Matching in Vertical Markets: Evidence from the Digital Advertising Industry

Andrea Chiantello, Francesco Decarolis, Maris Goldmanis, Antonio Penta, and Ksenia Shakhgildyan *

March 30, 2026

Abstract

The digital advertising industry relies on specialized intermediaries to facilitate advertisers' access to online ad space. This paper studies vertical complementarities in this market by estimating a structural model of advertiser–agency matching using novel data and methods for many-to-many matching in large markets. We adopt the standard concept of *pairwise stability* to describe the equilibrium matching between agencies and advertisers. In taking it to the data, we allow agents to not necessarily be aware of the entire market, and have instead different *consideration sets*. We introduce a machine-learning approach to infer agents' consideration sets, and we quantify the value created by matches. The results show that advertisers benefit from being affiliated with the same agency network as their competitors, even though they do not necessarily prefer to share the same agencies. We also document the role of industry specialization, exclusive contracting, and relationship persistence. Finally, we illustrate the policy relevance of our framework by evaluating the potential effects of the Omnicom–IPG merger, showing how network-level consolidation can generate heterogeneous advertiser gains.

JEL: C72, D44, L81.

Keywords: Vertical Markets, Online Advertising, Delegation, Matching, Sponsored Search Auctions.

*Chiantello, Bocconi University, *andrea.chiantello@studbocconi.it*. Decarolis, Bocconi University and IGIER, *francesco.decarolis@unibocconi.it*. Goldmanis, Department of Economics, Royal Holloway, University of London, *Maris.Goldmanis@rhul.ac.uk*. Penta, ICREA-UPF, BSE and TSE Digital Center, *antonio.penta@upf.edu*. Shakhgildyan, Bocconi University and IGIER, *ksenia.shakhgildyan@unibocconi.it*. The BSE acknowledges the financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in RD (CEX2019-000915-S). We gratefully acknowledge financial from the ERC Consolidator Grant - CoDiM -#101002867.

1 Introduction

Digital advertising is a large and rapidly growing industry. In 2024, global digital ad spending was worth a record of 792 billion US dollars.[1] Taking advantage of the opportunities offered by digital advertising requires advertisers to adopt novel technologies and acquire specialized skills. Most firms, and especially those that are not in the tech sector, rarely invest to develop these capabilities in-house, but rely instead on outsourcing these functions to specialized intermediaries. These intermediaries, commonly referred to as Digital Marketing Agencies (DMAs), offer a disparate variety of services to their clients, ranging from managing their bidding strategies on online ad auction platforms to the design and management of their ad campaigns more broadly. Despite the relevance of this vertical relationship between advertisers and intermediaries, still, only a few studies have analyzed its drivers and implications.

Besides the diversity of services DMAs offer to their clients, the complexity of the vertical relationship we aim to understand is further increased by the peculiar industrial relationships which characterize the organization of this market. In particular, while at the level of single agencies this market appears to be relatively fragmented, it is undergoing a significant transformation marked by increasing concentration and consolidation among intermediaries.[2] As a result, many of the DMAs belong to one of the seven major *agency networks*—Omnicom Group, Interpublic Group (IPG), Publicis Group, Dentsu Inc, WPP, Havas, and MDC—which together control a substantial share of advertisers' access to online advertising space.[3] These networks have engaged in a steady stream of mergers and acquisitions (M&As), reshaping the landscape through horizontal and vertical integration.

A salient example is a \$13.5 billion merger between Omnicom and IPG, two of the largest global agency networks, that cleared in September 2025.[4] Both operate extensive portfolios of DMAs and serve hundreds of advertisers across nearly all major U.S. industries. Such a merger significantly reduces the number of agency networks and could raise concerns about coordination, market foreclosure, and loss of competition in the upstream market for digital advertising services. At the same time, it could unlock efficiencies through data sharing, improved algorithmic bidding, and network synergies. Advertisers may benefit from sharing intermediaries with competitors due to multiple reasons such as data pooling and better algorithm performance (McAfee, 2011), budget coordination (Balseiro and Candogan, 2017), keyword splits and branded keyword bidding (Decarolis and Rovigatti, 2021), as well as bid coordination (Decarolis, Goldmanis and Penta (2020); Decarolis et al. (2023)).

These aspects of the market open the door to a rich set of potential patterns for advertisers' preferences, who might in principle display both complementarities or substitutabilities over the matchings between other

[1] Statista.

[2] According to the US Census Bureau's 2016 annual report, there are 12,000 marketing agencies in the US, with 8,000 of them having less than four employees and 10,800 less than 20. The advertising agencies with more than 100 employees are 285 and account for 60% of the employment in the sector.

[3] Six after the late-2025 acquisition of IPG by Omnicom, which falls outside our sample period but further illustrates the ongoing consolidation trend.

