsiya-kadriv-vishchoi-kvalifikatsii>

the same person. The uncertainty is higher if only initials are used. But everything is even more complicated in the case of publications by authors who are not native English speakers. It is possible to find numerous alternative transliterations of Cyrillic names for the same authors in our data set. Moreover, speaking of Ukrainian names, one should take into account the tradition of “translating” given names, and sometimes even last names, into Russian. For example, an author *Bosovskaya* can be also mentioned as *Bosovska*; *Orlovskaya* – *Orlovska*; *Mostenskaya* – *Mostenska*. Many of the first names can be transliterated to English using Ukrainian or Russian Cyrillic versions; some of the most used are: *Mykola* – *Nikolay*, *Oleksandr* – *Aleksandr*, *Kateryna* – *Ekaterina*, *Olena* – *Elena*. It can be instantly noted that the names in these pairs correspond to different initials: *M* – *N*, *O* – *A*, *K* – *E*, *O* – *E*, respectively. The space of possible alternatives is also expanded by using different short versions of names: *Olena* → *Lena*, *Oleksiy* → *Alex*, *Anastasiya* → *Nastya*, *Tetyana* → *Tanya*, etc. Sometimes, the same name can be written in many ways, and each of them is automatically recognized as a separate name. Moreover, metadata for Ukrainian Economics journals can be deposited not only in English, but also in Ukrainian (Russian). Last but not least, inaccurate usage of Latin and Cyrillic alphabets is a problem. The homoglyphs – letters that look the same on a screen but are coded differently – are used arbitrarily. After all, 40 versions of the name *Eugen* are found in the data set. Taken together, all these peculiarities of metadata of Ukrainian (non-native English) publications complicate the process of disambiguating the names of authors.

Due to the numerous nuances listed above, it is too difficult to perform a full disambiguation procedure automatically. That can be done only partially and only in a semi-automatic way. The following criteria and approaches are used:

- Identical names found in different papers are considered as related to one person since the limited data set that corresponds to a particular subject area is studied. The assumption that there is a low probability of duplicated names within our data set is confirmed by manually checking randomly selected records. Of course, exceptions are possible. Authors’ records with identical first and second names are separated if both appear in the same paper.
- The existence of common co-authors for two authors is considered as an argument to merge corresponding records.
- A manually created list of Ukrainian given names together with “synonymical” forms (Latin and Cyrillic) was used to find candidates for merging⁷. A gender



label is initially assigned to each name (manually). A few examples to demonstrate the variety of names are shown below.

o *ALEKSEI; ALEKSEY; ALEKSII; ALEXEI; ALEXEJ; ALEXEY; АЛЕКСЕЙ; АЛЕКСЕY; ОЛЕКСЕY; ОЛЕКСII; ОЛЕКСIY; ОЛЕКСYI; ОЛЕКСII; ОЛЕКСIY; ОЛЕXIJ; ОЛЕXIY; ALEXEI; АЛЕКСЕЙ; ОЛЕКСИЙ; ОЛЕКСII; АЛЕКСY; ОЛЕКСII; ОЛЕXI; ОЛЕXI; ALEXSEY; ОЛЕКСIY; ОЛЕКСIY; – Male*

o *CHRISTINA; CHRISTINE; CHRYSTYNA; CRISTINA; HRISTINE; KHRYSTYNA; KRISTINA; KRISTINE; KRISTYNA; KRYSTYNA; ХРИСТИНА; КРИСТИНА; КРИСТИНА; KHRISTINA; KRISZTINA; – Female*

- The records are merged if no contradictions appear. To give an example, all names from the following list: *KAFKA S.M.; KAFKA SOFIYA; КАФКА С.М.; KAFKA S.M.; KAFKA SOFIIA; KAFKA S.* are merged to get a single author's record. But an ambiguity exists for authors from the list *VERHUN A.; VERGUN ANDRIJ IVANOVYCH; VERHUN A.; VERHUN ANTONINA; VERHUN ANDRIJ; VERHUN A.*[®] corresponds either to *ANDRIJ* or to *ANTONINA* – no merging is performed in this case.

