A Quantitative Review of Matching Papers in Economics: Evolution, Diversity, and Gender*

Climent Quintana-Domeque   Alejandro Robinson-Cortés   Ye Yuan

April 7, 2025

Abstract

Matching markets offer a unique lens through which to study the evolution of the economics profession—both in terms of its research output and the composition of its contributors. We conduct a quantitative review of the field's development, using text analysis on RePEc data to identify matching-related papers and track changes in scale, diversity, and authorship. Since 1975, the volume of matching research has grown faster than the discipline overall, with its share of economics papers increasing tenfold. Matching papers have become more widely published across journals, including a notable rise in Top-5 outlets. Text-based clustering reveals a diverse intellectual structure, with distinct strands such as macro-labor, family economics, and market design emerging organically. Patterns of collaboration have also changed. Matching papers showed early and widespread co-authorship, but since 2010, the profession has surpassed the field in collaboration intensity. Gender representation has improved, but matching lags behind: women make up 30% of authors in economics by 2020, but less in matching, with a consistent 10-year delay. Cross-gender collaborations have become more common in economics overall, though less rapidly in matching.

Keywords: Matching markets, History of economics, Co-authorship patterns, Gender representation, Text analysis

*Climent Quintana-Domeque, Department of Economics, University of Exeter: c.quintana-domeque@exeter.ac.uk; Alejandro Robinson-Cortés, Department of Economics, University of Exeter: a.robinson-cortes@exeter.ac.uk; Ye Yuan, Department of Economics, University of Exeter: yy429@exeter.ac.uk. This review article has been prepared for the *Oxford Research Encyclopedia of Economics and Finance*.

1 Introduction

This review paper explores the evolution, diversity, and gender representation in the production of scientific articles within the economics profession, with a specific focus on research related to matching markets. The economics of matching markets is a natural field to study. On one hand, matching is inherently about how different economic units come together—in our context, how researchers collaborate with each other. On the other hand, research on matching markets cuts across various fields within economics, including market design, family economics, and macro-labor economics. Our aim is to quantify and visualize the breadth and diversity of a vast body of research—and the researchers behind it—that explores different facets of a common question: how do economic agents come together and match with each other to interact and exchange in markets?

The foundations of the matching literature have a long history. They may be traced back at least to the work by Leonid Kantorovich and Tjalling Koopmans on assignment problems in the mid-20th century (Kantorovich, 1939, 1942; Koopmans, 1947; Kantorovich, 1948; Koopmans and Beckmann, 1957; Kantorovich and Rubinstein, 1958).[1] In 1962, Lloyd Shapley and David Gale pioneered the study of matching markets with non-transferable utility in their landmark paper "College admissions and the stability of marriage." In 1971, Lloyd Shapley and Martin Shubik elaborated the theory with transferable utility. Since then, the economic analysis of matching markets has evolved into several subfields, giving rise to multiple strands of literature. The work of Koopmans and Kantorovich was jointly recognized with the Nobel Prize in 1975 "for their contributions to the theory of optimum allocation of resources". At least six other Nobel laureates, including Lloyd Shapley, have made significant contributions to the matching literature, spanning different applications and environments.

[1]The origins of this line of research can be traced even further back to Gaspard Monge's work on optimal transport (Monge, 1781).

In 1992, Gary Becker was awarded the Nobel Prize, in part, for contributions to the study of family economics and marriage. Becker (1973) built on the work of Koopmans and Beckmann (1957) to analyze household formation, sorting, and assortative matching in the marriage market. In 2010, Peter Diamond, Dale Mortensen, and Christopher Pissarides were awarded the Nobel Prize for their work on search and matching frictions. In a series of papers, they advanced the analysis of price dispersion, efficiency in search environments, and the development of modern search and matching theory in labor economics (Diamond, 1982; Pissarides, 1985; Mortensen and Pissarides, 1994). Departing from the earlier foundations laid out by Koopmans, Gale, Shapley, and co-authors, their work introduced dynamics and search frictions into the analysis of matching markets—placing particular emphasis on the labor market. In 2012, Lloyd Shapley and Alvin Roth were awarded the Nobel Prize "for the theory of stable allocations and the practice of market design." Roth brought the theoretical insights of Gale and Shapley (1962) into practical application, most notably through his analysis of market inefficiencies in the National Resident Matching Program (NRMP) (Roth, 1984, 2003).[2]

Despite the rich theoretical and empirical body of research following these foundational contributions, little is known about the meta-evolution of the field—how it has grown, diversified, and evolved demographically within the economics profession. This paper fills that gap by providing a panoramic view based on a comprehensive dataset constructed from RePEc (Research Papers in Economics). Using text-based classification algorithms applied to bibliographic data, we analyze the evolution and intersections of different strands of the matching literature and examine collaboration patterns and gender dynamics within the research community.

[2]David Gale passed away in 2008. Had it been awarded during his lifetime, he is believed to have shared the prize with Roth and Shapley (Roth, 2012).

Our quantitative review covers a vast body of work spanning multiple fields and subfields within the profession over several decades. Numerous review articles, handbooks, and monographs analyze and synthesize distinct strands of the matching literature. We provide a brief introductory overview in Section 2. Rather than delving into the nuanced connections within this extensive body of research, we adopt a bird's-eye view of the literature. The main analysis of the paper is then based on text analysis of bibliographic information—including titles, abstracts, keywords, journals, and publication years—from hundreds of thousands of papers. The first order of business in our analysis is to design a systematic procedure to identify matching papers in RePEc's database. We develop an ad-hoc algorithm based on a careful selection of matching-related terms and compare its performance with off-the-shelf machine learning classification methods. Section 3 explains our approach, as well as the steps we followed to evaluate its performance. With our sample of matching papers in hand, in Section 4 we study the evolution, diversity, and gender representation of the field.

Our analysis is structured around four main areas. First, we document the growth and evolution of the field between 1975 and 2020. While the volume of research published in the economics profession increased rapidly during this period, the field of matching grew even more rapidly. In that period, the number of RePEc papers published per year increased from around 6,500 to over 135,000 while the share of papers about matching increased tenfold—from around 2 per 10,000 papers in 1975 to 20 per 10,000 in 2020.

Second, we explore publication patterns and journal diversity. The number of journals publishing economics research grew from approximately 200 in 1975 to nearly 2,000 in 2020. The proportion of journals publishing matching-related papers increased from fewer than 1% to 8%. Among Top-5 journals, matching papers became more frequent, its share rising from 1% in 1995 to 5% in 2020—particularly after 2005. In non-Top-5 journals, the share of matching papers grew rapidly in the 1990s and plateaued in the 2000s at around 0.175%.

Third, we investigate the intellectual diversity and internal structure of the matching literature using hierarchical clustering based on the textual similarity of papers' titles and abstracts. This analysis reveals a rich tapestry of subfields, ranging from matching theory and macro-labor to applications in marriage, school choice, housing, kidney exchange, etc. The analysis shows how the distinct research agendas within the field, in part reflected in the distinct Nobel Prizes, emerge organically from text-analysis of the language used in titles and abstracts.

Finally, we turn to collaborations and gender representation. Collaboration dynamics have undergone major shifts during the past decades in the economics profession, the matching community being no exception. Co-authorship has become the norm in economics, with matching papers initially exhibiting earlier and higher levels of co-authorship than the rest of the discipline. While over half of matching papers were co-authored as early as 1990, the profession in general only reached that threshold after 2005. However, the rate of collaboration in the discipline has outpaced that within the matching community since 2010. In 2020, the average number of authors per paper was around 30% higher in the profession as a whole than in the matching community.

Regarding gender, the share of female authors has increased steadily in the profession, from 10% in 1975 to 30% in 2020. Among matching papers, the share of women authors has also increased but consistently lagged the profession for about 10 years. Looking at the intersection of gender and collaboration, our analysis shows a rising trend towards cross-gender collaboration in economics, albeit with a delayed uptake in the matching literature. Taken together, our analysis reflects deep changes in the culture and production technology behind research in the economics profession, the matching community included.

2 A brief literature review

In this section, we offer a brief literature review intended to guide students and researchers new to the field of matching.[3] It is recommended for readers interested in learning more about the various strands of the matching literature. Those already familiar with the literature may choose to skip this section. The quantitative component of our review begins in Section 3, where we describe the data and methods used in our analysis.

Matching models capture person-specific goods and relationships. Matching is about choosing and being chosen. At their core, all matching models aim to answer the question: *who matches with whom?* The baseline model for most matching markets is the standard marriage market model, with two sides and one-to-one matching.[4]

Matching models vary along several dimensions but are typically classified based on two main criteria: whether transfers are allowed and whether the matching process involves frictions. A central topic in the literature is assortativeness, or sorting—the extent to which similar individuals are more likely to match with one another. Measuring assortativeness and linking it to the underlying fundamentals of a market has been a major focus of empirical research. This brief review is organized around these core themes.

2.1 Classification of matching models

2.1.1 With or without transfers?

In non-transferable utility (NTU) settings, payoffs are fixed or exogenously determined. Who matches whom is determined by a mechanism that does not involve transfers, such as a matching algorithm. In contrast, in transferable utility (TU) settings, payoffs are determined

[3] For a recent handbook of the economics of matching, see Che, Chiappori, and Salanié (2024). For an undergraduate-level textbook, see Haeringer (2017).

[4] Other variants include one-side matching (e.g., matching roommates), many-to-one (e.g., matching students to universities), and many-to-many (e.g., matching PhD students to PhD supervisors).

endogenously in equilibrium, and they may affect who matches whom. A more general case is the imperfectly transferable utility (ITU) setting, which nests the NTU and the TU setting.

Historically, NTU matching traces back to Gale and Shapley (1962) and their celebrated Deferred Acceptance algorithm.[5] NTU matching markets may be studied under a decentralized lens (e.g., Roth and Vate, 1990) or a centralized one, with the latter being more commonly associated with market design. Real-world applications to centralized clearinghouses include the redesign of the NRMP (Roth and Peranson, 1999), and the development of mechanisms for school choice (Abdulkadiroğlu and Sönmez, 2003) and kidney exchange (Roth, Sönmez, and Ünver, 2004).[6] Another key centralized NTU model is the housing market of Shapley and Scarf (1974), which introduced a core-based approach to object allocation without monetary transfers and inspired subsequent work on priority-based mechanisms.

The seminal papers in TU matching in economics are Koopmans and Beckmann (1957), Shapley and Shubik (1971), and Becker (1973), which study models where transfers can be used to support efficient and stable matchings.[7] ITU models, where utility is neither fully transferable nor fixed, were pioneered by Crawford and Knoer (1981) and Kelso and Crawford (1982), and revived by Hatfield and Milgrom (2005), with further developments by Legros and Newman (2007). These models generalize both TU and NTU frameworks, capturing complementarities and substitutabilities in preferences, and have become central to the analysis of matching with contracts and complex market environments.

[5] Classic monographs include Knuth (1976); Gusfield and Irving (1989); Roth and Sotomayor (1990).

[6] For review articles and general-interest accounts, see Roth (2002); Niederle, Roth, and Sönmez (2008); Roth (2008, 2015, 2016). For handbooks of matching with applications to market design, see Vulkan, Roth, and Neeman (2013); Echenique, Immorlica, and Vazirani (2023).

[7] See Sattinger (1993) for an early survey of matching and assignment models in labor economics.

2.1.2 With or without frictions?

The second axis of distinction is whether matching occurs in a frictionless environment, where any participant is assumed to have perfect information about all possible matches, even in large markets, or in the presence of search frictions.[8] Matching models without frictions, where matches occur instantaneously and costlessly, have been applied to a host of different contexts. Chade, Eeckhout, and Smith (2017) highlight applications to marriage markets, hierarchies, international trade, finance, CEO selection, foreign direct investment, and development (see references therein).

Frictions have been incorporated into both NTU models (e.g., Bergstrom and Bagnoli, 1993; Burdett and Coles, 1997; Smith, 2006) and TU models (e.g., Shimer and Smith, 2000; Shimer, 2005; Atakan, 2006; Eeckhout and Kircher, 2010; Hagedorn, Law, and Manovskii, 2017). Matching models with frictions have been particularly relevant to the study of labor markets. Seminal papers in the search literature are Mortensen (1982, 1988). As emphasized by Chiappori (2017), while Mortensen explicitly refers to the marriage market, most of their use has been in studying the labor market, where frictions are in part responsible for unemployment (Pissarides, 2000). Search models play a central role in the labor economics literature (e.g., Postel-Vinay and Robin, 2002; Lise, Meghir, and Robin, 2013; Lise and Robin, 2017). For a review on search-theoretic models of the labor market, see Rogerson, Shimer, and Wright (2005), for search and matching models, see Chade et al. (2017), and for directed search models, see Wright, Kircher, Julien, and Guerrieri (2021).[9]

[8]For a review on frictional matching models, see Smith (2011) and Chade and Kircher (2023).

[9]See also the early survey on the matching function in labor economics by Petrongolo and Pissarides (2001), and the monographs by Mortensen (2005); Mortensen and Pissarides (2011). For applications of matching models to the marriage market, see Chiappori (2017), which covers a range of topics, from bidimensional matching models—dating back at least to Tinbergen (1956)—to search models and macroeconomic applications, such as Fernández, Guner, and Knowles (2005), who study the implications of marital sorting for household income inequality.

2.2 Assortativeness and the econometrics of matching models

A central theme in the matching literature is that of assortative matching or sorting. Positive (negative) assortative matching is the tendency of individuals with similar (different) characteristics to form partnerships with each other. Becker (1973)'s foundational insight is that, in one-dimensional environments, positive complementarities in joint surplus lead to positive assortative matching (PAM) under TU.[10] He also notes that under NTU, PAM arises when preferences on both sides are strictly increasing in types.[11] In principle, assortativeness may be inferred from observing data on matching patterns. Nonetheless, Echenique, Lee, Shum, and Yenmez (2013) show that it is not possible to distinguish between TU and NTU models using only information on observed matches. Likewise, recent work by Chiappori, Costa Dias, Meghir, and Zhang (2025) shows that the use of different measures of assortativeness can generate different conclusions.

The question of what we can identify about a matching model from observed data lies at the heart of empirical work in matching. A seminal contribution in this area is by Choo and Siow (2006), who provide a structural econometric model of marriage markets under transferable utility. Recent developments and refinements of TU matching econometrics are discussed in Chiappori and Salanié (2016) and Galichon and Salanié (2023). For the econometrics of ITU models, see Galichon, Kominers, and Weber (2019), and for NTU settings, see the review by Agarwal and Somaini (2023).[12]

[10] Lindenlaub (2017) generalizes the concept of assortativeness to multi-dimensional types.

[11] Interestingly, the NTU condition is distinct from, and not implied by, the TU condition. Cases in which the unique stable matching is PAM under NTU but exhibits negative assortative matching (NAM) under TU are provided by Smith (2011) and Lee and Yariv (2018).

[12] See also the early reviews by Fox (2009), on the structural econometrics of matching models, and by Graham (2011), on the econometrics of assignment problems. For a monograph exploring the connections between empirical matching models and optimal transport, see Galichon (2016). Also see Bonnet, Galichon, Hsieh, O'Hara, and Shum (2022) for the connection between two-sided matching and random-utility discrete-choice models.

3 Data and Methods

This section presents the data and methods we use for our quantitative review. First, we describe RePEc's dataset and discuss why we selected it as our main data source. Second, we explain how we identify papers that are about matching. Third, we discuss and validate our approach to identify matching papers. Fourth, we explain the text-based procedure we use to measure diversity in the matching literature across fields and disciplines. We finalize by reviewing how we extract the data for gender of authorship.

3.1 Description of RePEc data

RePEc is the largest bibliographic database and central index for disseminating economics related research. Publishers and institutions maintain their own RePEc archives, and support the services and technical infrastructure voluntarily. Bibliographic information in RePEc is stored in a standardized metadata format known as ReDIF (Research Documents Information Format), where each bibliographic item is described in a plain text template, containing fields such as title, author, abstract, keywords, etc. The whole database contains over 4.8 million bibliographic items.[13]

RePEc is a middle-point between universal indices that cover scientific research in general, such as Google Scholar, Web of Science, or Scopus, and more specialized series within economics, such as the SSRN, NBER, CEPR and IZA working paper series.[14] One advantage of RePEc for our purposes is that it also indexes economics research in disciplines adjacent to economics. This is particularly relevant to our study, as the literature on matching markets

[13]The database consists in a collection of hundreds of thousands of text files, comprising over 90GBs of data. For instructions on how to access the database, visit https://ideas.repec.org/getdata.html. For a detailed description about the history and functioning of RePEC, see Zimmermann (2013).