[4] The Federal Trade Commission took action to resolve antitrust concerns related to Omnicom Group Inc.'s \$13.5 billion acquisition of The Interpublic Group of Companies, Inc. (IPG).

advertisers and marketing agencies. An advertiser, for instance, might evaluate either positively or negatively the possibility of hiring an intermediary who also manages the ad campaigns of its rival advertisers: Although sharing a marketing agency might create conflicts – as suggested, for instance, by both theoretical (Villas-Boas, 1994) and empirical (Silk and King III, 2013) studies on the advertising industry before the digital revolution – the benefits in the realm of digital advertising can be disparate, including access to better data and the possibility of bid coordination in the auctions where online ad space is sold.[5]

Our empirical analysis employs a large dataset that links, among the top 6,000 US advertisers, every firm active in the digital advertising market to one (or more) specific advertising agency for all major US industries between 2014 and 2020. Deals between firms and intermediaries are modeled as the result of a (many-to-many) matching game. Our framework allows us to identify the forces driving advertisers' choices of intermediaries, and quantify how market concentration—both at the agency and network levels—affects the value created by matches. In particular, we estimate how advertisers value whether their competitors are also clients of the same agency or network, past relationships, and being an exclusive client.

Although matching games are commonly used for modeling specific applications such as marriage, school applications, kidney exchanges, and employment choices, they are relatively less used for vertical markets where matching between firms takes place along the supply chain. This is likely due to three challenges: first, the lack of systematic data covering the vertical matches in the industry; second, the inherent complexity associated with analyzing matching problems in settings, such as the present one, in which parties are not constrained to a single match (hence the many-to-many matching framework); finally, and related to the second point, in large markets it is unreasonable to assume that agents are fully aware of all existing agents and possible matching opportunities, as they are realistically restricted by their *consideration sets*. Such consideration sets, however, are unobserved, and need to be inferred from the data jointly with the other equilibrium objects.

In this work, we overcome all three difficulties. First, we rely on a novel firm-level dataset we specifically developed for this study, by combining several data sources on the digital ad industry. Second, we exploit the maximum score estimator proposed by Fox (2018), which is ideal to deal with large markets such as the one we consider. Third, to address the problem that the whole market under consideration is too large as a matching market, in the empirical analysis we estimate agents' consideration sets as part of the model. As discussed below, this estimation approach exploits the features of the stability notion (Jackson and Wolinsky, 1996) used to solve the game to avoid the severe curse of dimensionality which would plague other approaches, if applied to large markets. Crucially, the method does not require observing monetary transfers between the advertisers and the agencies, which are rarely observed in the data. Finally, in order to overcome the key challenge posed by the unobservability of the consideration sets, as well as to ease the computational complexity, we use the Support Vector Machine (SVM) method to predict the consideration

[5] As shown by Decarolis, Goldmanis and Penta (2020), in the main existing auction formats, bid shading (and, hence, price reduction) is in fact feasible for an intermediary which handles ad purchases on behalf of multiple advertisers that are interested in the same type of ad space.

sets. In particular, we make use of the network structure to infer possible links between agencies and advertisers. This estimation procedure also makes the problem computationally tractable, and allows us to structurally estimate the preference parameters for advertisers and intermediaries which determine the matching between them, as well as to evaluate counterfactual changes in the market structure.

Our results show that competing advertisers value being affiliated with the same agency network as their rivals, even though they do not necessarily prefer to share the same marketing agency. This pattern can be understood by considering the internal organization of digital marketing intermediaries. While creative activities and strategic decisions related to campaign design are typically handled at the *agency level*, key scale-intensive functions—most notably algorithmic bidding and data analysis—are centralized at the *network level* within specialized units known as agency trading desks (ATDs; see Section 3). As a consequence, advertisers benefit from network-level spillovers arising from data pooling and shared bidding technologies, which can improve targeting efficiency and bidding performance even in the presence of competition among advertisers. Consistent with this interpretation, our estimates indicate that advertisers place positive value on these network-level complementarities, in line with the mechanisms highlighted in Decarolis, Goldmanis and Penta (2020) and Decarolis et al. (2023).

Beyond this central finding, the estimates also provide a quantitative assessment of several additional forces shaping matching patterns in the digital advertising market, including the role of industry specialization among intermediaries, the value (or lack thereof) of exclusive relationships, advertisers' diversification across agencies, and the persistence of past relationships. Together, these results offer a comprehensive picture of how organizational structure and network effects interact to determine value creation in vertically related digital markets.

We conclude by assessing the potential effects of the Omnicom–IPG merger, using our estimated model to evaluate how the consolidation would affect advertisers' and agencies' valuations under the existing matching structure. The analysis highlights that mergers between agency networks can generate meaningful gains for advertisers through network-level complementarities—such as enhanced data sharing and improved algorithmic bidding—while also revealing substantial heterogeneity in these effects across networks. In particular, the results indicate that the magnitude and distribution of advertiser gains depend critically on the internal organization and client composition of the merging networks, rather than being uniform across holding companies. More broadly, this exercise illustrates how our framework provides a practical tool for merger evaluation in vertical markets, even in settings where monetary transfers are unobserved. By explicitly accounting for the many-to-many nature of matching and the layered organizational structure of intermediaries, our approach offers a new quantitative lens through which to analyze consolidation and competitive dynamics in digital markets.

The rest of the paper is organized as follows: Section 2 reviews the relevant literature concerning vertical markets, marketing agencies, and matching games. Section 3 describes the institutions and data. Section 4 introduces the theoretical framework. Section 5 discusses identification. Section 6 presents the estimation approach and the results. Section 7 presents the results of the Omnicom-IPG merger evaluation. Section 8 concludes.