The list of authors' names was processed using the own Python code to find the list of candidates for merging and to mark them as more or less probable. The final merging was manually confirmed using the results of this preliminary automatic procedure. Additional manual checks were performed for particular cases, where candidates are considered as important players due to a large number of publications or co-authors. Merging was not performed for the pairs where ambiguity remains, but even so, the initial set of 31.5 thousand authors' records was reduced to 23,094.

4 Gender disambiguation and analysis

As mentioned before, the gender label of an author is inferred from his/her given name. Since the majority of authors are from Ukraine (Mryglod et al., 2021), Slavic first names are predominantly found in our data set. Besides the list of Slavic names manually labeled by gender, free web resource Genderize[®] was partially used to detect gender for non-Slavic names[®]. Thus, 54.5% of 23,094 author records were marked by gender: 7,748 (33.5%) females and 4,865 (21%) males. According to this,

[®] It seems like the same name appears twice, since Cyrillic "A" and Latin "A" are used for the author's initial.

[®] Genderize.io | Determine the gender of a first name. <https://genderize.io/#overview>

[®] Only results for the names occurring at least 10 times with a probability of more than 0.9 were taken into account.



females appear in our data set approximately 1.59 times more often.

Some typical endings of Slavic surnames can be considered as gender-specific. Author records genderized on the previous step were used as a validation subset in order to check whether surnames' endings are distinctive enough for our data set. The gender of 19.85% females was repeatedly recognized using gender-specific endings of last names: *OVA, EVA, OBA, EBA, AЯ, AYA, AIA, INA, IHA*. In their turn, 18.19% of males were recognized using the following list of endings: *OV, EV, OB, EB, CB, YI, YJ, KY, KII, KIJ*. As it can be seen, such an approach is almost equally accurate for both genders: males are slightly less recognizable. Only eight (0.1%) female and seven (0.14%) male authors are re-marked incorrectly. Such a few examples can be considered as exceptions: male surnames with female-like endings *DIBROVA, COBA (SOVA), ELLAIA, JAYA, GLOVA, GECHBAIA, BOVA* and female surnames with male-like endings *KYI, MYSHELOV, GLAMBOSKY, LAZANYI, BOKII, URSAKII, SUSHYI, IVANOV*. Although gender can be assigned incorrectly to a particular person, we believe that the results are statistically correct.

Thus, gender labels were inferred from surnames for an additional 1,260 female and 828 male authors. Altogether, we continue with 63.7% of genderized authors' names. And having in mind that the number of male authors is slightly underestimated, one can state that 1.5 times more female authors are found.

Our finding is in line with the statement in (Larivière et al., 2013), where Ukraine is mentioned among other “countries with lower scientific output” that are characterized by more prevalent female authorship. But our research is not cross-disciplinary, it is initially related to the Economics area. Moreover, while Web of Science data was used in (Larivière et al., 2013), we exploit Crossref as a data source in our work. The remaining question is how different can be results obtained for top journals indexed in authoritative databases such as Scopus or Web of Science and for data beyond these sources.

It is shown that gender disparities are disciplinary-dependent (Nicola & D'Agostino, 2021). The Economics discipline is considered rather as a male-dominated one^① (Bayer & Rouse, 2016; Liu et al., 2020). For example, 20.3% female versus 63.4% male authors were found for an economics-related data set analyzed (Maddi & Gingras, 2021)^②. A similar proportion was reported by (Liu et al., 2020): “The proportion of men is 2.45 times higher than that of women”. On the contrary, more female authors are found in our data^③. Thus, it is natural to expect more papers from female authors (at least one author is recognized as female) – 76.3% than from

^① Gender in the global research landscape.

https://www.elsevier.com/_data/assets/pdf_file/0003/1083945/Elsevier-gender-report-2017.pdf

^② The gender is undetectable for the rest authors.