[14]As a comparison, while RePEc contains 4.8 million research items, SSRN contains 1.5 million (SSRN, 2025); the NBER working paper series, 32,000 (NBER, 2025); the CEPR working paper series, 19,000 (CEPR, 2025); the IZA discussion paper series, 16,000 (IZA, 2025); the Social Sciences Citation Index of Web of Science, 11.3 million (Clarivate, 2025), and Scopus, over 2.4 billion (Scopus, 2025).

spans disciplines. Our analysis is focused on published papers, which comprise 2.18 million items in RePEc's database. We further restrict attention to papers published between 1975 and 2020.[15]

3.2 Defining papers about matching

In this subsection we describe the procedure we follow to classify papers that are about matching. From a broad perspective, our starting point is what has become the standard definition of a matching market, which is one where the identities of the transacting parties are part of a good's description. In matching markets, agents must not only choose but also be chosen since prices alone do not coordinate outcomes (Roth, 2015). Operationalising this definition, however, is not without challenge. Many papers that study questions related to matching markets do not use the phrase "matching market" in their title or abstract. The literature has also developed their own terminology, e.g., "search and matching," "marriage market," "school choice," "housig markets," etc. Finally, the word "matching" on its own also has meanings that fall outside the scope of our study and are nevertheless prevalent in the economics literature, e.g., "propensity score matching," "price matching," etc.

To alleviate the above concerns, we take a holistic approach that aims to cast a wide net around many of the terms that are associated with papers about matching across different fields. Figure 1 describes the algorithm we follow. The first step is to filter for papers that contain the string "match" in the title, abstract, or keywords.[16] While this criterion may exclude some papers that are indeed about matching—including early seminal contributions

[15] We chose this sample period as we found few papers about matching markets prior to 1975 and to exclude the COVID pandemic period, during which publication trends were atypical (Bürgi and Wohlrabe, 2022). Data quality varies across RePEc's database. Table A1 in the Appendix provides an overview about the quality of the data.

[16] In a preliminary analysis we filtered for the string "matching" and realized that several terms used by papers about matching were being left out, e.g., "job match," "skill mismatch," etc. The string "match" may also appear as a stand-alone word or inside others, such as "matches", "matching", "matched" etc.

such as Gale and Shapley (1962), Shapley and Shubik (1971), and Roth (1984)—we find it crucial for eliminating papers that use matching-related terminology but are not explicitly focused on matching.

After filtering for the string "match," we identify a set of JEL codes that are associated with research on matching markets.[17] Papers containing at least one of these JEL codes in combination with the string "match" are included in our final sample. This dual-criteria approach addresses a key limitation: even the most directly relevant JEL code—C78 Bargaining Theory and Matching Theory—encompasses papers that may not be about matching. By requiring both the appropriate JEL classification and the presence of the string "match," we increase the precision of our selection. However, given the prevalence of missing JEL codes in the RePEc data—due to both poor data quality and journals not requiring them—JEL codes can only take us so far.[18]

The cornerstone of our approach is a selection of 93 terms, which we refer to as *matching-related terms*. The terms we chose are commonly associated with distinct strands of the literature about matching. Once we have filtered for the string "match," and looked up for matching-related JEL codes, we select papers that include at least one matching-related term in their title, abstract, or keywords.[19] Finally, in the last step of our algorithm, we determine a list of terms commonly used in unrelated contexts that contain the string "matching" (e.g., "string matching," "moment matching," etc.). Papers that contain a term in this list are not included in the sample.[20]

[17] Detailed information about the JEL codes we use is reported in Table A2 in the Appendix.

[18] See Table A1 in the Appendix. Many journals, including leading journals such as Econometrica, have not always included JEL codes in their publications.

[19] Table A3 in the Appendix reports the full list of terms we use. We included several variations and spellings of the same term, e.g., "labor market" and "labour market," "one to one" and "one-to-one," "Gale-Shapley" and "Gale and Shapley," "men-optimal" and "men optimal," etc. As an additional check, we considered an additional list of terms for the papers that contained "labor market." This list is reported in Table A4. Every paper that contains "labor market" in its title, abstract or keywords also contains at least one of these terms.

[20] The list of terms we use to remove papers in the last step is reported in Table A5 in the Appendix.

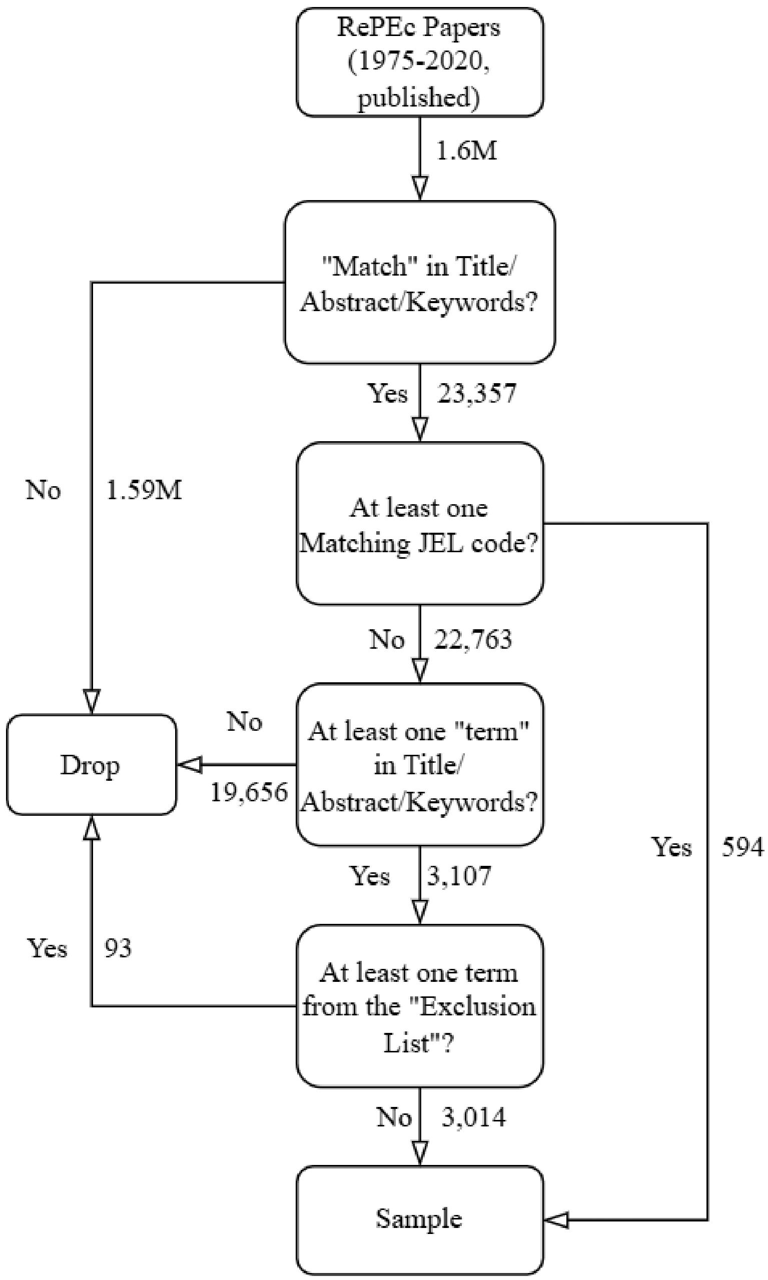

Figure 1: Classification algorithm to identify papers about matching

Figure 1 also reports the sample sizes at the different stages of the algorithm. The initial input sample of 1.6 million papers—the number of papers in RePEc published between 1975 and 2020—is reduced to a sample of 3,014 papers about matching. Our analysis and results are based on this sample. As we explain in the next section, we performed several calibration and validation steps to study the performance of our algorithm.

Table 1 reports the most frequent matching-related terms in the sample. Two terms appear in the title, abstract, or keywords of more than 10% of the papers: "labor market" and "matching model." On average, papers in our sample have 1.6 matching-related terms, and slightly more than half of them (53%) have a single term. The second column in the table reports the frequency of each term conditional on having only one. Finally, the last column reports, for each term, the percentage of papers that have it as a single term. Since papers are classified as being about matching if they contain *at least* one term in the list, the inclusion of some terms has more impact than others in the final selection of papers. For example, while "search and matching" is the third most common term, it uniquely identifies very few papers; 93% of the papers that have "search matching" have at least another matching-related term. By contrast, "perfect matching"—which is not a common term in the economics literature—is the 19th most frequent term in our sample.[21] However, 89% of the papers that contain it are only identified by it. A similar example is "real estate." As we shall see below in Subsection 4.2, these differences in terminology allow us to classify distinct strands of the matching literature.

[21] In graph-theoretic terms, a perfect matching refers to a pairing in a network in which every agent is matched to the agent who is matched to them. In economics, matchings that are not perfect matchings are often referred to as pre-matchings (e.g., Adachi, 2000).

Table 1: Top 20 Matching-Related Terms

#	*Keyword*	*Freq.*	*Cond. Unique*	*Is Unique*
1	labor market	20.3	21.3	33.8
2	matching model	10.5	9.4	28.6
3	search and matching	5.8	1.3	7.0
4	sorting	4.6	8.8	61.4
5	job match	4.5	5.3	37.5
6	marriage	3.8	4.8	41.2
7	job search	3.4	1.6	15.5
8	assortative	2.9	3.6	39.7
9	skill mismatch	2.8	3.0	34.5
10	matching market	2.5	1.2	15.6
11	two-sided market	2.4	1.6	21.7
12	search-model	2.3	2.3	31.9
13	matching friction	2.0	0.7	11.1
14	stable matching	1.9	0.9	14.6
15	real estate	1.8	4.1	74.7
16	housing market	1.7	2.8	52.3
17	search friction	1.6	1.0	20.0
18	perfect matching	1.5	4.1	89.0
19	matching efficiency	1.5	0.8	18.1
20	matching mechanism	1.4	1.7	39.1

Notes. The table reports the 20 matching-related terms with the highest frequency among the papers in our sample. The first column (*Freq.*) reports the percentage of papers in the sample that contain the matching-related term in its abstract, title, or keywords. The second column (*Cond. Unique*) reports the percentage conditional on having a single matching-related term. The third column (*Is Unique*) reports the percentage of papers with this term that do not have any other matching-related term.

3.3 Validating our algorithm

In this subsection we describe the process we followed to design the algorithm described above, and the steps we followed to validate it. The main advantage of our algorithm is its transparency, though its design warrants further discussion. The choice of terms to identify matching papers was an iterative process through which we took random samples in sequence, evaluated outcomes, and calibrated the list. In the process, we manually labeled 2,716 papers as being about matching or not by inspecting their titles, abstracts, and keywords. Labeling papers was not a trivial process.[22] Occasionally, we differed in our views on what should be considered a paper about matching. Likewise, given the brevity of abstracts, at times we were in doubt whether the main aspect of a paper was matching or if matching was just mentioned in passing. From this process, we concluded that labeling papers manually would invariably entail arbitrary choices.[23] Hence, we opted for a simple and transparent design. Given our data, our algorithm encapsulates what we consider to be our working definition of what constitutes a paper about matching.

We validated the performance of the algorithm in two ways. First, we computed its Type I and Type II Errors within the sample of papers we labeled manually. In our setting, the algorithm incurs in Type I Error if it fails to classify a paper as being about matching, which we deemed to be about matching. It incurs in Type II Error if it classifies a paper as being about matching when we did not labeled it as being about matching. In the last sample we took, after which we did not adjust the algorithm, it had a Type I Error of 12.09% and Type II Error of 5.5%.[24] That is, on average, we estimate that our algorithm captures 87.91% of

[22]We split labeling tasks among us and cross-validated our choices. We designed a bespoke labeling wizard to speed up the process. A screen capture of the wizard's UI may be found in the Appendix, see Figure A8.

[23]For the same reason, we concluded that employing research assistants or crowd-sourced platforms, such as Amazon Turk, would lead to a poor labeling. Our labeling task requires knowledge of specific fields within economics, beyond what is typically taught at the undergraduate level or standard core graduate-level courses.

[24]In the last stages of the design, we systematically labeled 4 random samples, each consisting of a random sample of 3% of the 1.6 million papers published between 1975 and 2020 ($N = 38,028$). From each sample, we filtered for papers that have "match" in the title, abstract, or keywords, and manually labeled the resulting

the papers that are about matching and contain the string "match" in their title, abstract, or keywords. And 5.5% of the papers it classifies as being about matching—because they have at least one matching-related JEL code or term—we do not consider to be about matching.[25]

Second, given our labeled sample of papers, we compared the performance of our algorithm to that of three standard off-the-shelf predictive models: Logistic Regression, Support Vector Machine (SVM) (Vapnik, 1995), and Random Forest (Breiman, 2001). We trained the models in a random sample of our labeled sample (80%), and tested their predictive performance in the remaining 20%.[26] The models generate a predicted value of a binary outcome. Given a predicted value and a threshold t, we classify the papers as being about matching if their predicted value is above the threshold. For a threshold $t = 0$, every paper is classified as being about matching, and the Type I Error equals 0 and the Type II equals 1. As the threshold increases and some papers stop being classified as being about matching, the Type I Error increases and the Type II Error decreases. That is, until $t = 1$, and no paper is classified as being about matching, and the Type I Error equals 1 and the Type II equals 0. Figure 2 plots both types of errors for the three predictive models as the threshold goes from 0 to 1. Table 2 reports the errors at and around the thresholds for which the predictive algorithms achieve the same Type I or Type II Error as our algorithm. As it can be observed, regardless of the threshold we choose, our algorithm has lower Type I and Type II Errors than

sample. In total, across the 4 random samples we took, we manually labeled 2,716 papers. The Type I and Type II Errors reported in the text correspond the last random sample, which consists of 706 papers. We stopped after four rounds as the marginal increase in sample size from adding new matching-related keywords became very small, and the Type I/II Errors did not vary significantly.

[25]Examples of Type I Errors are Green and Zhou (2002); Shi and Temzelides (2004), and of Type II Errors are Mendelberg, McCabe, and Thal (2017); Nunley, Pugh, Romero, and Seals (2015). We provide the abstracts of these papers in the Appendix.

[26]We trained the three models on a sample with 2,172 observations. For feature extraction, we applied TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to four text fields: title, abstract, keywords, and journal name. We obtained a total of around 250,000 features, of which we used the 10,000 with the highest frequency to train the models.

the predictive models.[27] Moreover, the size of the final sample size is very sensitive to the choice of threshold (last column in Table 2). The fact that our simple algorithm out-performs the predictive models is reassuring. Nevertheless, it should not be surprising as our training sample is relatively small in comparison to the high-dimensionality of the data.

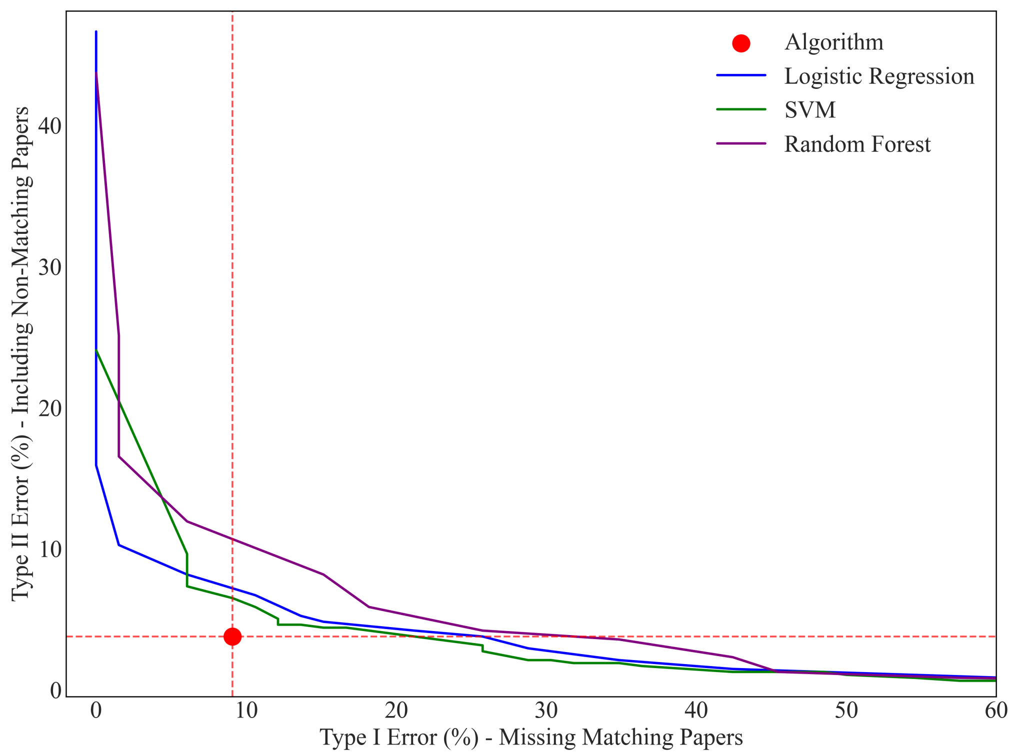

Figure 2: Type II and Type II Errors of Algorithm and Predictive Models

3.4 Measuring diversity across fields and disciplines

To analyze the intellectual diversity and disciplinary structure of the matching literature, we employed hierarchical clustering based on text similarity measures (Murtagh and Contreras, 2012). This methodology allows us to identify distinct research streams within the match-

[27]Note that the Type I and Type II Errors of our algorithm in the test sample are slightly lower than what we estimated in our last calibration step. This is to be expected as, rigorously, a fraction of the test data in this comparison exercise served as training data for our algorithm.