2 Related Literature

This work contributes primarily to three branches of the literature. First and most directly, it studies the behavior of advertisers in digital advertising. The economics literature on the digital advertising market is becoming more mature, but it is still primarily focused on the publishers' perspective and on what is the optimal mechanism to sell ad space. This is perhaps the reason why this literature has focused more on search than on display ads. As discussed below, the former entails ad space sold on the result page of the search engines, while the latter regards the ad space sold on any other website (from those of established newspapers to low-traffic web pages, all collectively indicated as publishers). The studies on the search ad auctions are closely related to the classical auction theory and mechanism design literature. They usually analyze how allocative efficiency and publishers' revenues change under different mechanism designs or model assumptions. Some papers treat the auctions for ads as classical multi-item auctions (Edelmann, Ostrovsky and Schwarz, 2007; Varian, 2007; Gomes and Sweeney, 2014; Shakhgildyan, 2025), while other works aim to account for peculiar aspects of the search auctions such as the fact that the bids are not updated every time an auction is run (Athey and Nekipelov, 2010) or that ranking positions yield heterogeneous clicks amounts for different keywords (Goldman and Rao, 2014). Another stream of literature takes the perspective of the advertiser and studies how the competition for ad slots affects their revenues (Agarwal and Mukhopadhyay, 2016) or what is the best strategy to maximize revenues while choosing the keyword portfolio under budget constraints (Baardman et al., 2019). In turn, Decarolis, Goldmanis and Penta (2020) and Decarolis et al. (2023) analyze how bid delegation to a common marketing agency undermines both auctioneer revenues and efficiency of the generalized second price auction.

If display advertising is somewhat less studied in the economics literature, it is more popular in the engineering and data science ones. An exhaustive literature review is conducted by Choi et al. (2017). The economics literature on the subject focuses again on the optimal mechanism design to allocate ads by exchanges (McAfee, 2011). Some works enquire about the relation between advertisers and other actors by theoretically investigating the optimal types of contracts (Balseiro and Candogan, 2017), but without studying how the players are chosen.

The second branch of literature is the one on vertical markets with intermediaries. Although middlemen are obviously fundamental in many markets, their presence and influence are often understudied. The most important actions carried out by economic intermediaries are, among others: setting prices and clearing markets, providing liquidity and immediacy, coordinating buyers and sellers, guaranteeing quality, and monitoring performance (Spulber, 1996). Recent works in the empirical industrial organization try instead to study which role intermediaries have, how they benefit the market, and try to quantify such effects. Since vertical markets often witness upstream actors interacting with many downstream counterparts and vice versa, network models are often employed (Condorelli, Galeotti and Renou, 2016) to capture the nature of the relation. In many cases, the whole network is described according to various measures (such as connectivity, density, or nestedness) to shed light on market functioning. There are numerous examples, from the macro level, on how value chains of countries predict growth (Hausmann and Hidalgo, 2011), to micro, on how the "nestedness" measure of a value chain in retail can predict failure of firms (Uzzi, 1996). Works of this kind are close to the network science literature and focus on networks as a whole. The current work, on the other hand, explicitly models the economic agents' profit functions, which paired with an equilibrium concept leads to a network formation game (or matching game).[6]

The third and final strand of related literature is on the estimation of matching games. To our knowledge, Day (2014) was the first to propose matching estimation as a way to analyze the relationship between advertising agencies and their clients, in her case in the context of traditional advertising for pharmaceutical products. Since on both ends of the market, there are firms making choices, matching models offer a reasonable modeling framework to estimate agents' preferences over the characteristics of their potential partners.[7] The revealed preference approach for the study of matching games encompasses several different estimation methods deriving from the discrete choice model literature. It is not a surprise, then, that the maximum likelihood and method of moments estimators, extensively used in the demand estimation literature (McFadden et al., 1973; Train and Weeks, 2005; Berry, Levinsohn and Pakes, 1995), are also the most common choice in 'large' matching games (Choo and Siow, 2006; Abdulkadiroğlu, Agarwal and Pathak, 2017). These two classes of estimators usually assume the idiosyncratic error term distribution to be known, while they let the deterministic part of the utility function be flexible.[8] The maximum score estimator we adopt takes the opposite direction: it specifies the deterministic part of the surplus function, and lets the random term be unrestricted. The maximum score estimator used in this works, which was developed by Fox (2018), builds on the seminal work of Manski (1975). It has been applied to many markets involving one-to-many matchings (Chen and Song, 2013; Fox and Bajari, 2013), and to a few which involve many-to-many matchings (Nosal, 2016; Fox, 2018) or even to entry dynamics (Ellickson, Houghton and Timmins, 2013). A fourth estimator is the inequality moment estimator (Pakes et al., 2015) which is similar to the one used here since they both rely on inequality conditions but differs substantially due to the form taken by the profit function.