^③ Of course, one has to remember that different datasets are used in these different case studies.



male authors – 48.4%. But let’s look more deeply into the individual contributions of female and male authors, as it was suggested in (Huang et al., 2020). The conclusion “female and male authors are largely indistinguishable when it comes to the number of publications per year” supported also by results presented by Liu et al., (2020), which is relevant to our data: authors of both genders publish approximately the same number of papers per year on average. To be more precise, 1.28 papers per year on average are published by male authors, and 1.34 by female authors.

Another interesting relevant issue is the analysis of gender mixing and the patterns for forming authorship teams. The annual change of shares of papers classified according to the gender of authors for our data is shown in Fig. 1. Five categories are defined here for 17,352 out of 25,933 papers (the rest cannot be classified due to the lack of gender information about authors): papers authored by a single person are labeled as *F solo* or *M solo* in correspondence to the author’s gender (F stands for female and M for male); *F coll* and *M coll* labels are used only for papers where gender for all authors is defined and the same (solo-gender collaboration); if not all authors are labeled by gender, but at least one female and one male are found, the paper is attributed to the *MIX coll* category (cross-gender collaboration). Since not all authors in the data set are labeled by gender, the gender spectrum of papers can be considered as not conclusive. However, the same rule is applied here to data that correspond to different years, therefore, the tendencies are considered informative.

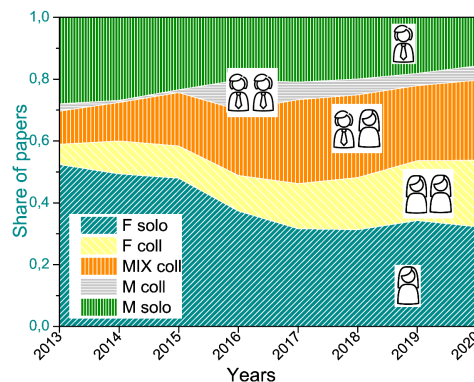


Figure 1. (Color online) Annual dynamics of shares of papers classified according to the gender of authors: a single female author (*F solo*); two or more authors, only females (*F coll*); two or more authors, at least one female and at least one male (*MIX coll*); two or more authors, only males (*M coll*); a single male author (*M solo*). The time window between 2013 and 2020 is chosen for visualization due to small annual statistics (less than 100 papers per year) before 2013.

As Fig. 1 shows, many papers correspond to the decreasing but still largest category of publications by a single female author. On the one hand, this is in line with the conclusion in (Boschini & Sjögren, 2007), where over-representation of



single female authors is reported. On the other hand, it was already mentioned that the number of papers by females is expected to be larger in principle, simply due to the larger number of female authors. Indeed, for the same data set, if authors' gender labels "F", "M" and "undefined" are randomly reshuffled first, the share of papers in the *F solo* category $\approx 34.8\%$ is very close (even slightly larger) to the real value ($\approx 33.8\%$), see Table 1. Again, gender equality can be found at the individual level. The similar average share of solo-authored papers per author's portfolio is found for both genders (approximately 24% if all genderized authors are taken into account; 28% (F) and 30% (M) if only genderized authors with at least 2 papers are considered). The similar conclusion is found (Kwiek & Roszka, 2021) for Polish authors in the field of Economics – the share of solo-publications in an individual portfolio is approximately equal for both male and female authors. According to this, a curious paradox can be observed: While the majority of papers in the Economics discipline are not collaborative (Mryglod et al., 2021), at the individual level, only a third (fourth if authors with a single paper are considered) paper is written without co-authors. No gender differences are also found in terms of the average number of authors per individual collaborative paper.

The homophily of co-authorship groups with respect to authors' gender in the Economics field was discussed in (Boschini & Sjögren, 2007). At first glance, our results deny this conclusion. The cross-gender category is the second-largest one among the gender-labeled papers in our data, see Fig. 1 and Table 1. However, it is easy to show that this share is smaller than expected. And indeed, solo-gender collaborative papers by female authors are over-represented in the real data set in comparison to reshuffled data. It is interesting to see that the shares of cross-gender

Table 1. Distribution of papers according to authors' gender: a single female author (*F solo*); two or more authors, only females (*F coll*); two or more authors, at least one female and at least one male (*MIX coll*); two or more authors, only males (*M coll*); a single male author (*M solo*).