Table 2: Comparison of Algorithm vs Machine Learning: Error Rates and Sample Sizes

Method	*Type I Error (%)*	*Type II Error (%)*	*Final Sample Size*
Algorithm	**9.09**	**3.77**	3,014
Logistic (t=0.250)	1.52	10.25	127,030
Logistic (t=0.330)	**9.09**	7.11	67,058
Logistic (t=0.520)	25.76	**3.77**	23,314
Logistic (t=0.750)	54.55	1.05	5,875
SVM (t=0.014)	7.58	6.69	77,976
SVM (t=0.016)	**9.09**	6.49	72,569
SVM (t=0.075)	22.73	**3.77**	31,620
SVM (t=0.250)	30.30	2.09	15,489
Random Forest (t=0.250)	6.06	13.39	90,263
Random Forest (t=0.270)	**9.09**	11.30	75,172
Random Forest (t=0.440)	31.82	**3.77**	17,764
Random Forest (t=0.500)	42.42	2.51	10,806

Notes: The table reports the Type I and Type II Errors of our algorithm and the three predictive models we trained. The error rates were calculated using a test sample ($n = 544$) different to the training sample. For each model, we report the errors at threshold values that are around the errors our algorithm achieves in the test sample. (See Figure 2 for the curves describing the error rates at all thresholds.) The last column reports the size of the sample of matching papers we obtain from applying each model to the full RePEc dataset (1.6 million papers) using distinct thresholds.

ing literature and examine how the field has differentiated across various applications and theoretical frameworks without imposing any external structure on the data.

Our clustering approach began with the construction of document vectors using the Term Frequency-Inverse Document frequency (TF-IDF) method. For each paper in our sample, we calculated the TF-IDF scores that weight each word according to their frequency in the abstract and their rarity across the corpus.[28] This measurement emphasizes distinctive terminology that characterizes specific research areas while down-weighting common words that appear throughout the literature. Formally, the TF-IDF score for term t in document (paper) d is calculated as:

$$\text{TF-IDF}(t, d) = TF(t, d) \times IDF(t), \tag{1}$$

where $TF(t, d)$ is the frequency (number of occurrences) of term t in document d, and $IDF(t)$ is the inverse document frequency, calculated as:

$$IDF(t) = \log\left(\frac{N}{n_t}\right), \tag{2}$$

with N equal to the total number of documents, and n_t the number of documents containing term t. While $TF(t, d)$ increases with the prevalence of a word t in a document d, $IDF(t)$ gives lower weight to terms that appear in more documents. Hence, TF-IDF scores capture words that are central to specific documents. After constructing the TF-IDF vectors for each

[28]Our text pre-processing pipeline consisted of several steps to ensure robust analysis. First, we cleaned the abstract text by removing non-alphabetic characters (e.g., dash symbols) and converting all text to lowercase. We implemented language normalization by converting British English spellings to American English (e.g., "labor" to "labor") using a comprehensive spelling dictionary to ensure consistency across the corpus. We then applied lemmatization using NLTK's WordNetLemmatizer to reduce words to their base forms (e.g., "matches" to "match"), helping to standardize variations of the same terms. Since some papers' abstracts are multilingual, we developed a stop-word filtering approach that combined stop-words from multiple languages (English, French, German, Italian, and Spanish). This stop-word list was further enhanced with domain-specific terms, such as common academic phrases (e.g., "paper", "study", "result") and generic prepositions. In total, we excluded over 200 non-informative words from our analysis.

paper, we computed the pairwise Euclidean distance between papers to measure their textual similarity and used hierarchical clustering.

Hierarchical clustering is an unsupervised learning method that sequentially clusters observations according to their similarity. The algorithm begins with each paper as its own singleton cluster, then iteratively merges the two most similar clusters based on their pairwise distance.[29] This bottom-up approach, known as agglomerative clustering, continues until all papers form a single cluster, creating a tree-like structure—known as a dendrogram—that reveals the hierarchical relationships within the data. When compared to other clustering algorithms, such as K-means or DBSCAN, an advantage of hierarchical clustering is that it does not require pre-specifying the number of clusters or density parameters. This makes hierarchical clustering more suitable for discovering natural groupings in the data (Murtagh and Contreras, 2012).[30]

3.5 Measuring gender authorship

To investigate the gender dynamics within the matching literature community, we had to determine the gender of authors in our dataset. Since the RePEc bibliographic database does not include gender information, we employed the Genderize API to predict authors' genders based on their names. This API assigns gender probabilities based on a large database of name-gender associations across multiple countries and languages, allowing us to classify authors as male or female with reasonable confidence. The Genderize API has 908,290,909 data entries collected from all over the web (Genderize, 2025). For example, for all names collected from the Olympic athletes in 2024 (11,074), the accuracy rate of the algorithm is 93.87%.

[29] We use Ward's minimum variance method to measure distances between clusters (Ward Jr., 1963; Kaufman and Rousseeuw, 1990).

[30] In earlier versions of the paper we used K-means, and found that varying the number of clusters generated partitions that became finer but at times overlapped each other in non-trivial ways. Rather than choosing a specific level of granularity for the clustering exercise, we opted to analyze the whole structure by using hierarchical clustering.

4 Main results

This section presents the main findings, highlighting key patterns in the field's growth and evolution, diversity within the field, and team size and gender composition in coauthorship. Based on a sample of 3,014 papers, we use bibliometric and text analysis methods to provide a comprehensive overview of how the matching literature has evolved over the past half century.

First, we track the growth of matching research in RePEc relative to the broader economics literature, analyzing trends in publication volume and journal placement, with particular attention to top-tier versus other outlets. Second, we map the structure of matching-related research, showing how the field has differentiated into distinct streams with varying conceptual frameworks, and specific terminology. Finally, we examine author participation, collaboration patterns, and gender representation, comparing these dynamics to broader trends in the economics profession.

4.1 Growth and evolution of the field

In this subsection we document patterns of growth and diffusion of the matching literature, as revealed through our bibliometric analysis. Figure 3 illustrates the relative prevalance of matching papers within the broader economics literature from 1975 to 2020. While the total number of RePEc papers has grown steadily over this period (from around 10,000 in 1975 to over 135,000 by 2020), the proportion of matching papers has followed an overall well-developed trajectory. In the field's early stages, during the 1970s and early 1980s, matching papers constituted a small fraction of the economics literature, with approximately 2 papers per 10,000 RePEc papers. There is a surge witnessed in the mid 1980s, with the proportion increased to 7 papers per 10,000 by 1985. This growth continued steadily through the 1990s, reaching at a significant peak around 2000, when the proportion jumped to over 20 matching papers per 10,000. Since then, the relative share has stabilized, maintaining the similar level around 20 papers through 2020.

The community of scholars engaged in matching research has also evolved distinctively, as shown in Figure 4. The number of authors publishing at least one matching paper per 10,000 RePEc authors (which denotes number of authors publishing at least one paper during the same period) increased dramatically from about 2.5 in 1980 to a peak of nearly 23 around 2005. The relative proportion of matching authors, however, has declined since 2005, dropping to about 17 authors per 10,000 by 2000, despite the continued growth in matching publications shown in Figure 3. This pattern implies a consolidation of matching research within a more specialized community of scholars rather than continuing expansion across the discipline.

The diffusion of matching literature across economics is analyzed in Figure 5. This figure shows the field's increasing prevalence within the economics profession. In 1975, less than 1 journal per 100 in RePEc published matching papers. By 2020, this ratio had increased to nearly 8 journals per 100. The most substantial expansion occurred during the 1990s, from less than 4 to more than 6 journals per 100. This positive growth reflects how matching papers have progressively extended their presence across a broader range of publication outlets; even if the number of new journals is still growing, matching related research keep emerging toward more general audiences.

Figure 6 shows a particularly notable trend in the publication of matching papers in top versus non-top jounals. While matching papers in non-Top-5 journals showed early growth in the 1980s and a substantial jump between 1990 and 2000, Top-5 journals (the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economic Studies*) exhibited a more gradual but persistent increase in matching content. Most significantly, the rate of matching papers in Top-5 journals has accelerated since around 2005, rising from about 2 papers per 100 to 5 papers per 100 by 2020. This increase demonstrates that the field's growing prominence and acceptance within mainstream economics, with elite journals increasingly recognizing the theoretical and empirical contributions of matching research.

Finally, in Table 3, we can see that the top 30 journals publishing the 3,014 matching papers in our sample include all five of the most prestigious economics journals. Two of these—the *American Economic Review* and the *Review of Economic Studies*—are also among the top five journals by volume of matching papers, with the *American Economic Review* ranking first overall and the *Review of Economic Studies* fifth. The *Journal of Political Economy* and *Econometrica* appear in the top 10, ranked 8th and 9th, respectively. The *Quarterly Journal of Economics* is the lowest among them in terms of matching paper output, ranking 29th.[31]

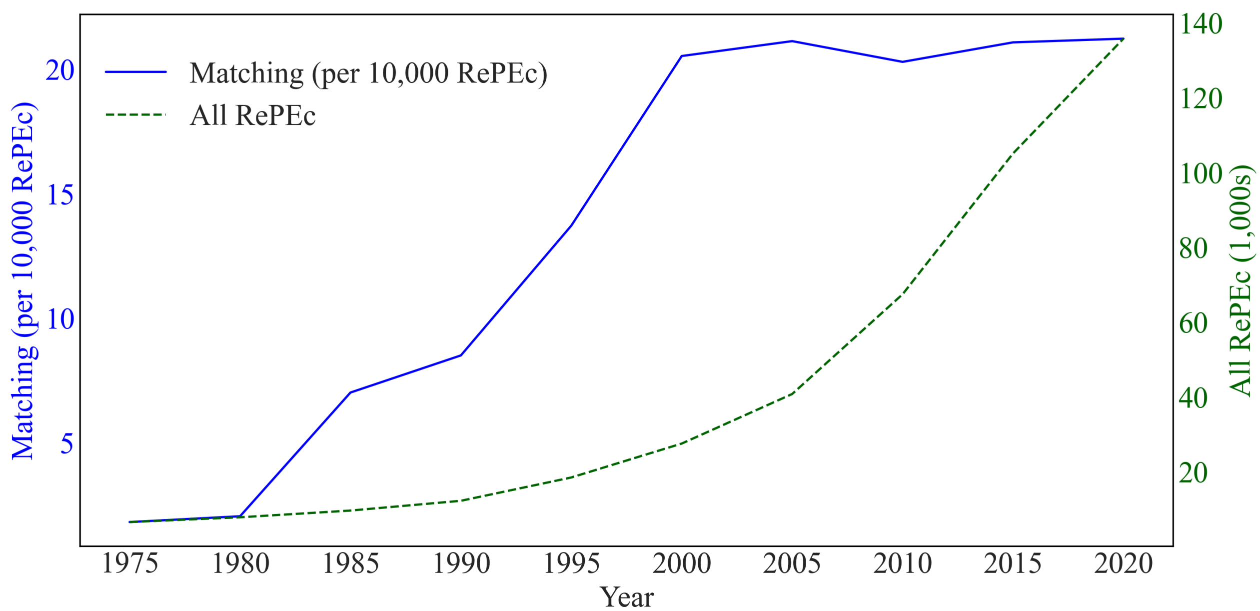

Figure 3: Number of Matching Papers Published over Time

[31]Interestingly, the *Review of Economics and Statistics*, once regarded as roughly comparable to the top five (Angrist, Azoulay, Ellison, Hill, and Lu, 2020; Angrist and Diederichs, 2024), does not appear among the top 30 journals publishing matching papers.