The choice of a maximum score estimator in this study is motivated by two reasons. First, the unrestricted distribution of the error term eases the computational problem, which is relevant in large, many-to-many

[6]Note that we do not model intermediaries as both sellers and buyers. We focus on the downstream section that sees only the advertiser and agencies. Multistage matching models that treat agents as both sellers and buyers are possible in theory but are substantially more complicated in practice (Azevedo and Hatfield, 2018). Translated to the digital advertising industry this means that deals between agencies and other intermediaries along the value chain such as demand-side platforms (DSPs), Ad exchanges, or Search engines are not directly addressed.

[7]Although an important strand of the empirical matching literature focuses on non-transferable utility models – see Agarwal (2017) for a recent overview, for applications related to the vertical links between firms a transferable utility framework is evidently the right approach. We thus follow Fox (2018), as discussed next.

[8]Most common choice is GEV type I, for the same reason provided by demand estimation literature.

matching markets such as this one. Importantly, many-to-many matching is the core of the work since its objective is to measure complementarities. Second, the advertisers and the agencies bargain for payment in exchange for service after choosing each other, but the data on payments are not available. Our model presumes an environment with Transferable Utility (TU), in which transfers are known to the agents but unobserved by the econometrician. Some alternative estimators circumvent the second problem by considering a joint surplus function but do not provide any solution to the first one (Chiappori and Salanié, 2016).

Many papers study the vertical relationship through a Nash bargaining model that exploits data on wholesale prices (Donna et al., 2018), intermediaries' marginal costs (Berto Villas-Boas, 2007) or both (Ho and Lee, 2019). Papers that study how intermediaries affect the market and the corresponding network formation cover different industries, including healthcare (Ho and Lee, 2019), food and beverage (Berto Villas-Boas, 2007), physical advertising (Donna et al., 2018), etc. The Nash bargaining model is a popular choice because, if such data are available, they can be used to estimate bargaining power parameters across the vertical relation and more importantly to split the generated surplus. Nevertheless, without data on wholesale prices or costs, it is to our knowledge not possible to split the surplus generated by the intermediaries between them and their clients, thereby making the Nash bargaining model less useful in our context. As a matter of fact, difficulties in analyzing vertical markets often arise from the scarcity of these data, which are enterprises' private information.

3 Institutions and Data

3.1 Digital Advertising

The market of digital advertising is growing rapidly. In 2019, according to industry reports by eMarketer, expenses on online ads surpassed for the first time those on print and television combined. Digital advertising can be divided into four main categories: sponsored search, display (or banners), video, and social media. Among all advertising formats, sponsored search auctions are the most prominent segment of digital advertising with a 40% revenue share.[9] This is the segment on which we focus in this project.

For Search advertising, things are somewhat simpler than they are for display ads. Search ads show up on the results page of a search engine such as Google, Yahoo!, or Bing after a user submits a query. Thus, the only seller for such ad spaces is the search engine itself, which literally runs an auction to allocate the space to advertisers every time a consumer queries a keyword.[10] The main source of complexity is the number of keywords for which any advertiser is allowed to bid. Considering that 63,000 searches every second are submitted to Google, potential advertisers need intermediaries, in the form of DMAs, to manage their buying process effectively.[11]

[9] Search advertising revenues.

[10] Social Media advertisement works in a similar fashion, with the social media platform being the only supplier of space and allocating it through auctions to the publishers, with the only difference being that ads are shown on the personal feed and not among query results.

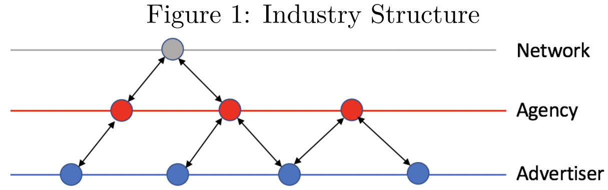
[11] Sometimes, however, the DSP service is provided by the search engine itself, in a process of disintermediation. Google

The reason behind such a heavy intermediation can be found in three characteristics regarding ads and their allocation: (i) the speed at which the auctions must be accomplished; (ii) the minuscule value and high volume of the items that are traded; and as a result, (iii) the need to use automated systems for bidding, evaluation, and execution of the trades (McAfee, 2011). Programmatic (automated, or algorithmic) bidding is also performed because the items on sale are difficult to evaluate. In fact, the value of a click on a given ad varies extensively according to who clicks on it. Advertisers, therefore, collect data on potential customers through proprietary cookies, the IP address of visitors, or simply by acquiring the data from third parties. Once all this information on the target audience is acquired, it is used to optimize the bidding process, selecting the right timing and location of the ads to show. In order to perform real-time bidding, automation is necessary.

3.2 Digital Marketing Agencies

Marketing agencies are the core of the advertising and marketing sector. They traditionally provide multiple services including account management, production of advertising material, media planning, buying (i.e., placing ads), as well as creative services. According to the US Census Bureau's 2022 annual report, there are 13,682 marketing agencies in the US, with 9,620 of them having less than four employees and 12,411 less than 20. The advertising agencies with more than 100 employees are 294 and account for 60% of the employment in the sector. While this data might suggest a rather fragmented industry, the reality is more nuanced as most agencies belong to one out of just seven large agency networks.[12] These are holdings, sometimes referred to as "DMA Networks": WPP, Publicis Group, Dentsu Inc, Interpublic, Omnicom Group, Havas, and MDC. DMA Networks' revenues range from less than 2 billion US dollars for MDC, up to more than 15 billion dollars for WPP, the market leader. Figure 1 presents a schematic representation of the industry structure.