	Entire data set		Reshuffled*	Not indexed in Scopus and/or WoS		Indexed in Scopus and/or WoS	
	# records	Share	Share	# records	Share	# records	Share
All papers	25933	-	-	22683	-	3250	-
All papers assigned to one of the categories	17352	100%	(13785 records on average)	15422	100%	1930	100%
F solo	5871	33.8%	34.8%	5684	36.9%	187	9.7%
F coll	3085	17.8%	10.4%	2841	18.4%	244	12.6%
MIX coll	4283	24.7%	28.5%	3236	21%	1047	54.3%
M coll	816	4.7%	4.1%	598	3.9%	218	11.3%
M solo	3297	19%	22.2%	3063	19.8%	234	12.1%

* Averages for 10 versions of the original set of publications labeled by categories after random reshuffling of authors' gender labels are provided.



papers and collaborative papers by male authors are remarkably higher for papers that are indexed in Scopus and/or WoS, see Table 1. To some extent, this is in agreement with the conclusion about the tendency to comparatively lower gender homophily in higher-impact journals (Holman & Morandin, 2019). Such gender mixing can be seen in a very positive way, since gender is one of the most important dimensions of team diversity, which, in its turn, is often considered as a powerful catalyst for creativity, see, e.g., (Farhoomand & Drury, 2001; Liao, 2010; Reynolds & Lewis, 2017) and references therein.

5 Authors ordering

Another aspect of forming collaboration teams – the ordering of authors – can be studied using our data. The position of the author's name in a list, which is ordered neither randomly nor alphabetically, can be considered as a basis for credit allocation. In this case, it is also reasonable to investigate the correlation between authors' gender and their roles (i.e., positions in co-authorship lists). For example, the statistics of female first-authored journal articles are studied in (Thelwall & Mas-Bleda, 2020).

Economics is often considered as one of the fields where alphabetization is common (see, e.g., Frandsen & Nicolaisen, 2010; Kuld & O'Hagan, 2017; Levitt & Thelwall, 2013; Waltman, 2012), although the alphabetization rate in the economy has declined somewhat over the past decade (Wohlrabe & Bornmann, 2022). Moreover, top Economic journals are characterized by the share of alphabetized articles that is even higher compared to other Economic journals (70% vs. 60%, correspondingly) (Kuld & O'Hagan, 2017; Levitt & Thelwall, 2013). Such a way of ordering can be interpreted as a declaration of equal authors' contribution. At the same time, the first position can still be perceived as special by external assessors. The so-called “alphabetical discrimination” is discussed in (Einav & Yariv, 2006; Kuld & O'Hagan, 2017): The staff members of a top U.S. economic department whose surnames start with letters from the first part of the alphabet are found to be more tenured. In some sense, this can be seen as a consequence of the Thomas theorem (Bornmann & Marx, 2020): even a groundless consideration of the first author as the principal one can cause further advantage in an academic career. Therefore, the following two questions are addressed further: (i) Is there any gender preference for the first position in a co-authorship list? and (ii) Can we state that the alphabetical ordering of authors is typical for the Ukrainian Economics discipline, in general?

Gender of first authors. Due to the larger share of female authors, it is expected also to get a larger share of collaborative papers, where the first positions in co-authorship lists are occupied by female authors. But it is interesting to calculate



the probability of being in the first position in a collaborative list for each author. The individual publication records for both male and female authors with at least three collaborative papers are analyzed for this purpose: the share of collaborative papers, where the given author is in the first position, is counted. According to our results, the probability to occupy the first position is equal to 0.48 for authors of both genders}. In this context, rather gender equality is found.