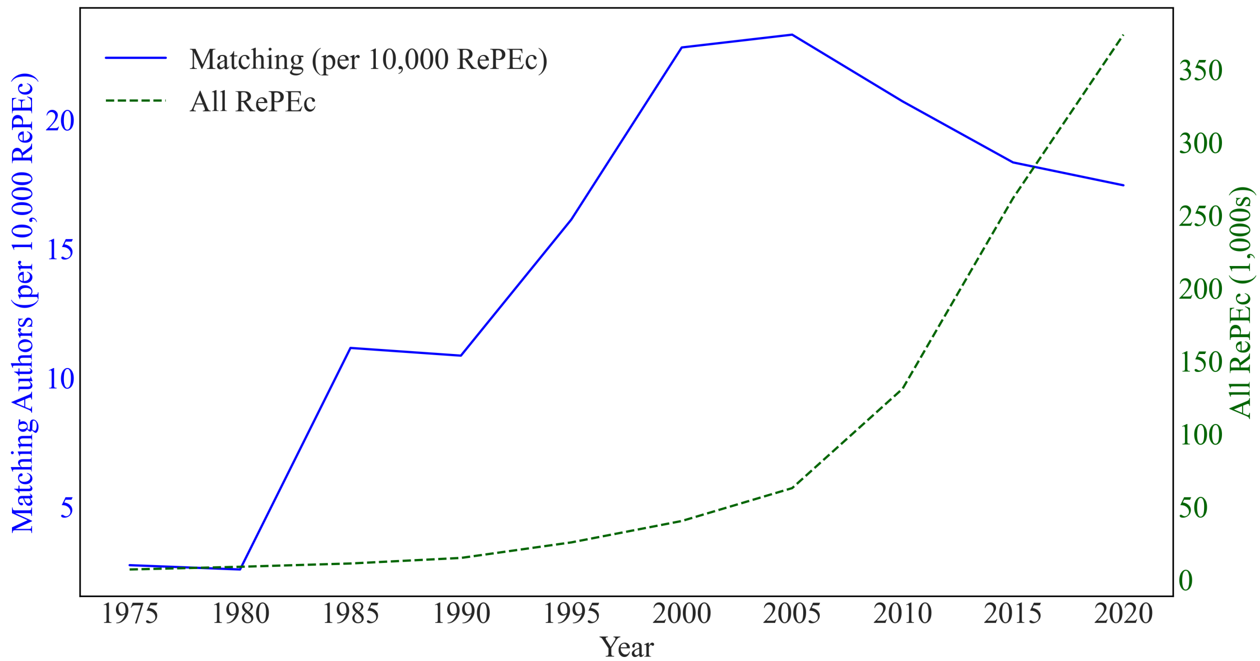

Figure 4: Number of Authors Publishing Matching Papers per 10,000 RePEc Authors

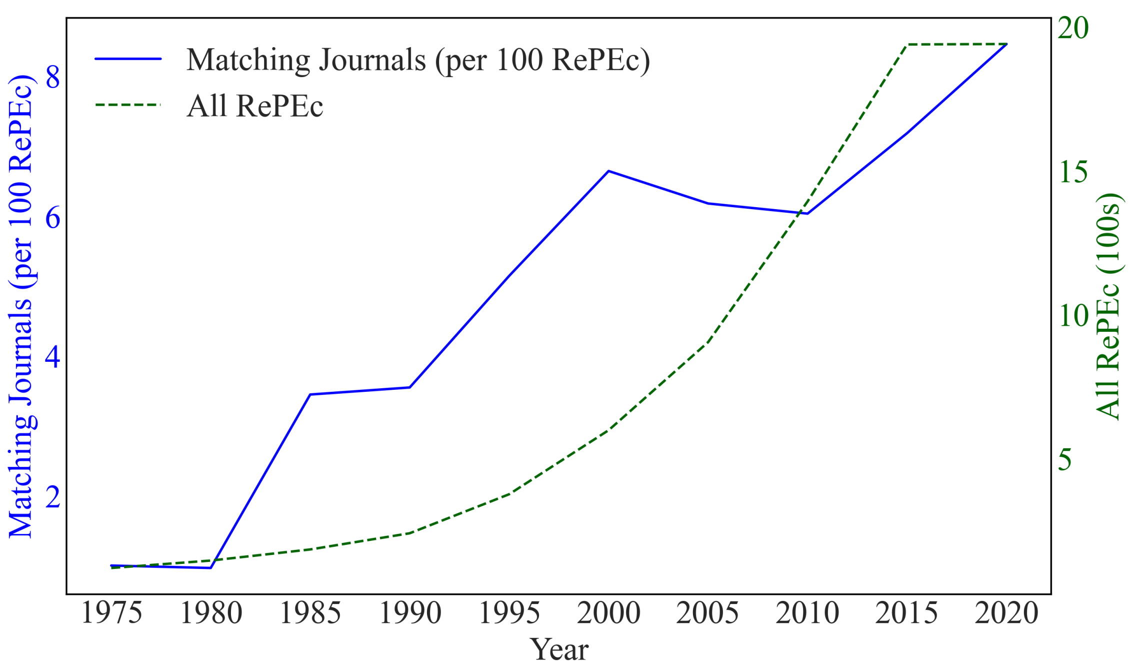

Figure 5: Number of Journals Publishing Matching Papers per 100 RePEc Journals

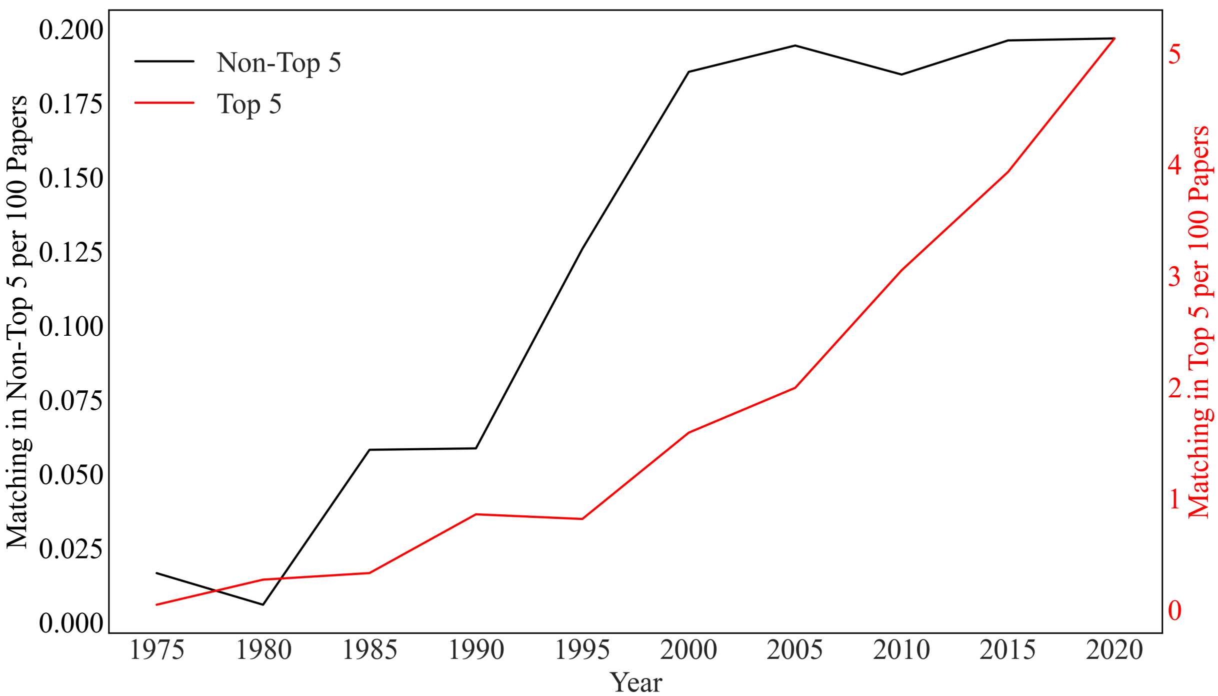

Figure 6: Number of Matching Papers per 100 Papers in Top-5 and Non-Top-5 Journals

Table 3: Top 30 Journals in the Matching Papers Sample

Journal	Count	Percent (%)
1. **American Economic Review**	102	3.38
2. Economics Bulletin	90	2.99
3. International Economic Review	71	2.36
4. Review of Economic Dynamics	71	2.36
5. **Review of Economic Studies**	67	2.22
6. Journal of Combinatorial Optimization	64	2.12
7. Journal of Labor Economics	60	1.99
8. **Journal of Political Economy**	51	1.69
9. **Econometrica**	47	1.56
10. PLOS ONE	31	1.03
11. Theoretical Economics	31	1.03
12. The B.E. Journal of Macroeconomics	30	1.00
13. American Economic Journal: Microeconomics	29	0.96
14. Applied Economics	28	0.93
15. Urban Studies	27	0.09
16. Southern Economic Journal	27	0.90
17. Journal of the European Economic Association	24	0.80
18. Journal for labor Market Research	23	0.76
19. Annals of Economics and Statistics	23	0.76
20. Economic Journal	22	0.73
21. International Journal of Game Theory	22	0.73
22. Canadian Journal of Economics	22	0.73
23. Revue économique	22	0.73
24. American Economic Journal: Macroeconomics	21	0.70
25. ILR Review	20	0.66
26. Journal of Money, Credit and Banking	20	0.66
27. Oxford Economic Papers	19	0.63
28. Canadian Journal of Economics/Revue canadienne d'économique	19	0.63
29. **The Quarterly Journal of Economics**	19	0.63
30. The B.E. Journal of Theoretical Economics	18	0.60

Notes: The table reports the thirty journals most common in our sample of matching papers. Top-5 journals are displayed in **bold**. The first column reports the number of papers in the sample published by each journal. The second column reports the same figure as a percentage of the number of papers in our sample.

4.2 Diversity within the matching literature

In this subsection we investigate the "different literatures" withing the matching literature. Figure 7 displays a dendrogram: a diagram that shows how clusters of papers are arranged through hierarchical clustering of the matching literature. The horizontal axis represents the Euclidean distance, with papers grouped at varying levels of similarity; greater distance indicates lower similarity and leads to earlier separation. We identify seven major clusters, which are colour-coded and labelled according to their primary research focus. The legend indicates the proportion of papers in each cluster. Inspecting the figure from left to right, we observe that Macroeconomics papers are first separated from Microeconomics papers, suggesting that research in these fields employs distinct language patterns. Within the Microeconomics branch, we identify clusters related to mathematical matching theory, school choice, housing markets, marriage markets, and general matching theory. The Macroeconomics branch mainly splits into macro-labor papers and other research broadly related to Macroeconomics within the matching field.

In Figure 8 we present a more detailed view of the clustering using a flowchart, which uses the same coloring as Figure 7. Each node displays the percentage of papers in that cluster, along with the top five terms with the highest TF-IDF scores.[32] The direction of the flowchart from left to right mirrors the dendogram as the distance in the x-axis decreases towards zero. The literature splits early into two broad branches—Macroeconomics (brown cluster, e.g. "job, labor, unemployment, worker, wage"), which represents 36% of the matching papers, and Microeconomics (blue cluster, e.g. "agent, marriage, preference, mechanism, equilibrium", which represents 64% of the matching papers)—suggesting distinct vocabularies across fields. The increasing level of specialization captured by the clustering can be appreciated in some of the smallest nodes reported in the flowchart. For example in the macro

[32]The hierarchical clustering does not use our list of matching-related terms. Once we have selected the sample of matching papers, we apply hierarchical clustering to the whole text data in abstracts.

branch, the most frequent terms in the smallest clustered shown in the flowchart are "curve, beveridge, shift, labor, outward", and on the three smallest on the micro side are "graph, vertex, perfect, pairdominating, dominating", "seller, buyer, price, intermediary, equilibrium", and the "money, currency, random, outside, holding." Each of these clusters is composed by around 30 papers (1% of the sample).

This hierarchical structure illustrates how the matching literature has diversified into distinct yet interconnected research streams. This analysis allows us to quantify the relative size of various subfields within matching-related research and to visualize how closely these subfields are related, based on the similarity of the language used in their scholarly writing.

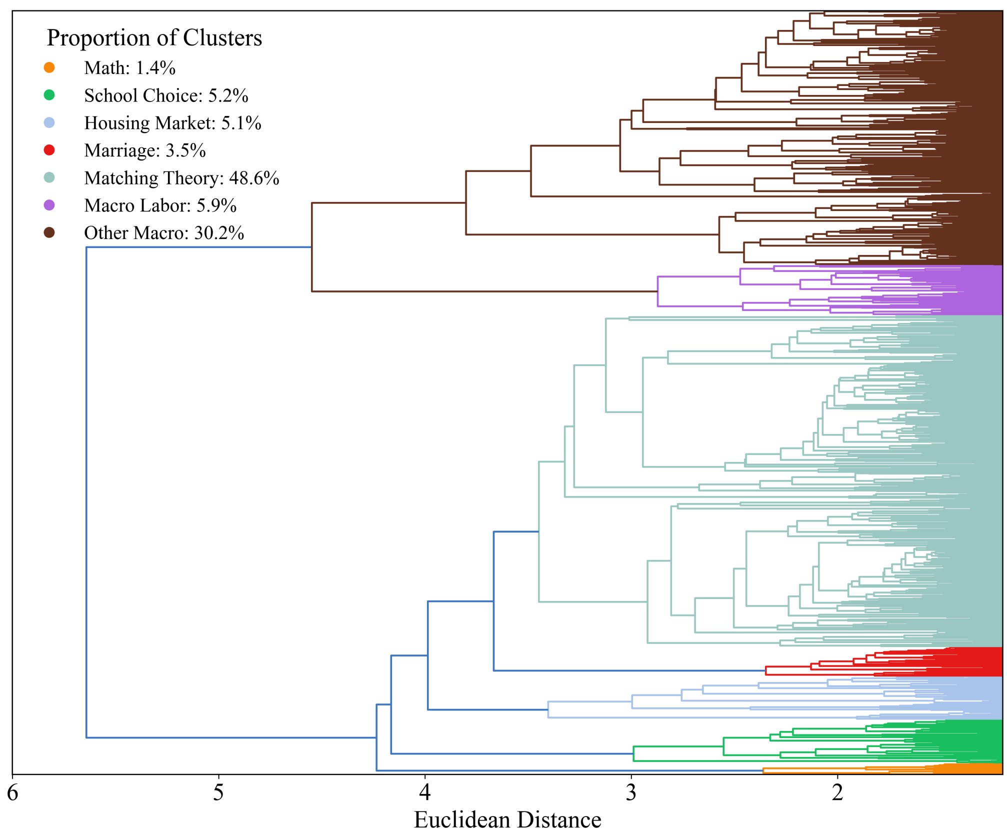

Figure 7: Hierarchical Clustering of Matching Papers

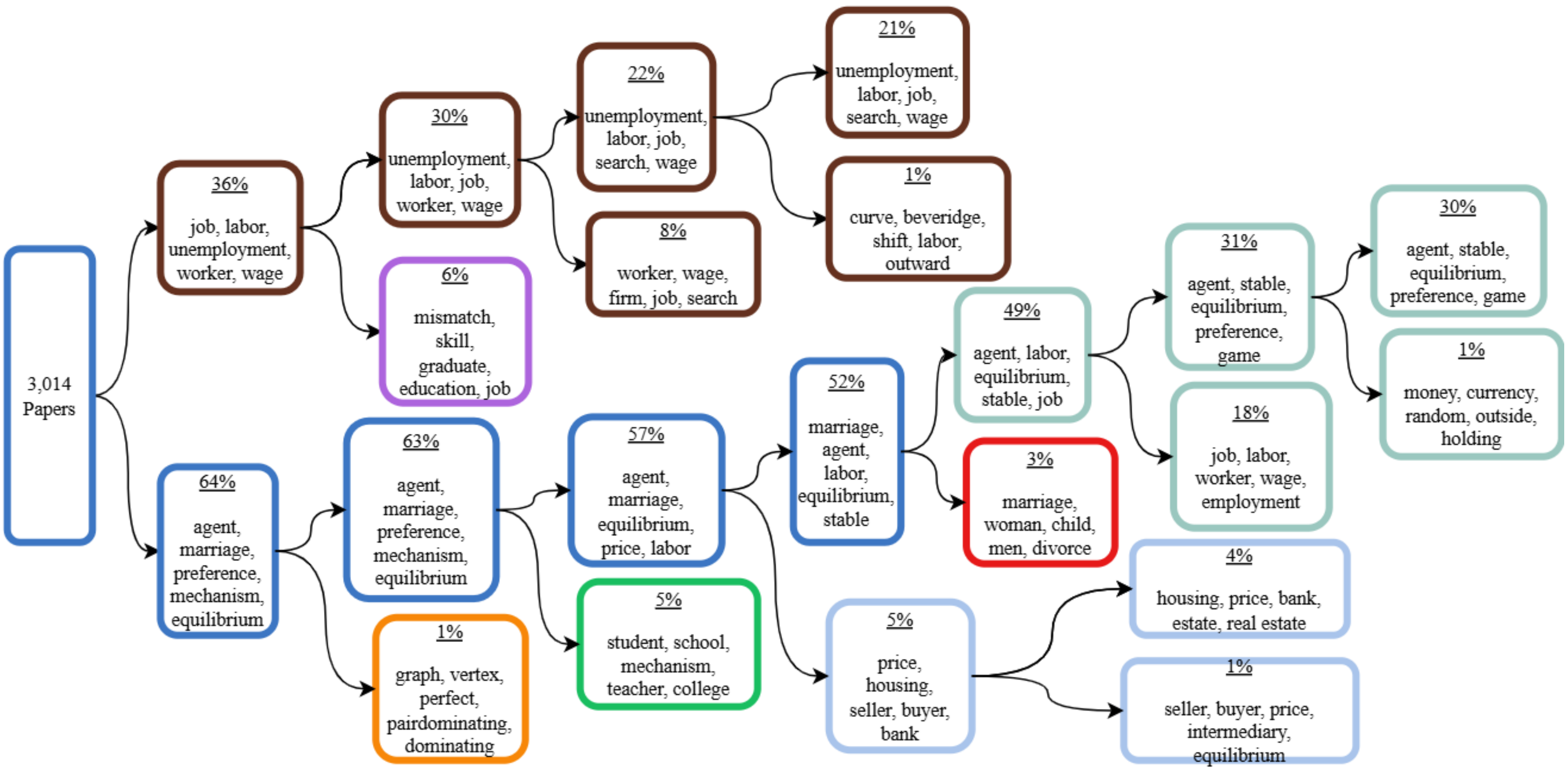

Figure 8: Hierarchical Clustering Dendrogram: flow chart

4.3 Collaboration dynamics and gender composition

Finally, in this last subsection, we focus on collaboration dynamics and gender composition. In Figure 9, we plot the average number of authors per paper from 1990 to 2020, comparing matching papers to all RePEc papers. Both series show a steady increase in collaboration over time, reflecting a broader trend toward co-authorship in economics. During the 1990s and early 2000s, matching papers and all RePEc papers followed a similar trajectory, with only slight differences in team size, from around 1.5 to 2. However, from the mid-2000s onward, their paths began to diverge. The average number of authors on all RePEc papers rose sharply, surpassing three authors per paper by 2020. In contrast, matching papers experienced a more moderate increase, reaching approximately 2.3 authors per paper by the end of the period. This suggests that, despite growing interest and output in the field, matching research has continued to be conducted in relatively smaller teams compared to the broader economics discipline.[33]

While Figure 9 is informative about the intensive margin of collaboration, it remains silent about the extensive margin, which is the focus of Figure 10. Here, we track the prevalence of co-authorship over the period 1990–2020. Up until around 2015, matching papers consistently exhibited a higher degree of collaboration, as measured by the share of co-authored papers. Co-authorship became the norm in matching research in the early nineties, whereas this trend only took hold in the broader economics literature about a decade later, with an initial gap in co-authorship rates of over 10 percentage points. Although the average team size is now larger among all economics papers compared to matching (Figure 9), the fraction of co-authored papers is currently similar—around 75%.[34]

[33] If we extend the analysis to cover the period from 1975 to 2020, as shown in Figure A2 in the online appendix, we introduce some additional noise and variability in the earlier years due to smaller sample sizes. Nonetheless, the key qualitative patterns remain intact.