Figure 1: Industry Structure



These parent networks own multiple agencies which act as independent companies, offering a wide range of services but remaining similar in their nature. While the very concept of a network of agencies was first introduced by Interpublic Group as a way of circumventing the industry's non-spoken rule that forbids a DMA to serve direct competitors, according to Silk and King III (2013) the main advantages of a common affiliation are support programs, technologies, and access to capital. Indeed, it is often the case that agencies

Ads (previously AdWords) is technically a DSP for ads on Google's result pages, while Bing Ads is the competitor's equivalent serving both Bing (Microsoft) and Yahoo! Search (Verizon).

[12] See 2022 SUSB Annual Data Tables by Establishment Industry.

belonging to the same network even compete for clients, and it does seem to be crucial for agencies to retain as much independence as they can, in order to appear credible to their clients.

DMA dimensions vary widely with the type of service they offer, and the industry is extremely dynamic. Similarly, DMAs clients range from big corporations to very small family-owned enterprises, even if for the former there is a strong case for managing marketing in-house because of the cost. Importantly, it is not the case that DMAs specialize in serving clients in some specific industry. In contrast, one DMA often serves advertisers from a broad set of industries. The billing methods used by DMAs are multiple, ranging from fixed fees per ad campaign (the most widely used method), to hourly rates, to a mixture of both. Performance-based billing is not the standard in the industry, and for this reason, firms are trying to internalize the service offered by the DMAs, which are sometimes considered to be rather opaque. Another trend occurring in this industry is project-based billing, a phenomenon that could explain why advertisers deal with so many different DMAs during a single year.

3.3 Data

The dataset used in this study extends to 2020 the dataset used by Decarolis and Rovigatti (2021) in their study of how intermediary concentration affects Google's revenues from search auctions. The dataset combines two sources: a dataset linking advertisers to DMAs (and DMAs to their agency networks) and a second one linking advertisers to sponsored keywords.

Redbooks is a comprehensive database on marketing agencies that links a set of advertisers—representing the largest U.S. corporations active in online marketing—to the full list of digital marketing agencies (DMAs) with which they work. In addition, it identifies whether each agency is independent or affiliated with one of the seven major agency networks, and reports the type of services provided, including whether the agency is active in sponsored search. To focus on the segment of the industry most closely tied to auction-based advertising, we restrict attention to DMAs flagged as paid-search agencies, yielding between roughly 100 and 150 agencies per year and between 248 and 482 associated advertisers.

Tables 1 and 2 provide a first overview of the resulting network structure. Table 1 reports summary statistics across different subsets of agencies (all agencies, digital agencies, and paid-search agencies), highlighting how the paid-search segment differs from the broader marketing industry. Although the overall industry appears highly fragmented, with thousands of agencies and a large mass of independents, the paid-search segment is markedly more concentrated at the network level. In particular, while individual paid-search agencies remain relatively small, advertisers tend to cluster within a limited number of large networks, with each network serving on average more than 75 advertisers.

Table 2 focuses on paid-search agencies and illustrates the many-to-many nature of the matching environment. On average, each advertiser works with more than one agency, and each agency serves multiple advertisers, with substantial dispersion in both dimensions. At the same time, advertisers typically affiliate with only one network, suggesting that networks represent the relevant organizational layer for coordination

Table 1: Summary Statistics by Agency Type

	All	Digital	PaidSearch
Number of Networks	7	7	7
Number of Agencies	3780	1146	223
Number of Independent Agencies	2698	894	179
Number of Advertisers	2717	1668	815
Avr. Number of Advertisers per Agency	2.67	3.50	5.41
Avr. Number of Agencies per Advertiser	3.72	2.40	1.48
Avr. Number of Agencies per Network	161.43	38.43	7.14
Avr. Number of Advertisers per Network	336	183.57	75.71
Avr. Number of Networks per Advertiser	1.86	1.57	1.23

Notes: The unique matches are counted across years, therefore each match is only counted once if it occurs in multiple years. The columns report the different subsets of marketing agencies on which the dataset is filtered for the production of the statistics: All, Digital, and Paid Search. Where the column "All" uses no filters for agency type.

and data aggregation, even when advertisers split their activities across multiple agencies. This combination of sparse bilateral relationships and dense network-level connections motivates our modeling approach, which allows for many-to-many matching and explicitly distinguishes between agency-level and network-level interactions.

Together, Tables 1 and 2 underscore that the digital advertising market cannot be accurately described as a collection of isolated bilateral relationships. Instead, it is best understood as a networked market in which advertisers interact with multiple intermediaries embedded within larger organizational structures. These features play a central role in shaping competitive interactions and are a key input into the matching model developed in the next section.

Table 2: Summary Statistics for Paid Search Agencies

	Mean	Median	Min	Max	Std
Number of Advertisers per Agency	5.41	3.00	1.00	42.00	7.56
Number of Agencies per Advertiser	1.48	1.00	1.00	10.00	1.01
Number of Agencies per Network	7.14	6.00	2.00	14.00	4.18
Number of Advertisers per Network	75.71	72.00	14.00	154.00	47.73
Number of Networks per Advertiser	1.23	1.00	1.00	5.00	0.56

Notes: The unique matches are counted across years, therefore each match is only counted once if it occurs in multiple years.