Alphabetization. The alphabetical or non-alphabetical order of authors is determined by their last names. For the authors with identical last names, initials are considered. Since the majority of author names hypothetically are Slavs, their alphabetization can be performed in two ways: for their original names written in Cyrillic or for the corresponding versions in Latin. Therefore, the possible alphabetical order of each sequence of names is checked twice: for the names as they are in metadata and for their transliterations. To give an example, the names in the list *MARTYNENKO VALENTYNA; ZAMOTA IRINA* are found to be *ordered* by Latin alphabet (*MARTYNENKO; ZAMOTA*), but their transliterated versions are *not ordered* by Cyrillic alphabet: (*МАРТИНЕНКО; ЗАМОТА*).

The more authors in the collaboration list, the less the probability of accidental alphabetical authorship is (see, e.g., Kuld & O'Hagan, 2017; Waltman, 2012). Since small co-authorship lists are dominant in the Economics discipline, in many cases one cannot be sure whether the names are sorted alphabetically intentionally or unintentionally. But non-alphabetical sorting is an unambiguous indicator of other priority scheme usage. Therefore, we count the share of papers with authors' names ordered neither by Latin nor Cyrillic alphabets. Corresponding numbers for different publication samples are provided in Table 2. For example, it can be seen that in almost half of all collaborative articles, authors are not sorted by the last names. Moreover, knowing the exact numbers of authors in the rest of the papers, one can suggest that another 21.3% of the papers are sorted in alphabetic order accidentally^⑩ (Kuld & O'Hagan, 2017; Waltman, 2012). Therefore, we conclude that the level of alphabetization of authors' names in Ukrainian Economic papers turned out to be lower than it was reported for other publication sets. Moreover, the share of non-alphabetized articles indexed in Scopus or Web of Science databases is even larger. One can only speculate about the reasons for such features of Ukrainian Economics research. While the first position in the list of authors is not encouraged officially, still it is considered more beneficial due to its greater visibility (it is a common



^⑩ Considering all collaborative papers in the initial data set, one can find that the total number of alphabetically sorted papers is 6,641. The probability for a paper to be unintentionally ordered alphabetically depends on the number of authors n : $P(n)=1/n!$. Correspondingly, the number of papers that are intentionally alphabetized might be smaller: $5,194(1-1/2!)=2,597$ duo-authored papers; $1,284(1-1/3!)=1,070$ trio-authored papers; $135(1-1/4!)\approx 129$ quarto-authored papers;

practice to mention just the first author to refer to the co-authored publication) and its special perception within a number of disciplines. The name of the first author appears at the beginning of the reference.

Table 2. The numbers of non-alphabetically ordered papers and estimated volume of potentially alphabetized papers (adjusted values) for different data samples.

	Number of papers (collaborative only)	Number (%) of papers with definitely non-alphabetical ordering of authors' names	Adjusted number (%) of papers that are potentially alphabetically ordered by intention
Entire data sets	13237	6596 (49.8%)	3824 (28.9%)
Papers marked by gender categories	8184	4057 (49.6%)	2384 (29.1%)
Papers indexed in Scopus and/or WoS	2592	1672 (64.5%)	601 (23.2%)
Papers NOT indexed in Scopus and/or WoS	10645	4924 (46.3%)	3223 (30.3%)
Cross-gender papers	4283	2369 (55.3%)	1169 (27.3%)
Solo-gender papers	3901	1688 (43.3%)	1215 (31.1%)

6 Discussion and conclusions

The analysis of Ukrainian journals within the Economics discipline, started in the previous work (Mryglod et al., 2021), is continued in this study. Revealing the typical features of this particular segment of scholarly literature is important for solving many practical issues related to the development of assessment procedures at the national level. However, another goal of this work is to reinforce the call for complete and qualitative metadata. Crossref database is used here to describe one of SSH disciplines for poorly studied European countries. Publication metadata related to Ukrainian Economic journals is collected from the Crossref database. An attempt was made to conduct an analysis with an emphasis on gender effects at the level of individual authors. However, the procedure for disambiguating authors' names can be done only partially. A number of peculiarities of processing author names related to the usage of Cyrillic and local traditions of parallel usage of different forms of names and even surnames are highlighted. Since the gender of an author is inferred from the full first name, even partial merging of authors' records allows one to increase the statistics of publications with authors labeled by gender. Moreover, a manually created list of gender-specific endings for Slavic last names was used to enlarge the number of genderized authors. Altogether, 63.7% of 23,094 author records were labeled by gender, and the number of female authors is found to be 1.5 times larger than male authors. This result contradicts the statements about the masculinism nature of Economics research. Alternatively, female dominance in Ukrainian Economics research may be considered as a hint about its specific thematic spectrum. According to (Thelwall et al., 2019), keywords related to qualitative and exploratory methods are statistically associated with female scholar authors, while