[34] If we extend the analysis to cover the period from 1975 to 2020, as shown in Figure A3 in the appendix, similar qualitative patterns emerge.

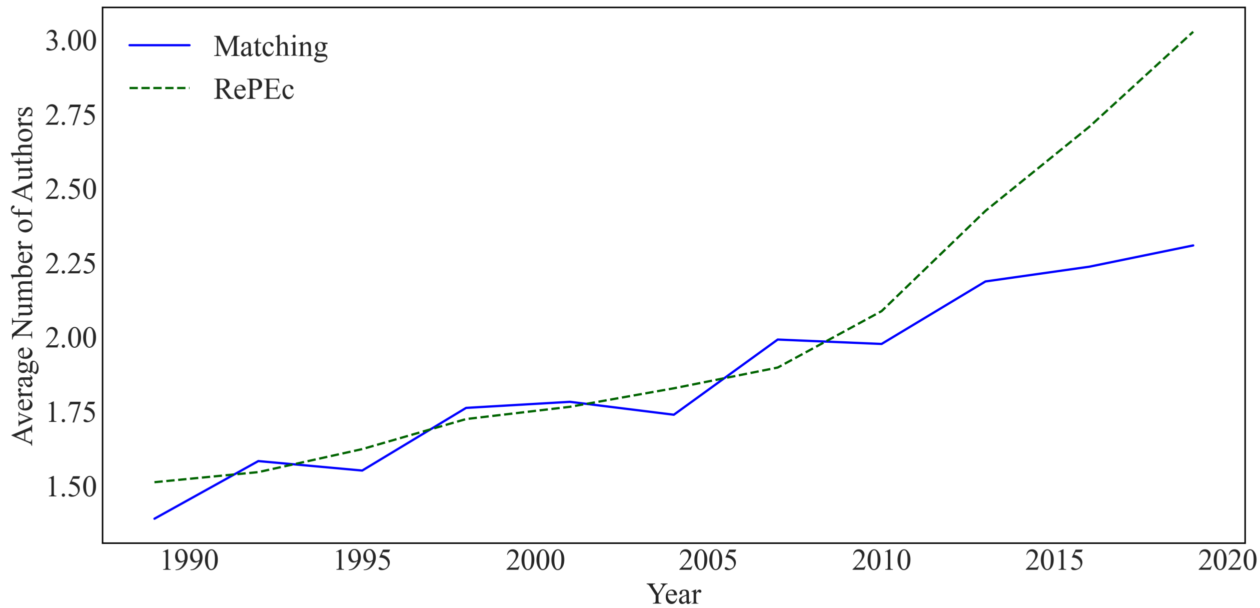

Figure 9: Average Number of Authors per Paper, 1990-2020, by 3-year bins

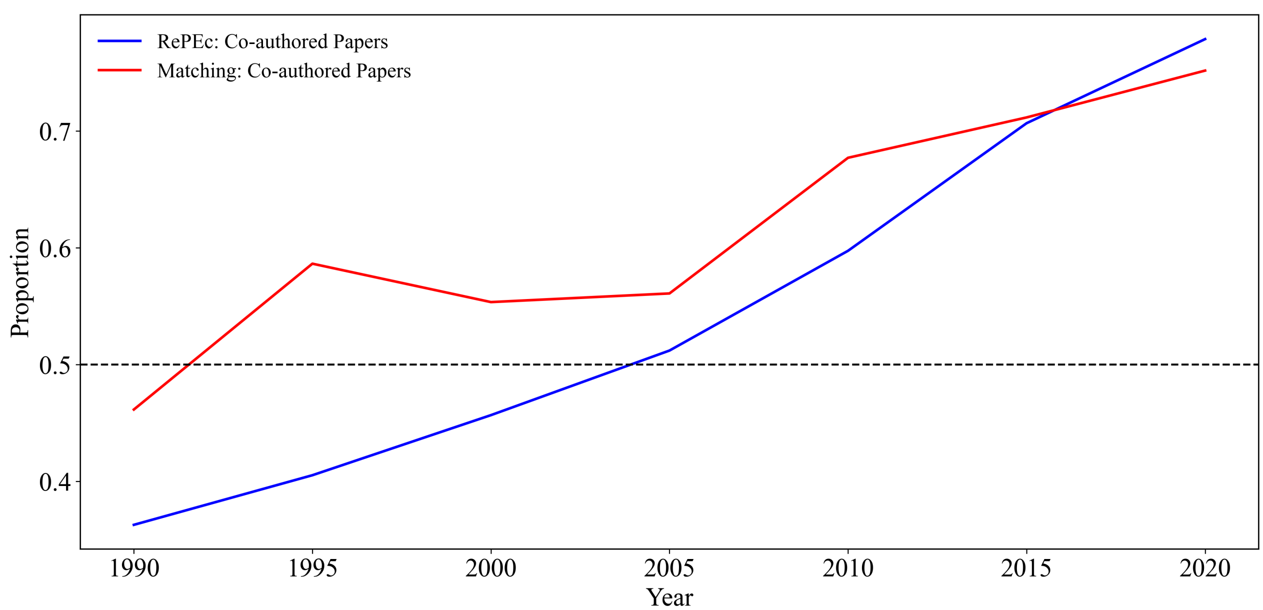

Figure 10: Evolution of Co-authored Papers, 1990–2020, by 5-year bins

In Figure 11, we examine the evolution of female authorship in matching and all economics papers in RePEc since 1990. The figure shows that matching has consistently lagged behind the broader economics literature in terms of women's representation. In 1990, only 10% of matching papers included a female author, compared to around 20% for all economics papers. Although female participation in authorship has increased over time in both groups, the gap has remained fairly stable at about 10 percentage points, equivalent to a lag of about 10 years. By 2020, women were authors on approximately 25% of matching papers and 35% of all papers in RePEc.[35]

We now examine the gender composition of papers in Figure 12, distinguishing between male-only, female-only, and mixed-gender co-authorship. Several patterns are worth highlighting. First, there is a clear decline in all-male authorship: both RePEc and matching papers show a steady decrease in the proportion of male-only authored papers. Matching papers began at a higher level (around 90% in 1990) and remain more male-dominated in 2020 (around 50%). RePEc papers started lower (around 80%) and dropped more sharply, falling below 40% by 2020. Second, mixed-gender authorship has risen: RePEc papers show a strong upward trend, with mixed-gender teams becoming the most common category by 2020 (close to 50%). Matching papers also show growth in this category, reaching around 35% by 2020, but still lagging behind RePEc. Third, the share of female-only authored papers remains low and relatively stable across both groups. By 2020, all-female papers account for just over 10% in both matching and RePEc, with matching slightly lower. Finally, there is a persistent gender gap in the matching field: throughout the period, matching papers consistently lag behind the broader RePEc papers in gender diversity, reflected in both a higher share of male-only authorship and a lower share of mixed-gender teams.

[35] If we extend the analysis to cover the period from 1975 to 2020, we get very noisy estimates for matching papers pre-1990, as shown in Figure A4 in the appendix. Both the growth of the economics research community and the significant increase in the share of women in the profession are consistent with the analysis by Ductor, Goyal, and Prummer (2023) using data drawn from the EconLit database.

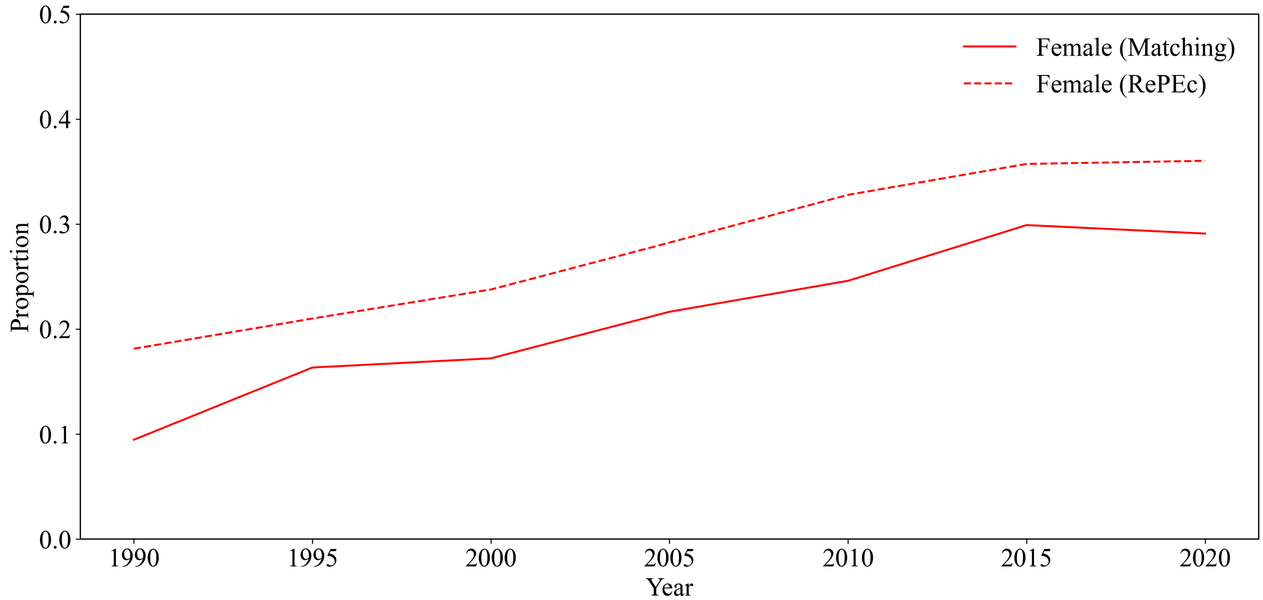

Figure 11: Evolution of Female Economists, 1990-2020, by 5-year bins

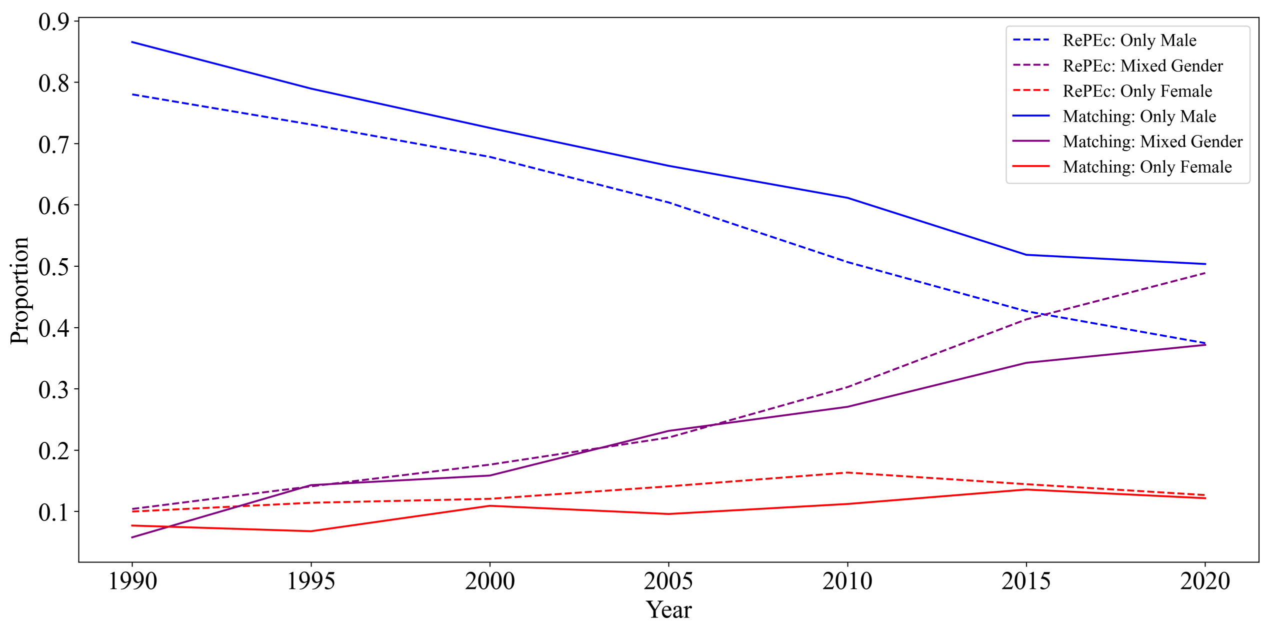

Figure 12: Evolution of Gender Composition in Economics Papers, 1990–2020, by 5-year bins

Finally, we examine collaboration patterns by disaggregating them into same-gender (all-male or all-female) and cross-gender (male and female) collaborations. In Figure 13, we focus on all RePEc papers. Four clear trends emerge over the period from 1990 to 2020. First, there has been a sustained decline in single-authored papers by men, which fell from 55% in 1990 to just 15% in 2020. Second, mixed-gender collaborations have seen a consistent rise, increasing from 10% in 1990 to 50% in 2020. Notably, the crossover point occurred in the late 2000s, when the share of male single-authored papers and mixed-gender collaborations both reached 30%. Third, the proportion of papers authored exclusively by male teams remained relatively stable, hovering around 25% throughout the period. Finally, papers authored either solely by women or by all-female teams consistently represented a small share, remaining at or below 10% across the three decades. These patterns highlight both the growing role of gender diversity in academic collaboration and the persistent underrepresentation of women in sole and same-gender coauthored publications.

We then replicate the analysis for the sample of matching papers in Figure 14. The findings are qualitatively similar, though the magnitudes differ. Single-authored papers by men declined markedly, dropping from 50% in 1990 to 20% in 2020. In contrast, mixed-gender collaborations increased substantially, rising from just 5% in 1990 to 38% by 2020. In the late 2000s, both single-authored and mixed-gender papers represented approximately 25% of publications. Over the same period, the share of papers produced exclusively by male teams remained stable at around 38%. Meanwhile, the share of publications involving only female authors—either single-authored or in all-female teams—persistently remained below 10%. Together, these trends underscore a broader movement toward more gender-integrated scholarly collaboration, while reaffirming the continued low representation of women in single-gender authorship.

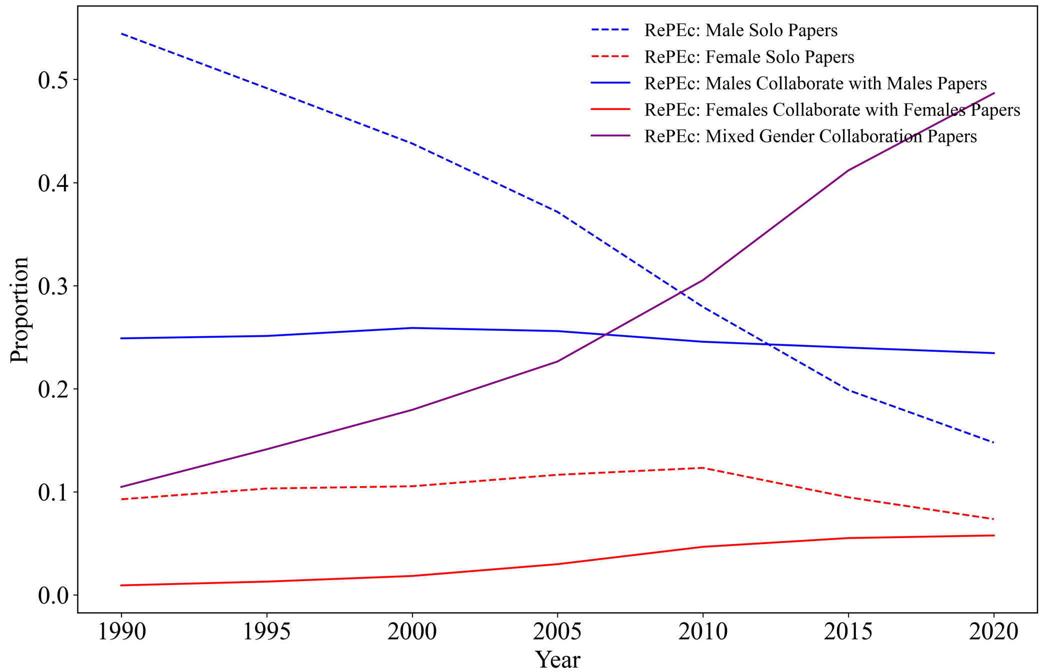

Figure 13: Gender Composition Trends in all RePEc by Collaboration Type, 1990-2020, by 5-year bins