Table 3 reports the number of advertisers and DMAs active in each year between 2014 and 2020, disaggregated by advertisers' industry as classified by Redbooks. The table highlights substantial heterogeneity both across industries and over time. Some industries—such as Financial Services, Media, Technology, and Automotive—consistently host a large number of advertisers and agencies, reflecting the intensive use of sponsored search advertising in these sectors. Other industries are much smaller and more stable, with only a handful of advertisers and DMAs active in each year.

In interpreting these figures, it is important to note that advertisers are uniquely assigned to a single industry, whereas DMAs can—and often do—operate across multiple industries. As a result, the same agency may appear multiple times in Table 3, serving different sets of advertisers in different industries within the same year. This cross-industry presence of agencies implies that industry classifications cannot be used to restrict the set of potential advertiser–agency matches. Instead, industries primarily serve as a useful lens to describe market composition and heterogeneity, while the relevant competitive interactions among advertisers are captured at the level of advertising markets constructed from keyword clusters in the next section.

Table 3: Number of Advertisers and DMAs Active Each Year in Each Industry

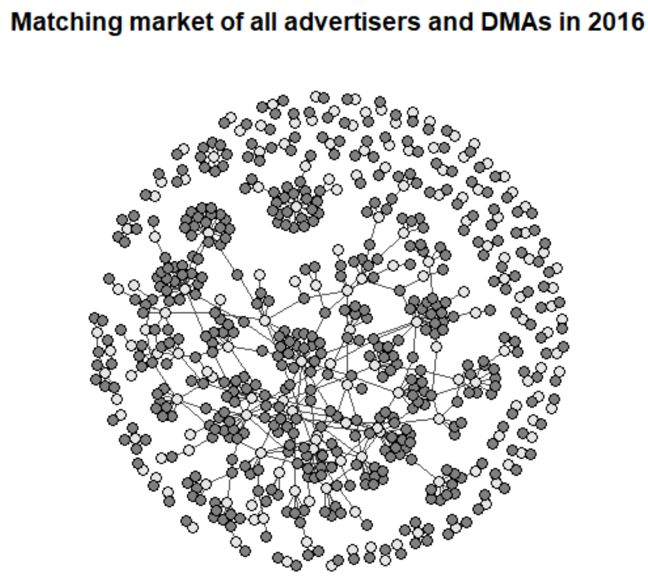
Year	2014		2015		2016		2017		2018		2019		2020	
Industries	Adv	DMA	Adv	DMA	Adv	DMA	Adv	DMA	Adv	DMA	Adv	DMA	Adv	DMA
Accessories, Cosmetics & Jewelr	7	7	6	6	8	9	7	8	7	10	7	10	7	7
Apparel Wear & Shoes	17	20	14	15	20	18	18	19	12	12	11	12	9	10
Associations, Institutions, Unio	8	7	11	12	12	13	10	9	4	4	5	5	4	4
Automotive & Transport	19	20	23	28	19	24	14	20	21	23	19	24	17	18
Education	5	5	6	7	8	9	7	8	3	3	5	5	2	2
Financial Services	46	33	51	40	61	45	20	21	32	22	30	22	25	22
Food & Beverage	7	11	8	7	11	15	9	10	8	16	11	18	6	6
Food Processing & Agriculture	20	22	17	21	28	29	23	24	20	24	23	23	18	15
Government	5	5	6	6	8	8	8	8	6	8	7	8	4	5
Groceries & Food Retailers	2	3	5	6	4	5	2	3	2	2	4	4	4	4
Hardware & Construction	15	18	19	22	22	23	9	11	6	9	6	8	4	7
Housewares & Appliances	10	13	13	13	13	12	8	8	6	8	7	10	6	7
Industrial	27	25	28	25	28	27	12	16	15	15	16	20	14	14
Media	51	44	63	54	66	56	48	45	34	33	36	34	36	30
Pharmaceutical & Health	14	16	21	23	16	19	10	11	14	17	16	17	19	20
Recreation	22	22	21	22	33	29	23	25	16	17	21	22	16	14
Restaurants	19	20	23	21	25	24	16	15	6	11	9	15	12	17
Retail	10	12	11	14	11	14	9	12	9	13	9	13	9	9
Technology	34	31	39	36	39	34	29	28	29	27	27	25	17	15
Telecom	9	13	9	14	9	16	3	7	5	8	5	7	3	4
Travel, Leisure, Real Estate	36	31	34	28	37	31	18	18	14	16	15	16	12	13
Utilities	3	3	3	3	4	4			2	2	3	3	4	3
All Industries	386	147	431	158	482	166	303	138	271	107	292	111	248	96

Notes: Adv column reports the number of advertisers clients of DMAs in a given market and year. DMA column reports the number of DMAs. The number of DMAs is almost always higher than advertisers.