other keywords related to quantitative methods are more related to male authors. In some sense, such sensitivity of gender representation to the topic selection is in line with the conclusion in (West et al., 2013): considerable differences in this context were observed for different Economics subfields. This reinforces our previous conclusions about the specific patterns of collaborativeness in Ukrainian Economics research. Still, an important caution exists: this study is one of the rare examples where data beyond internationally recognized databases is used. Therefore, the guess about the different nature of locally-oriented and internationally-oriented topics chosen for Economics research remains relevant.

Gender mixing is analyzed to find the evidence that gender plays a role when forming collaboration teams. All papers labeled according to five gender-related categories (solo-publications by males; solo-publications by females; solo-gender collaborations of males; solo-gender collaborations of females; and cross-gender collaboration) are considered. While only one third¹⁵ of individual papers are written without coauthors and there is a tendency towards more collaborative papers (Mryglod et al., 2021), the share of solo-publications remains high. One third of all papers are found to be solo-publications authored by female authors. The corresponding share for male authors is slightly smaller than expected. Finally, while the share of cross-gender teams is larger than the shares of solo-gender teams (see Fig. 1), the results compared with randomly reshuffled data indicate that the share of cross-gender teams is considerably higher than it can be expected only for publications in the journal indexed in Scopus and/or WoS, see Table 1.

It is shown that the level of alphabetization of authors' names in Ukrainian Economic papers is comparatively low. This is especially true for articles indexed in Scopus or Web of Science databases.

Interestingly, different results of gender mixing are found for papers published in journals indexed in Scopus or Web of Science, compared to the rest of publications. Remarkably, while the largest share of papers solo-authored by female authors is expected due to the greater general number of female authors, this category of papers indexed in the international databases is the least represented one. Most papers in internationally recognized journals are characterized by cross-gender collaboration. This can be seen as the manifestation of the so-called reactivity of the Ukrainian Economics discipline (see Aistleitner et al., 2019; Sasvári et al., 2019). One can speculate about the adaptive publishing behavior: a different publishing or even research strategy is chosen depending on the level of recognition and audience of the target journal. The similar conclusion can be drawn for different shares of papers where authors are listed alphabetically. Interestingly, while a high level of



¹⁵ One fourth if authors with single papers are considered as well

alphabetization is found for the Economics discipline in general and even higher for Economic publications in internationally recognized journals, the opposite pattern is observed in Ukrainian Economics research.

To conclude, the results of another case study are presented. Besides the findings specifically related to Ukrainian research, some key aspects related to the processing of non-English metadata are highlighted. It is worth emphasizing once more that many complications become irrelevant if unique digital identifiers are commonly used.

Funding information

This work was supported in part (OM) by the National Research Foundation of Ukraine, project No. 2020.01/0338.

Author contributions

Olesya Mryglod (olesya@icmp.lviv.ua): Conceptualization (Equal), Data curation (Lead), Methodology (Equal), Resources (Lead), Software (Lead), Visualization (Lead), Writing – original draft (Lead), Writing – review & editing (Equal); Serhii Nazarovets (serhii.nazarovets@gmail.com): Conceptualization (Equal), Methodology (Equal), Resources (Supporting), Visualization (Equal), Writing – original draft (Supporting), Writing – review & editing (Equal); Serhiy Kozmenko (kozmenko.uabs@gmail.com): Conceptualization (Supporting), Methodology (Supporting), Visualization (Equal), Writing – original draft (Supporting), Writing – review & editing (Equal).