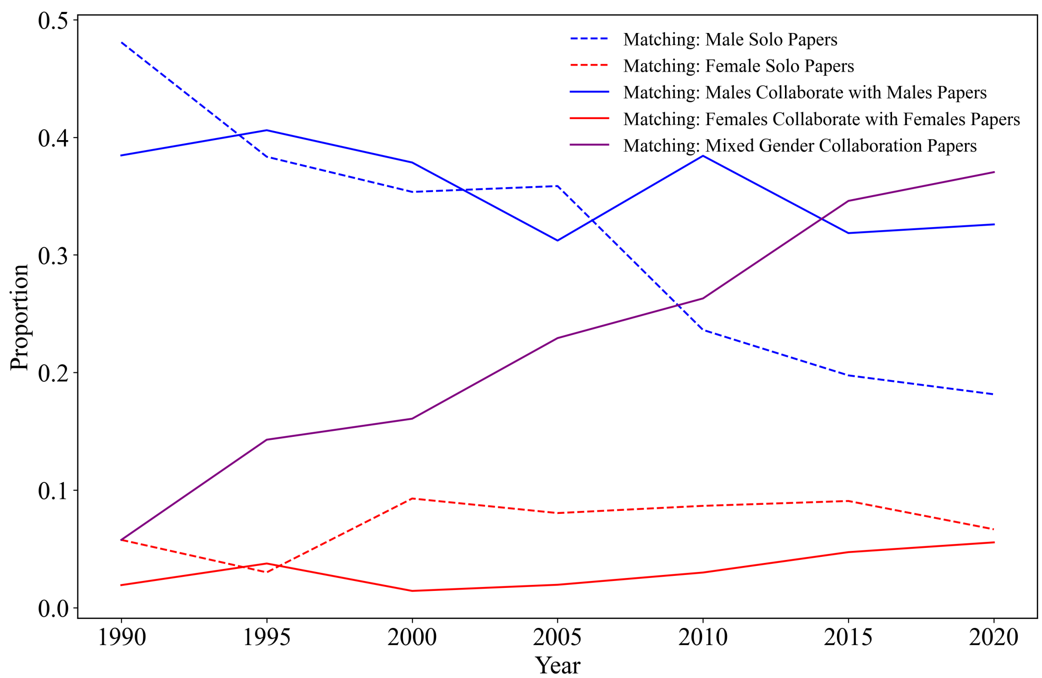

Figure 14: Gender Composition Trends in Matching Papers by Collaboration Type, 1990-2020, by 5-year bins

Taken together, three key messages emerge from this analysis. First, collaboration has increased across the board, but the pace and scale of growth differ: matching research continues to involve relatively smaller teams and has not mirrored the broader economics literature's shift toward larger coauthor groups. Second, although female authorship has increased over time, matching papers remain more male-dominated, with higher rates of male-only authorship and lower female participation compared to the broader field. Third, the gains in female authorship have been driven primarily by a rise in mixed-gender collaborations rather than an increase in all-female teams—a pattern especially pronounced in the broader economics literature, where mixed-gender teams have become the dominant form of coauthorship. This suggests a broader tendency toward gender-integrated collaboration over same-gender partnerships. In contrast, while mixed-gender collaboration has also grown within the matching subfield, it remains less prevalent, reinforcing the pattern of slower progress toward gender-integrated collaboration in this area.

5 Conclusion and discussion

This paper provides the first large-scale, bibliometric and text-based mapping of the matching literature within economics. Drawing on a sample of over 3,000 matching papers from an initial dataset of 1.6 million RePEc entries published between 1975 and 2020, we document how the field has grown, fragmented into distinct research streams, and evolved in terms of authorship and collaboration.

Our findings reveal several notable trends. First, matching research has achieved widespread visibility, with increasing representation across journals—including Top-5 outlets—and growing institutionalization as a core area of inquiry. However, its expansion has stabilized in recent decades, and it remains relatively concentrated within a specialized group of contributors.

Second, the matching literature displays rich internal diversity, spanning both macroeconomic and microeconomic domains, with applications ranging from labor markets to school choice and marriage. Clustering analysis reveals that these subfields have often evolved in relative isolation, suggesting a degree of conceptual fragmentation within the broader matching domain.

Third, collaboration has increased across the field, but at a slower pace than in economics overall. Matching papers are still more likely to be authored by smaller and more male-dominated teams. While female authorship has grown, much of this growth has come through mixed-gender collaborations rather than an increase in all-female teams. These trends point to a shift toward gender-integrated authorship, but one that has diffused unevenly—matching research remains less gender-integrated than the discipline as a whole.

Despite the depth of our analysis, several open questions remain. As a descriptive study based on publication metadata, our approach does not identify the causal mechanisms behind observed patterns. Why has mixed-gender collaboration diffused more slowly in the matching community? To what extent do institutional, disciplinary, or network factors shape participation and visibility? Future research could examine interactions across thematic streams—through citations, co-authorship networks, or hybrid models—and further investigate the rise in male–female collaborations. Are these trends driven by changing preferences, the growing presence of women in the field, or evolving returns to demographic diversity? More broadly, applying similar analyses to other subfields could reveal whether these patterns are unique to matching or reflect structural shifts in the discipline. Extending the approach to dimensions such as geography, institutional affiliation, or methodology could offer deeper insight into how economics evolves as a collaborative enterprise.

References

Abdulkadiroğlu, A. and T. Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747. 6

Adachi, H. (2000): "On a characterization of stable matchings," *Economics Letters*, 68, 43 – 49. 13

Agarwal, N. and P. Somaini (2023): "Empirical Models of NTU Matching," in *Online and Matching-Based Market Design*, ed. by I. Echenique and Vazirani, Cambridge University Press, chap. 25. 8

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. F. Lu (2020): "Inside Job or Deep Impact? Extramural Citations and the Influence of Economic Scholarship," *Journal of Economic Literature*, 58, 3–52. 23

Angrist, J. and M. Diederichs (2024): "Dissertation Paths: Advisors and Students in the Economics Research Production Function," *NBER Working Paper*, 33281. 23

Atakan, A. E. (2006): "Assortative Matching with Explicit Search Costs," *Econometrica*, 74, 667–680. 7

Becker, G. S. (1973): "A theory of marriage: Part I," *Journal of Political Economy*, 81, 813–846. 2, 6, 8

Bergstrom, T. and M. Bagnoli (1993): "Courtship as a Waiting Game," *Journal of Political Economy*, 101, 185–202. 7

Bonnet, O., A. Galichon, Y.-W. Hsieh, K. O'Hara, and M. Shum (2022): "Yogurts Choose Consumers? Estimation of Random-Utility Models via Two-Sided Matching," *The Review of Economic Studies*, 89, 3085–3114. 8

Breiman, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32. 16

Burdett, K. and M. G. Coles (1997): "Marriage and Class," *Quarterly Journal of Economics*, 112, 141–168. 7

Bürgi, C. and K. Wohlrabe (2022): "The influence of Covid-19 on publications in economics: bibliometric evidence from five working paper series," *Scientometrics*, 127, 5175–5189. 10

CEPR (2025): "CEPR Discussion Papers," Available at: https://cepr.org/publications/discussion-papers. Accessed: 2025-03-12. 9

Chade, H., J. Eeckhout, and L. Smith (2017): "Sorting through Search and Matching Models in Economics," *Journal of Economic Literature*, 55, 493–544. 7

Chade, H. and P. Kircher (2023): "Matching with Search Frictions," in *Online and Matching-Based Market Design*, ed. by I. Echenique and Vazirani, Cambridge University Press, chap. 19. 7

Che, Y.-K., P.-A. Chiappori, and B. Salanié, eds. (2024): *Handbook of the Economics of Matching*, vol. 1, Amsterdam: Elsevier. 5

Chiappori, P.-A. (2017): *Matching with Transfers: The Economics of Love and Marriage*, Princeton: Princeton University Press. 7

Chiappori, P.-A., M. Costa Dias, C. Meghir, and H. Zhang (2025): "Changes in Marital Sorting: Theory and Evidence from the US," *Journal of Political Economy*, forthcoming. 8

Chiappori, P.-A. and B. Salanié (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54, 832–861. 8

Choo, E. and A. Siow (2006): "Who marries whom and why," *Journal of Political Economy*, 114, 175–201. 8

Clarivate (2025): "Social Sciences Citation Index (SSCI)," Available at: https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/web-of-science-core-collection/social-sciences-citation-index/. Accessed: 2025-03-20. 9

Crawford, V. P. and E. M. Knoer (1981): "Job Matching with Heterogeneous Firms and Workers," *Econometrica*, 49, 437–450. 6

Diamond, P. A. (1982): "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 90, 881–894. 2

Ductor, L., S. Goyal, and A. Prummer (2023): "Gender and Collaboration," *The Review of Economics and Statistics*, 105, 1366–1378. 32

Echenique, F., N. Immorlica, and V. V. Vazirani, eds. (2023): *Online and Matching-Based Market Design*, Cambridge University Press. 6

Echenique, F., S. Lee, M. Shum, and M. B. Yenmez (2013): "The Revealed Preference Theory of Stable and Extremal Stable Matchings," *Econometrica*, 81, 153–171. 8

Eeckhout, J. and P. Kircher (2010): "Sorting and Decentralized Price Competition," *Econometrica*, 78, 539–574. 7

Fernández, R., N. Guner, and J. Knowles (2005): "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality," *Quarterly Journal of Economics*, 120, 273–344. 7

Fox, J. T. (2009): "Structural Empirical Work Using Matching Models," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Palgrave Macmillan. 8

Gale, D. and L. S. Shapley (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly*, 69, 9–15. 1, 2, 6, 11

Galichon, A. (2016): *Optimal Transport Methods in Economics*, Princeton, NJ: Princeton University Press. 8

Galichon, A., S. D. Kominers, and S. Weber (2019): "Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility," *Journal of Political Economy*, 127, 2875–2925. 8

Galichon, A. and B. Salanié (2023): "Structural Estimation of Matching Markets with TU," in *Online and Matching-Based Market Design*, ed. by I. Echenique and Vazirani, Cambridge University Press, chap. 26. 8

Genderize (2025): "Genderize API," Available at: https://genderize.io/our-data. Accessed: 2025-04-02. 20

Graham, B. S. (2011): "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," in *Handbook of Social Economics, Volume 1B*, ed. by J. Benhabib, M. O. Jackson, and A. Bisin, Elsevier, North-Holland, 965–1052. 8

Green, E. J. and R. Zhou (2002): "Dynamic Monetary Equilibrium in a Random Matching Economy," *Econometrica*, 70, 929–969. 16

Gusfield, D. and R. W. Irving (1989): *The stable marriage problem: structure and algorithms*, MIT Press. 6

Haeringer, G. (2017): *Market Design: Auctions and Matching*, The MIT Press. 5

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85, 29–65. 7

Hatfield, J. W. and P. Milgrom (2005): "Matching with Contracts," *American Economic Review*, 95, 913–935. 6

IZA (2025): "IZA Discussion Papers," Available at: `https://www.iza.org/content/publications`. Accessed: 2025-04-01. 9

Kantorovich, L. (1939): "Mathematical methods in the organization and planning of production," *Leningrad University*, reprint in Management Science 6:366–422, 1959. 1

——— (1942): "On the translocation of masses," *Doklady Akademii Nauk SSSR*, 37, 227–229, in Russian. English translation in *Journal of Mathematical Sciences*, 133(4):1381–1382, 2006. 1

——— (1948): "On a problem of Monge," *Uspekhi Matematicheskikh Nauk*, 3, 225–226, in Russian. English translation in *Journal of Mathematical Sciences*, 133(4):1383, 2006. 1

Kantorovich, L. and G. Rubinstein (1958): "On a space of completely additive functions," *Vestnik Leningradskogo Universiteta*, 13, 52–59. 1

Kaufman, L. and P. Rousseeuw (1990): *Agglomerative Nesting (Program AGNES)*, John Wiley Sons, Ltd, chap. 5, 199–252. 20

Kelso, A. S. and V. P. Crawford (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50, 1483–1504. 6

Knuth, D. E. (1976): *Marriages Stables*, Montreal: Les Presses de l'Université de Montreal. 6

Koopmans, T. C. (1947): "Optimum utilization of the transportation system," in *Proceedings of the International Statistics Conference*, vol. 5, 136–146. 1

Koopmans, T. C. and M. Beckmann (1957): "Assignment problems and the location of economic activities," *Econometrica*, 25, 53–76. 1, 2, 6

Lee, S. and L. Yariv (2018): "On the Efficiency of Stable Matchings in Large Markets," Working Papers 2018-4, Princeton University. Economics Department. 8

Legros, P. and A. F. Newman (2007): "Beauty Is a Beast, Frog Is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, 75, 1073–1102. 6

Lindenlaub, I. (2017): "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 84, 718–789. 8

Lise, J., C. Meghir, and J.-M. Robin (2013): "Matching, Sorting and Wages," *Review of Economic Dynamics*, 19, 63–87. 7

Lise, J. and J.-M. Robin (2017): "The Macrodynamics of Sorting between Workers and Firms," *American Economic Review*, 107, 1104–1135. 7

Mendelberg, T., K. T. McCabe, and A. Thal (2017): "College Socialization and the Economic Views of Affluent Americans," *American Journal of Political Science*, 61, 606–623. 16

Monge, G. (1781): "Memoire sur la Theorie des Deblais et des Remblais," *Histoire de l'Académie Royale des Sciences*, 666–704. 1

Mortensen, D. T. (1982): "Property Rights and Efficiency in Mating, Racing, and Related Games," *American Economic Review*, 72, 968–979. 7

——— (1988): "Matching: Finding a Partner for Life or Otherwise," *American Journal of Sociology*, 94, S215–S240. 7

——— (2005): *Wage Dispersion: Why Are Similar Workers Paid Differently?*, Cambridge, MA: MIT Press. 7

MORTENSEN, D. T. AND C. A. PISSARIDES (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61, 397–415. 2

——— (2011): *Job Matching, Wage Dispersion, and Unemployment*, Oxford: Oxford University Press. 7

MURTAGH, F. AND P. CONTRERAS (2012): "Algorithms for hierarchical clustering: an overview," *WIREs Data Mining and Knowledge Discovery*, 2, 86–97. 17, 20

NBER (2025): "About the National Bureau of Economic Research (NBER)," Available at: https://www.nber.org/about-nber. Accessed: 2025-03-12. 9

NIEDERLE, M., A. E. ROTH, AND T. SÖNMEZ (2008): "matching and market design," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Basingstoke: Palgrave Macmillan. 6

NUNLEY, J. M., A. PUGH, N. ROMERO, AND R. A. SEALS (2015): "Racial Discrimination in the Labor Market for Recent College Graduates: Evidence from a Field Experiment," *The B.E. Journal of Economic Analysis Policy*, 15, 1093–1125. 16

PETRONGOLO, B. AND C. A. PISSARIDES (2001): "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 39, 390–431. 7

PISSARIDES, C. A. (1985): "Short-run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages," *American Economic Review*, 75, 676–690. 2

——— (2000): *Equilibrium Unemployment Theory*, Cambridge, MA: MIT Press, 2nd ed. 7

Postel-Vinay, F. and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, 2295–2350. 7

Rogerson, R., R. Shimer, and R. Wright (2005): "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 43, 959–988. 7

Roth, A. E. (1984): "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, 92, 991–1016. 2, 11

——— (2002): "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, 70, 1341–1378. 6

——— (2003): "The Origins, History, and Design of the Resident Match," *JAMA*, 289, 909–912. 2

——— (2008): "What Have We Learned from Market Design?" *The Economic Journal*, 118, 285–310. 6

——— (2012): "David Warsh on Lloyd Shapley and his colleagues," Blog post on *Market Design*, accessed: 2025-04-04. 2

——— (2015): "Who Gets What and Why: The New Economics of Matchmaking and Market Design," *American Economic Review*, 105, 385–409. 6, 10

——— (2016): *Who Gets What — and Why*, New York: First Mariner Books. 6

Roth, A. E. and E. Peranson (1999): "The redesign of the matching market for American physicians: Some engineering aspects of economic design," *American Economic Review*, 89, 748–780. 6

Roth, A. E., T. Sönmez, and M. U. Ünver (2004): "Kidney Exchange," *The Quarterly Journal of Economics*, 119, 457–488. 6

Roth, A. E. and M. Sotomayor (1990): *Two-Sided Matching: A Study in Game-Theoretic Modelling and Analysis*, Econometric Society Monographs. 6

Roth, A. E. and J. H. V. Vate (1990): "Random paths to stability in two-sided matching," *Econometrica*, 1475–1480. 6

Sattinger, M. (1993): "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31, 831–880. 6

Scopus (2025): "Scopus: Scopus Content," Available at: https://www.elsevier.com/en-gb/products/scopus/content. Accessed: 2025-03-20. 9

Shapley, L. and H. Scarf (1974): "On cores and indivisibility," *Journal of Mathematical Economics*, 1, 23 – 37. 6

Shapley, L. S. and M. Shubik (1971): "The assignment game I: The core," *International Journal of Game Theory*, 1, 111–130. 1, 6, 11

Shi, S. and T. Temzelides (2004): "A MODEL OF BUREAUCRACY AND CORRUPTION," *International Economic Review*, 45, 873–908. 16

Shimer, R. (2005): "The Assignment of Workers to Jobs in an Economy with Coordination Frictions," *Journal of Political Economy*, 113, 996–1025. 7

Shimer, R. and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68, 343–369. 7

Smith, L. (2006): "The Marriage Model with Search Frictions," *Journal of Political Economy*, 114, 1124–1144. 7

——— (2011): "Frictional Matching Models," *Annual Review of Economics*, 3, 319–338. 7, 8

SSRN (2025): "SSRN: Social Science Research Network," Available at: https://www.ssrn.com/index.cfm/en/. Accessed: 2025-03-12. 9

Tinbergen, J. (1956): "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, 77, 155–175. 7

Vapnik, V. (1995): *The nature of statistical learning theory*, New York: SpringerVerlag. 16

Vulkan, N., A. E. Roth, and Z. Neeman, eds. (2013): *The Handbook of Market Design*, Oxford: Oxford University Press. 6

Ward Jr., J. H. (1963): "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58, 236–244. 20

Wright, R., P. Kircher, B. Julien, and V. Guerrieri (2021): "Directed Search and Competitive Search Equilibrium: A Guided Tour," *Journal of Economic Literature*, 59, 90–148. 7

Zimmermann, C. (2013): "Academic Rankings with RePEc," *Econometrics*, 1, 249–280. 9

Appendix

Examples of Type I Error.