The network structure in Figure 2 helps visualize the data. Figure 2 pools together all the matchings observed across all industries in 2016. This graph indicates the presence of two types of DMAs: on the edges of the graph, we notice a universe of independent DMAs dealing with one or a few clients, while, at the center, the biggest component of the network includes more interconnected DMAs dealing with multiple clients. There are plenty of advertisers who match to more than one DMA and plenty of DMAs matching with more than one advertiser.

Observing that advertisers use multiple DMAs is not surprising. In fact, the advertisers in the Redbooks data are large corporations, needing different types of marketing services and are active across multiple markets. DMAs, in turn, are specialized both by the types of services they offer and by the markets in which they have expertise. Indeed, an industry typically contains multiple separate markets. Moreover, it is not uncommon that advertisers deal with DMAs for a single marketing campaign, like the launch of a specific

Figure 2: Connections between all advertisers and DMAs active in 2016



Notes: advertiser in grey, DMAs in white. The figure is a bipartite graph with advertisers only connected to DMAs and vice versa. A link between nodes represents a match.

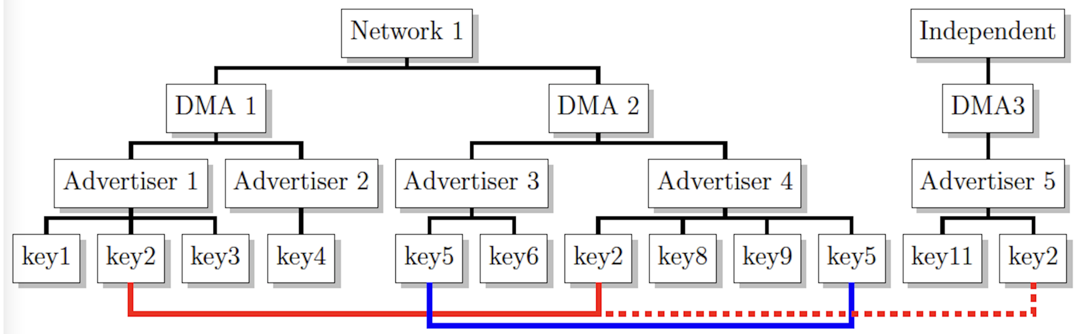
product and therefore deal with many DMAs throughout the same year.

Next, we combine the Redbooks data with SEMrush data covering Google's sponsored search auctions in the U.S. over the same period. SEMrush is one of the main providers of online advertising data and reports, at the keyword level, the set of advertisers winning ad slots, search volumes, and average prices (cost-per-click). Merging these two sources allows us to observe, for each advertiser, both the organizational structure of its intermediation relationships—agencies and agency networks—from Redbooks, and the set of advertising markets in which it is active—from SEMrush.

Figure 3 illustrates the resulting data structure. Advertisers form the central layer of the dataset, linking upward to agencies and agency networks, and downward to the keywords on which they bid. These connections give rise to two distinct but complementary sources of interaction. On the one hand, advertisers are connected to one another through shared agencies and shared agency networks, capturing organizational and technological spillovers in intermediation. On the other hand, advertisers are linked through overlapping keyword participation, which defines the competitive environment in the auctions for ad space. The figure highlights how these two dimensions jointly determine the relevant competitive and organizational structure of the market.

This merged data structure is crucial for our analysis. It allows us to construct measures of competitive overlap across advertisers based on keyword-level activity, while simultaneously accounting for the network of intermediaries through which advertisers access auction markets. In particular, it enables us to distinguish between competition occurring within agencies and competition mediated at the network level, a distinction that plays a central role in the matching model and in the interpretation of our empirical results.

Figure 3: Data Structure



Notes: Data structure: keywords (SEMrush), advertisers (Redbooks/SEMrush), agencies and networks (Redbooks). Solid lines represent examples of coalitions: within DMA (blue) and network (red).

3.4 Keyword clusters

Decarolis and Rovigatti (2021) propose a methodology to cluster together keywords so that they resemble a market: each cluster contains keywords that are related from the point of view of consumers and for which advertisers are in competition for the scarce ad space.[13] These clusters represent the markets where the competition takes place for the ad slots. In this work we employ similar methodology in order to define in which markets each advertiser operates.[14] Technically, the clustering algorithm assigns keywords to thematic clusters based on their semantic content. For example, we expect keywords such as "camera lens", "professional camera" and "photography kit" to belong to the same thematic cluster.[15]

Two caveats are important. First, we shall stress that we are considering the competition for ad spaces and not between the products. Hence, as far as advertisers are in competition for the limited ad space, the fact that both substitutes and complements are likely to be pooled together is not a problem. That is, even firms that sell rather different products might be competing to grab the attention of the same consumer as he browses the web. Second, albeit these markets are constructed using Google's sponsored search auctions alone, the paramount importance of this form of online advertising ensures that we can use these markets to approximate more generally the degree of overlap between advertisers. Indeed, it is reasonable to assume that advertisers competing for search ad space will also be competing with display ad and social media ad

[13]The authors employ the so-called GloVe (Global Vectors for Word Representation) algorithm, an unsupervised learning algorithm for learning vector representation of words based on their statistical properties (see Pennington, Socher and Manning (2014) for further details).