Acknowledgements

The authors thank anonymous peer reviewers for improving the paper.

References

- Aistleitner, M., Kapeller, J., & Steinerberger, S. (2019, December). Citation patterns in economics and beyond. *Science in Context*, 32(4), 361380. doi: 10.1017/s0269889720000022
- Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of scopus and web of science. *Journal of Data and Information Science*, 4(1), 1–21.
- Bayer, A., & Rouse, C. E. (2016, November). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, 30(4), 221–242. doi: 10.1257/jep.30.4.221
- Bornmann, L., & Marx, W. (2020, February). Thomas theorem in research evaluation. *Scientometrics*, 123(1), 553–555. doi: 10.1007/s11192-020-03389-6



- Boschini, A., & Sjögren, A. (2007, April). Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics*, 25(2), 325–365. doi: 10.1086/510764
- Cainelli, G., Maggioni, M. A., Uberti, T. E., & De Felice, A. (2012). Co-authorship and productivity among Italian economists. *Applied Economics Letters*, 19(16), 1609–1613.
- Cheberkus, D., & Nazarovets, S. (2019, November). Ukrainian open index maps local citations. *Nature*, 575(7784), 596596. doi: 10.1038/d41586-019-03662-6
- Einav, L., & Yariv, L. (2006, February). What's in a surname? the effects of surname initials on academic success. *Journal of Economic Perspectives*, 20(1), 175–188. doi: 10.1257/089533006776526085
- Farhoomand, A., & Drury, D. H. (2001). Diversity and scientific progress in the information systems discipline. *Communications of the Association for Information Systems*, 5. doi: 10.17705/1cais.00512
- Frandsen, T. F., & Nicolaisen, J. (2010, October). What is in a name? Credit assignment practices in different disciplines. *Journal of Informetrics*, 4(4), 608–617. doi: 10.1016/j.joi.2010.06.010
- Gomide, J., Kling, H., & Figueiredo, D. (2017). Name usage pattern in the synonym ambiguity problem in bibliographic data. *Scientometrics*, 112(2), 747–766.
- Holman, L., & Morandin, C. (2019, April). Researchers collaborate with same-gendered colleagues more often than expected across the life sciences. *PLOS ONE*, 14(4), e0216128. doi: 10.1371/journal.pone.0216128
- Huang, J., Gates, A. J., Sinatra, R., & Barabasi, A.-L. (2020, February). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), 4609–4616. doi: 10.1073/pnas.1914221117
- Kim, J., Kim, J., & Owen-Smith, J. (2021). Ethnicity-based name partitioning for author name disambiguation using supervised machine learning. *Journal of the Association for Information Science and Technology*, 72(8), 979–994.
- Kuld, L., & O'Hagan, J. (2017, November). Rise of multi-authored papers in economics: Demise of the 'lone star' and why? *Scientometrics*, 114(3), 1207–1225. doi: 10.1007/s11192-017-2588-3
- Kwiek, M., & Roszka, W. (2021, August). Gender-based homophily in research: A large-scale study of man-woman collaboration. *Journal of Informetrics*, 15(3), 101171. doi: 10.1016/j.joi.2021.101171
- Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013, December). Bibliometrics: Global gender disparities in science. *Nature*, 504(7479), 211–213. doi: 10.1038/504211a
- Levitt, J. M., & Thelwall, M. (2013, July). Alphabetization and the skewing of first authorship towards last names early in the alphabet. *Journal of Informetrics*, 7(3), 575–582. doi: 10.1016/j.joi.2013.03.002
- Liao, C. H. (2010, November). How to improve research quality? examining the impacts of collaboration intensity and member diversity in collaboration networks. *Scientometrics*, 86(3), 747–761. doi: 10.1007/s11192-010-0309-2
- Liu, J. W., Song, Y. L., & Yang, S. (2020, July). Gender disparities in the field of economics. *Scientometrics*, 125(2), 1477–1498. doi: 10.1007/s11192-020-03627-x