1. Green, E.J. and Zhou, R. (2002), "Dynamic Monetary Equilibrium in a Random Matching Economy". *Econometrica*, 70: 929-969.

    Abstract: This article concerns an infinite horizon economy where trade must occur pairwise, using a double auction mechanism, and where fiat money overcomes lack of double coincidence of wants. Traders are anonymous and lack market power. Goods are divisible and perishable, and are consumed at every date. Preferences are defined by utility-stream overtaking. Money is divisible and not subject to inventory constraints. The evolution of individual and economywide money holdings distributions is characterized. There is a welfare-ordered continuum of single price equilibria, reflecting indeterminacy of the price level rather than of relative prices.

2. Shi, S. and Temzelides, T. (2004), "A Model Of Bureaucracy And Corruption". *International Economic Review*, 45: 873-908.

    Abstract: We analyze bureaucracy and corruption in a market with decentralized exchange and "lemons." Exchange is modeled as a sequence of bilateral, random matches. Agents have private information about the quality of goods they produce and can supplement trade with socially inefficient bribes. Bureaucracy is modeled as a group of agents who enjoy centralized production and consumption. Transaction patterns between the bureaucracy and the private sector are fully endogenous. Centralized production and consumption in the bureaucracy give rise to low power incentives for the individual bureaucrats. As a result, private agents might bribe bureaucrats, whereas they do not bribe each other. An equilibrium with corruption and an equilibrium without corruption can coexist. We discuss some welfare implications of the model.

Examples of Type II Error.

1. Mendelberg, T., K. T. McCabe, and A. Thal (2017): "College Socialization and the Economic Views of Affluent Americans," *American Journal of Political Science*, 61, 606–623.

Abstract: Affluent Americans support more conservative economic policies than the nonaffluent, and government responds disproportionately to these views. Yet little is known about the emergence of these consequential views. We develop, test, and find support for a theory of class cultural norms: These preferences are partly traceable to socialization that occurs on predominantly affluent college campuses, especially those with norms of financial gain, and especially among socially embedded students. The economic views of the student's cohort also matter, in part independently of affluence. We use a large panel data set with a high response rate and more rigorous causal inference strategies than previous socialization studies. The affluent campus effect holds with **matching**, among students with limited **school choice**, and in a natural experiment; and it passes placebo tests. College socialization partly explains why affluent Americans support economically conservative policies.

2. Nunley, J., Pugh, A., Romero, N. and Seals, R. (2015): "Racial Discrimination in the Labor Market for Recent College Graduates: Evidence from a Field Experiment". *The B.E. Journal of Economic Analysis & Policy*, 15(3), 1093-1125.

Abstract: We present experimental evidence from a correspondence test of racial discrimination in the **labor market** for recent college graduates. We find strong evidence of differential treatment by race: black applicants receive approximately 14% fewer interview requests than their otherwise identical white counterparts. The racial gap in employment opportunities is larger when comparisons are made between job seekers with credentials that proxy for expected productivity and/or **match** quality. Moreover, the racial discrimination detected is driven by greater discrimination in jobs that require customer interaction. Various tests for the type of discrimination tend to support taste-based discrimination, but we are unable to rule out risk aversion on the part of employers as a possible explanation.

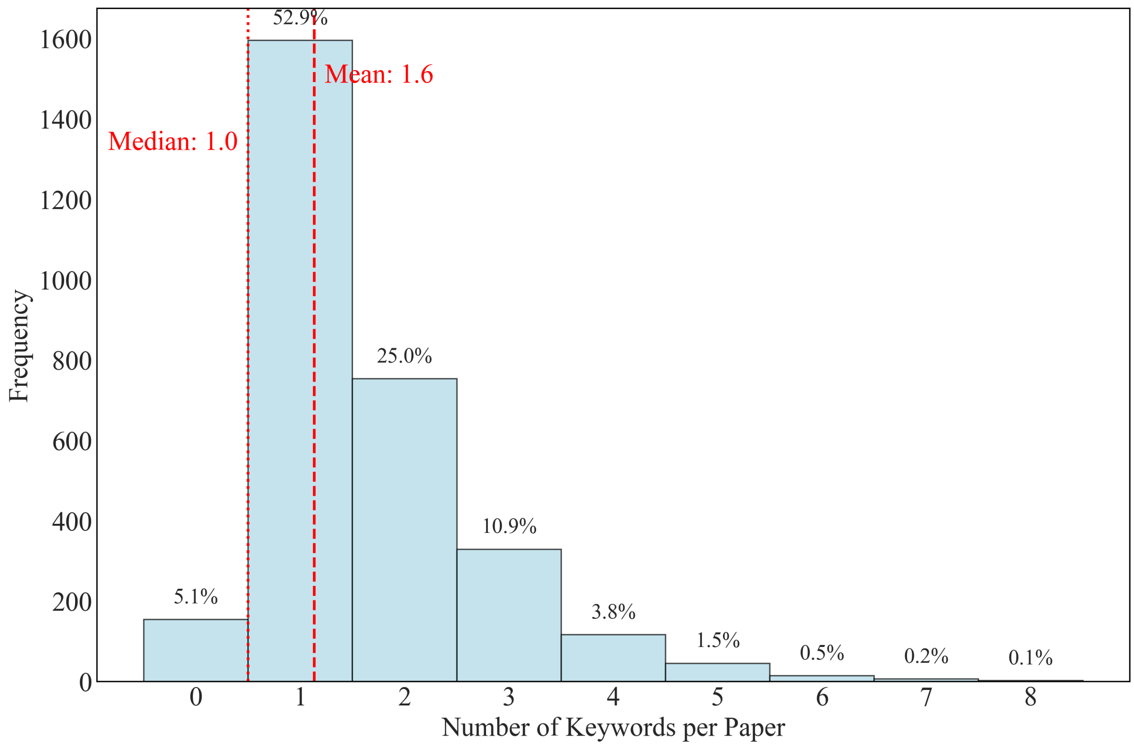

Figure A1: Number of Matching-related Terms per Paper

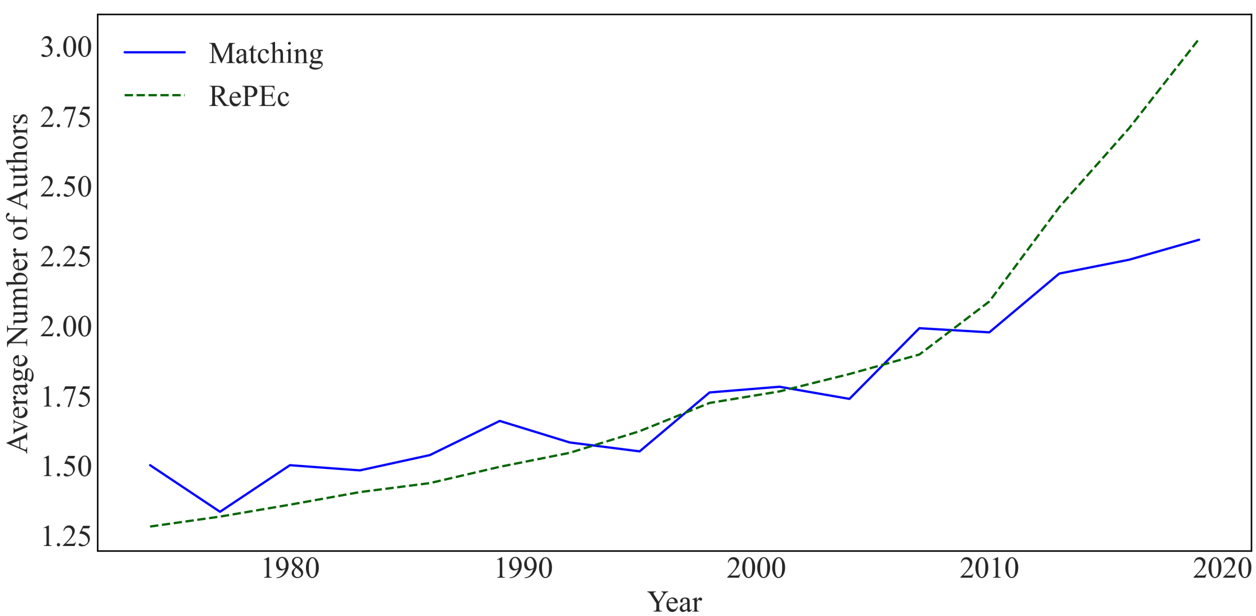

Figure A2: Average Number of Authors per Paper, 1975-2020, by 3-year

Table A1: Size and quality of RePEc dataset (published papers)

	Exactly one "match" no missing ('75–'20)	*Exactly one "match" ('75–'20)*	*At least one "match" ('75–'20)*	*All ('75–'20)*	*All*
Sample size	7,241	20,007	23,357	1,611,122	2,179,019
Missing...					
Title	0	4	4	671	790
	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.0%)
Abstract	0	377	378	343,501	458,910
	(0.0%)	(1.9%)	(1.6%)	(21.3%)	(21.1%)
Keywords	0	12,735	14,319	949,345	1,226,614
	(0.0%)	(63.7%)	(61.3%)	(58.9%)	(56.3%)
Journal	4	15	16	2,772	23,236
	(0.1%)	(0.1%)	(0.1%)	(0.2%)	(1.1%)
No. of pages	298	573	644	119,831	173,298
	(4.1%)	(2.9%)	(2.8%)	(7.4%)	(8.0%)
Year	0	0	0	0	6,312
	(0.0%)	(0.0%)	(0.0%)	(0.0%)	(0.3%)
Author Name(s)	17	59	61	23,169	27,285
	(0.2%)	(0.3%)	(0.3%)	(1.4%)	(1.3%)
Author Workplace(s)	2,193	12,473	14,254	1,026,215	1,290,961
	(30.3%)	(62.3%)	(61.0%)	(63.7%)	(59.2%)
JEL Codes	4,059	16,024	18,401	1,317,746	1,662,646
	(56.1%)	(80.1%)	(78.8%)	(81.8%)	(76.3%)
URL/DOI	27	484	533	64,029	68,608
	(0.4%)	(2.4%)	(2.3%)	(4.0%)	(3.1%)
Has "match" in...					
Title	136	766	3,303		
Abstract	6,741	18,872	22,189		
Keywords	364	369	1,814		
Has "matching" JEL Code	212	353	594	7,719	9,804

Notes. The table reports descriptive statistics of five samples, one per column. The first column corresponds to papers with exactly one instance of "match" in the title, abstract or keywords, and no missing values in any of these variables, the second to papers with exactly one instance of "match", the third to papers with at least one instance of "match." The next to last column corresponds to the sample of papers published between 1975 and 2020, and the last column to all published papers in RePEc. The first row reports the sample size of each sample. The first panel reports the number and percent (in parenthesis) of missing values in each sample of several variables. For the first three samples, the second panel reports the number of papers in which "match" appears in the title, abstract, or keywords. The last row reports the number of papers in each sample that have a "matching" JEL code (see Table A2).

Table A2: JEL Codes Used to Identify Matching Papers

Parent Classification	*JEL Code*	*JEL Code Description*
C7	C78	Bargaining Theory; Matching Theory
D4	D47	Market Design
I1	I11	Analysis of Health Care Markets
I2	I21	Analysis of Education
J1	J12	Marriage; Marital Dissolution; Family Structure; Domestic Abuse
J4	J41	Labor Contracts
	J44	Professional Labor Markets; Occupational Licensing
	J45	Public Sector Labor Markets
J6	J61	Geographic Labor Mobility; Immigrant Workers
	J63	Turnover; Vacancies; Layoffs
	J64	Unemployment: Models, Duration, Incidence, and Job Search
R3	R30	General Regional Economics: General
	R31	Housing Supply and Markets
	R32	Other Spatial Production and Pricing Analysis
	R33	Nonagricultural and Nonresidential Real Estate Markets
	R38	Government Policy
	R39	Other

Notes: Parent classifications are as follows: C7 (Game Theory and Bargaining Theory), D4 (Market Structure, Pricing, and Design), I1 (Health), I2 (Education and Research Institutions), J1 (Demographic Economics), J4 (Particular Labor Markets), J6 (Mobility, Unemployment, Vacancies), and R3 (Real Estate Markets, Spatial Production Analysis, and Firm Location).

Table A3: Terms Used to Identify Matching Papers

1. labor market	32. transferable utility	63. coauthorship
2. search and matching	33. gale and shapley	64. one-sided matching
3. matching model	34. dynamic matching	65. justified envy
4. job match	35. top trading cycle	66. assignment market
5. job search	36. mechanism design	67. online marketplace
6. assortative	37. endogenous matching	68. match between workers and jobs
7. marriage	38. firm-work	69. e-commerce platform
8. sorting	39. frictionless	70. matching of firms and workers
9. skill mismatch	40. college admission	71. theory of matching
10. one to one	41. assignment model	72. student-advisor
11. matching market	42. centralized matching	73. marital matchmaking
12. two-sided market	43. partner matching	74. one-sided market
13. search-model	44. ride sharing	75. doctors and hospitals
14. matching friction	45. student optimal	76. women optimal
15. stable matching	46. kidney exchange	77. men optimal
16. real estate	47. allocation mechanism	78. matching of jobs and workers
17. housing market	48. decentralized matching	79. student placement
18. search friction	49. organ transplant	80. TU-matching
19. perfect matching	50. match value	81. worker proposing
20. matching efficiency	51. bilateral matching	82. match between schools and students
21. many to one	52. mate-matching	83. marital search
22. matching mechanism	53. non-transferable utility	84. extremal matching
23. matching theory	54. matching platform	85. match between a firm and its managers
24. buyers and sellers	55. matching with contracts	86. boston algorithm
25. deferred acceptance	56. exchange program	87. match between jobs and people
26. school choice	57. job-education match	88. child-adoption match
27. worker-firm	58. school admission	89. matching of persons and jobs
28. assignment problem	59. roommate problem	90. early admission
29. market design	60. core allocation	91. object allocation
30. many to many	61. unstable matching	92. matching with transfers
31. strategy-proof	62. marital matching	93. immediate acceptance

Notes. The table reports the 93 matching-related we use to identify matching papers. They are sorted by their frequency in the final selection of papers.

Table A4: Additional Labor Market Matching Terms

1. job center	5. matching function	9. workers and jobs
2. job exit	6. matching in the labor market	10. workers and vacancies
3. job switch	7. mismatch	
4. labor market matching	8. workers and firms	

Notes. The table reports 10 additional terms for papers with “labor market” as the only matching related term detected. These terms are sorted alphabetically, as the robustness check to avoid papers including “labor market” but not focusing on the matching process in this market covered into the matching sample. We found all papers with only “labor market” in our sample passed the check.