- Maddi, A., & Gingras, Y. (2021, February). Gender diversity in research teams and citation impact in economics and management. *Journal of Economic Surveys*, 35(5), 1381–1404. doi: 10.1111/joes.12420
- Mryglod, O. (2012). Ukrainian scientific academic periodicals: The level of visibility. *Science of Ukraine in the Global Information Space* (6), 36–44.
- Mryglod, O., Nazarovets, S., & Kozmenko, S. (2021, June). Universal and specific features of ukrainian economic research: publication analysis based on crossref data. *Scientometrics*, 126(9), 81878203. doi: 10.1007/s11192-021-04079-7
- Müller, M.-C., Reitz, F., & Roy, N. (2017). Data sets for author name disambiguation: an empirical analysis and a new resource. *Scientometrics*, 111(3), 1467–1500.
- Nicola, A. D., & D'Agostino, G. (2021, March). Assessment of gender divide in scientific communities. *Scientometrics*, 126(5), 3807–3840. doi: 10.1007/s11192-021-03885-3
- Reynolds, A., & Lewis, D. (2017). Teams solve problems faster when they're more cognitively diverse. <https://hbr.org/2017/03/teams-solve-problems-faster-when-theyre-more-cognitively-diverse>. ([On-line; accessed 05-July-2022])
- Sasvári, P., Nemeslaki, A., & Duma, L. (2019, March). Exploring the influence of scientific journal ranking on publication performance in the hungarian social sciences: the case of law and economics. *Scientometrics*, 119(2), 595–616. doi: 10.1007/s11192-019-03081-4
- Schläpfer, F. (2010, June). How much does journal reputation tell us about the academic interest and relevance of economic research? empirical analysis and implications for environmental economic research. *Gaia – Ecological Perspectives for Science and Society*, 19(2), 140–145. doi: 10.14512/gaia.19.2.13
- Thelwall, M., Bailey, C., Tobin, C., & Bradshaw, N.-A. (2019, February). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics*, 13(1), 149–169. doi: 10.1016/j.joi.2018.12.002
- Thelwall, M., & Mas-Bleda, A. (2020, August). A gender equality paradox in academic publishing: Countries with a higher proportion of female first-authored journal articles have larger first-author gender disparities between fields. *Quantitative Science Studies*, 1(3), 1260–1282. doi: 10.1162/qss_a_00050
- Treeratpituk, P., & Giles, C. L. (2012). Name-ethnicity classification and ethnicity-sensitive name matching. In Proceedings of the AAAI conference on artificial intelligence 26: 1141–1147.
- Truc, A., Claveau, F., & Santerre, O. (2021, January). Economic methodology: a bibliometric perspective. *Journal of Economic Methodology*, 28(1), 67–78. doi: 10.1080/1350178x.2020.1868774
- Vaio, G. D., & Weisdorf, J. L. (2009, March). Ranking economic history journals: a citation-based impact-adjusted analysis. *Cliometrica*, 4(1), 1–17. doi: 10.1007/s11698-009-0039-y
- Waltman, L. (2012, October). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, 6(4), 700–711. doi: 10.1016/j.joi.2012.07.008
- Wei, G. Y. (2018, March). A bibliometric analysis of the top five economics journals during 2012–2016. *Journal of Economic Surveys*, 33(1), 25–59. doi: 10.1111/joes.12260



Research Paper

- West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013, July). The role of gender in scholarly authorship. *PLoS ONE*, 8(7), e66212. doi: 10.1371/journal.pone.0066212
- Wohlrabe, K., & Bornmann, L. (2022, March). Alphabetized co-authorship in economics reconsidered. *Scientometrics*, 127(5), 2173–2193. doi: 10.1007/s11192-022-04322-9
- Zhao, S. X., Yu, S., Tan, A. M., Xu, X., & Yu, H. Y. (2016, May). Global pattern of science funding in economics. *Scientometrics*, 109(1), 463–479. doi: 10.1007/s11192-016-1961-y



This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