Table A5: Terms Used to Exclude Non-Matching Market Papers

1. DNA	9. PSM	17. Mean matching
2. Climate matching	10. Price matching	18. Moment matching
3. Feature matching	11. String matching	19. Matching fund
4. History matching	12. Matching grant	20. Matching test
5. Maturity matching	13. Parameter matching	21. Ontology matching
6. Synthetic matching	14. Boolean matching	22. Statistical matching
7. Sequencing platform	15. Propensity score matching	23. Coarsened exact matching
8. Reconciliation and matching	16. Difference-in-Differences matching	

Notes: This table presents terms used to identify papers that contain the word ”match” but are not related to economic matching markets. Papers containing any of these terms in their title, abstract, or keywords were excluded from our final sample.

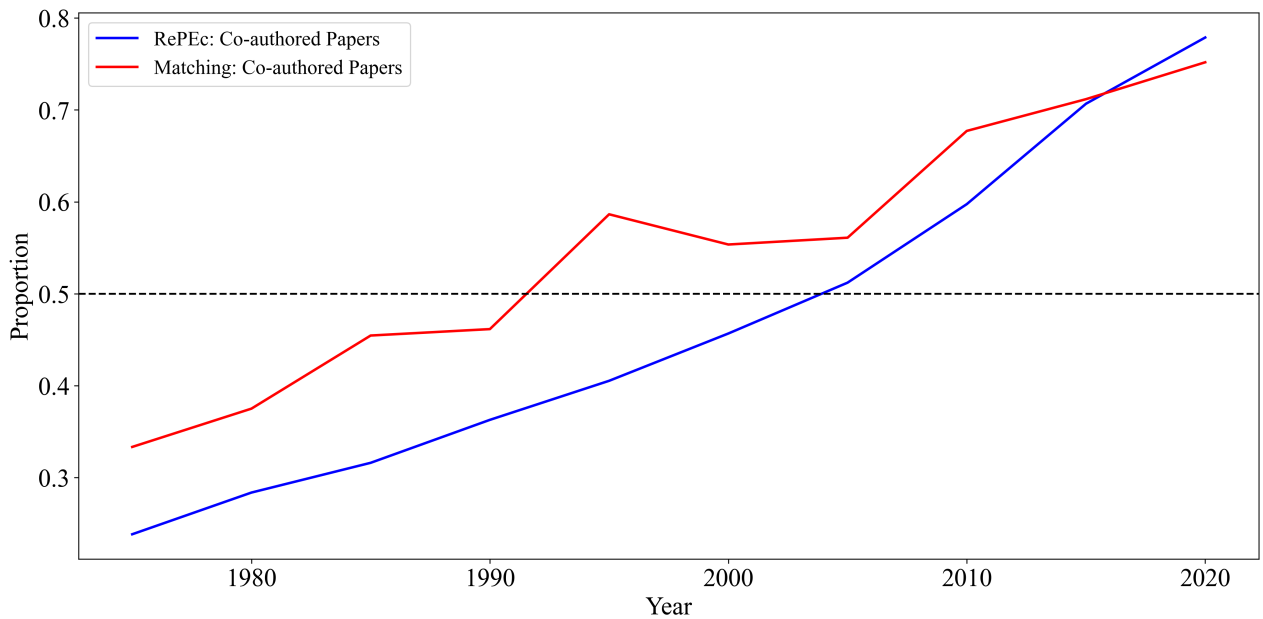

Figure A3: Evolution of Co-authored Paper: RePEc (by 5-year) vs. Matching (by 5-year) Data (1975–2020)

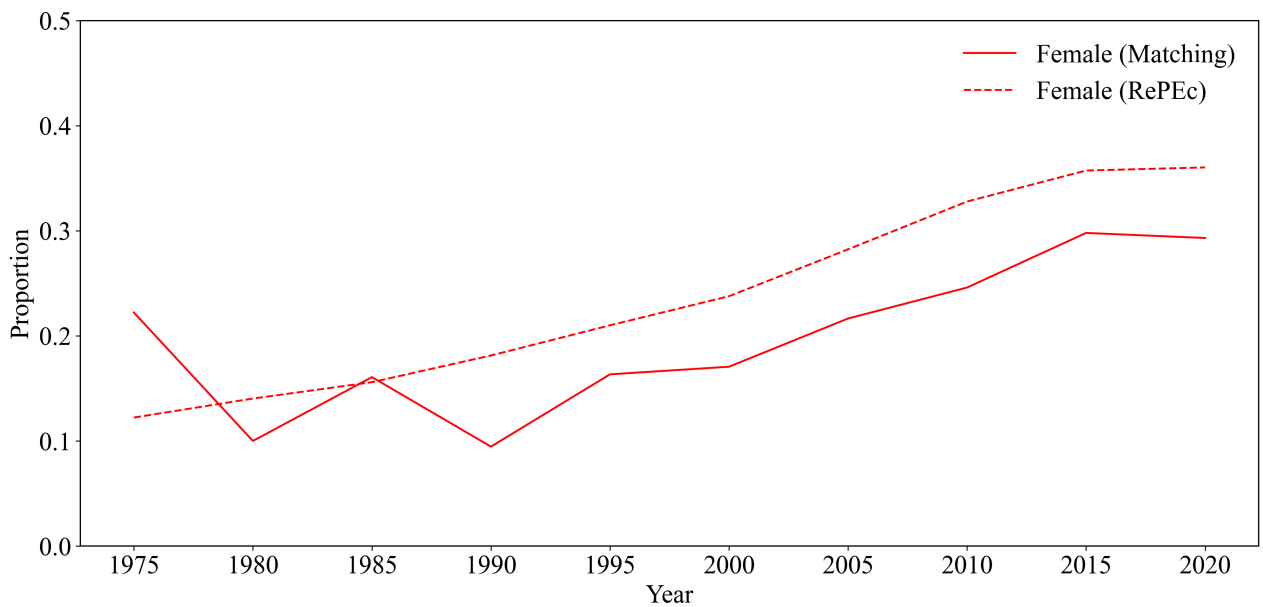

Figure A4: Evolution of Female Economists: RePEc vs. Matching Data (1975–2020), by 5-year

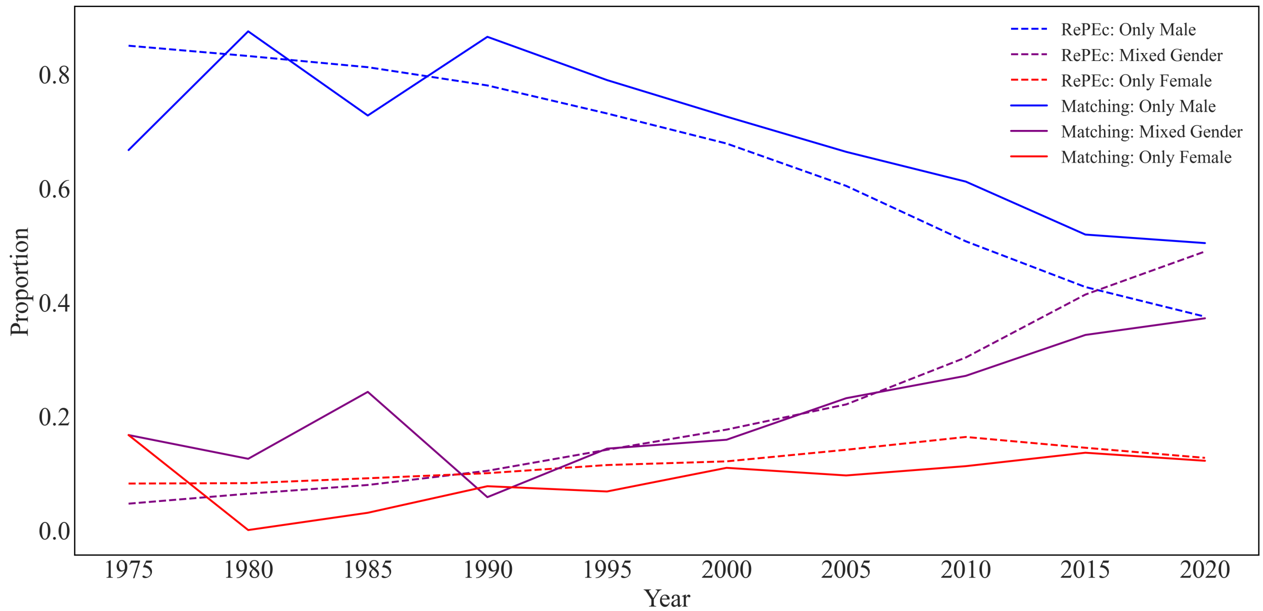

Figure A5: Evolution of Gender Composition in Economics Papers: RePEc vs. Matching Data (1975–2020)

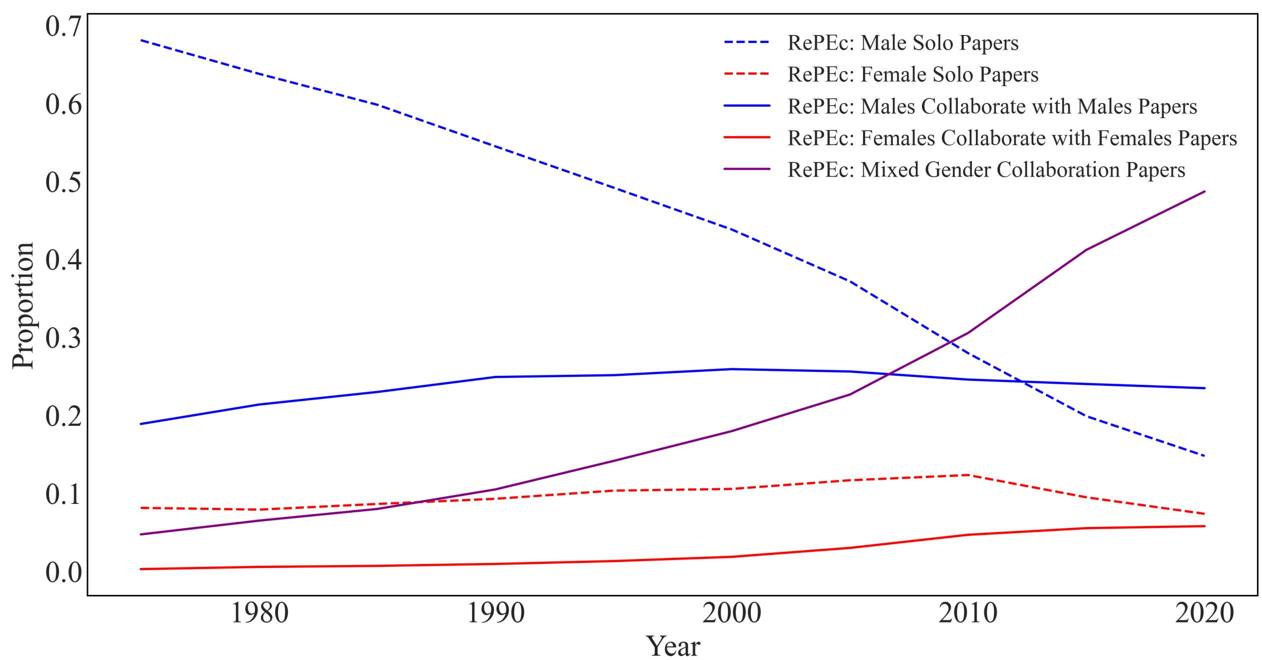

Figure A6: Detailed Gender Composition Trends in RePEc Papers by Collaboration Type, 1975-2020 by 5-year

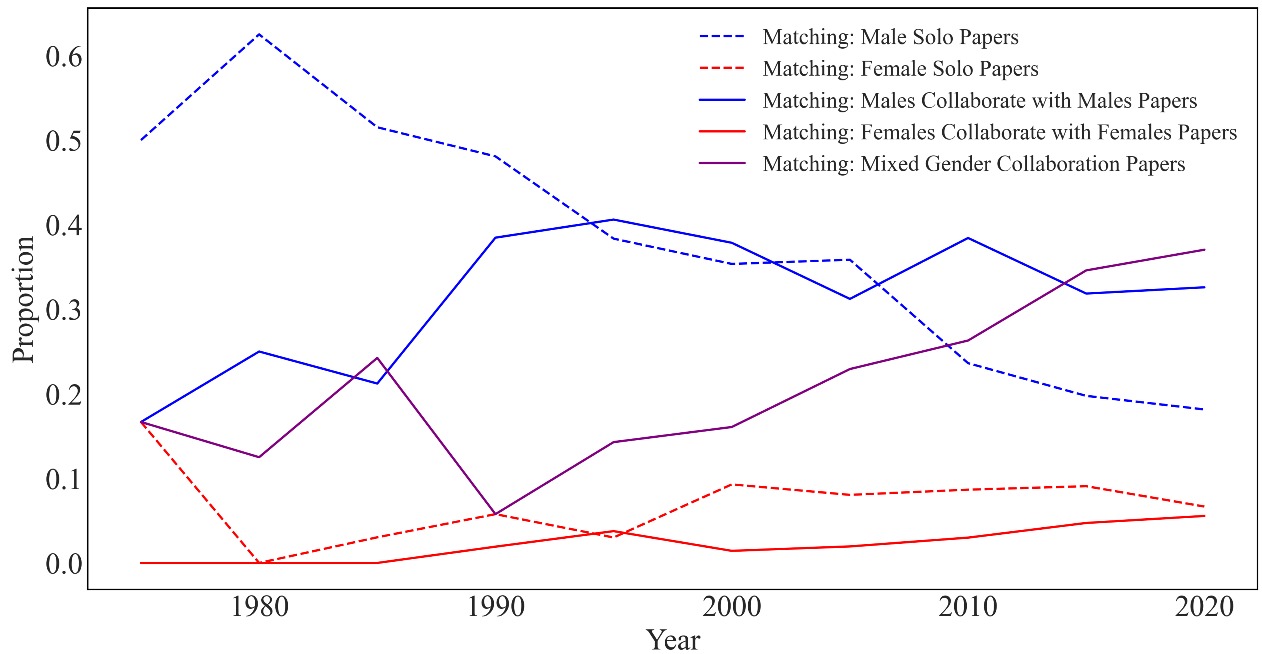

Figure A7: Detailed Gender Composition Trends in Matching Papers by Collaboration Type, 1975-2020 by 5-year

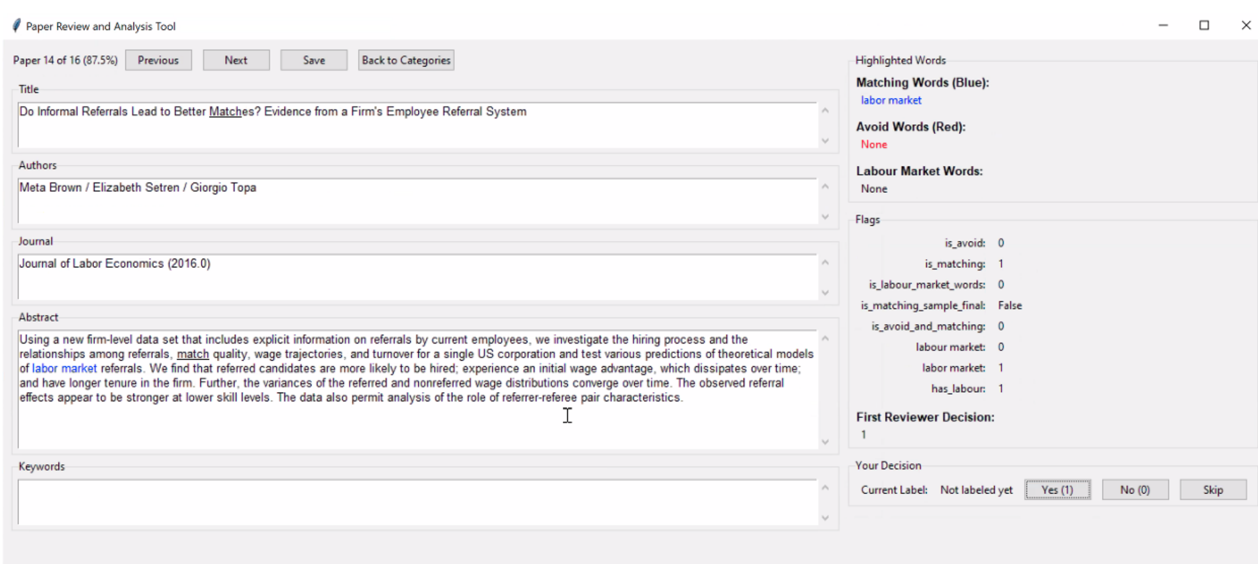

Figure A8: UI example