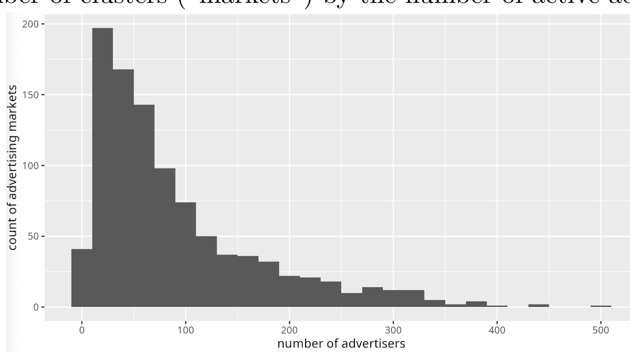
[14]However, we only consider the 1st layer not to make the markets too narrow.

[15]In particular, we apply Spherical K-means clustering to the keyword vectors. The number of clusters is set to 1,000 to strike a balance between computational efficiency and clustering granularity. Whenever a keyword is made up of more than a single word, its embedding representation is defined as the sum of the embeddings of the tokens making it up.

spaces since the target audience and the type of ad will be similar. Video and audio ads might, instead, represent a somewhat distinct market, but we have no data to address them separately.

Figure 4 shows the distribution of the number of advertisers bidding in each advertising cluster (market). The distribution is highly right-skewed. Most markets are relatively small, with a limited number of advertisers competing for ad space, while a long right tail captures a small number of highly contested markets with dozens or even several hundred active advertisers. In the 2016 snapshot shown here, the bulk of markets host fewer than 50 advertisers, but there is substantial heterogeneity: some clusters attract more than 200 advertisers, and a few extreme cases exceed 400 participants. Similar patterns emerge in other years, indicating that advertising markets are typically narrow and specialized, yet coexist with a limited set of broad, highly competitive markets that concentrate a large share of advertiser participation.

Figure 4: Number of clusters ("markets") by the number of active advertisers (2016)



Notes: the histogram shows how the markets in 2016 are populated by different numbers of advertisers. Two advertisers active in the same market are considered competitors in their purchase of internet ad space.

Before moving on, it is important to remark once more different notions of markets and matching markets. Both can be described as collections of advertisers and DMAs. However, while a *matching market* is the set of all advertisers and DMAs which are active in a given year and consider each other as potential partners, *markets* are composed of those sets of advertisers which are active on at least one keyword belonging to the keyword clusters. The usage of these two notions will be different: *markets* will be used to build measures that define, for any advertiser, which are its competitors among other advertisers; while *matching markets* will be used to determine for each advertiser and agency pair, which are their potential matches.[16]

[16] As a reminder, we use a SVM algorithm to identify the *"consideration set"*: the set of potential matches out of all possible matches in a given year. We discuss the method in detail in Section 6.1. Importantly, the industry cannot be used as a restriction on the matching market since it is commonly observed that a single agency matches with advertisers operating in a broad set of industries.

4 Model

4.1 Model

The theoretical model in this section builds on Azevedo and Hatfield (2018), suitably adapted to a finite number of agents (borrowing some of the notation and definitions from Jackson and Wolinsky (1996)), in order to be able to take the model to the data. Note that, in the finite model, existence of equilibrium is not guaranteed in general (the general results of Azevedo and Hatfield (2018) do rely on the continuous setup). For our purposes however, this is not a issue, since we focus on the estimation of preferences under the assumption that the observed market is in equilibrium.

Consider a matching market with given sets of advertisers and DMAs.[17] Let I be the set of advertisers in the market, indexed by i and let J be the set of DMAs, indexed by j. Each possible matching is then a bipartite graph on I and J. Since the set of nodes remains fixed, we identify each matching with its set of edges, which is a subset of $I \times J$. We refer to members of matchings as matches. Given a matching $\Phi \subseteq I \times J$, for each $i \in I$ we let Φ_i be the set of DMAs that are matched with advertiser i ($\Phi_i = \{j \in J \,|\, (i,j) \in \Phi\}$). Similarly, for each $j \in J$, we let Φ^j be the set of advertisers matched to DMA j ($\Phi^j = \{i \in I \,|\, (i,j) \in \Phi\}$). We allow many-to-many matchings (each advertiser can be matched with multiple DMAs, and vice versa), but we rule out unmatched agents.[18] Thus, the set of all possible matches is $\Omega = \left\{\Phi \in \mathcal{P}(I \times J) \,\middle|\, \Phi_i \neq \emptyset \, \forall i \in I \text{ and } \Phi^j \neq \emptyset \, \forall j \in I\right\}$. We use "$ij$" as a shorthand for (i,j), "$\Phi + ij$" as a shorthand for $\Phi \cup \{(i,j)\}$ and "$\Phi - ij$" as a shorthand for $\Phi \backslash \{(i,j)\}$.[19]

Each advertiser and each DMA derive some inherent value from each possible matching. Let advertiser i's matching valuation function be $\nu_i : \Omega \to \mathbb{R}$ and let DMA j's matching valuation function be $\nu^j : \Omega \to \mathbb{R}$. These valuation functions may depend in a nonlinear way on the whole matching, and hence they may accommodate general *externalities*. In particular, it is not assumed that they are additively separable in valuations of single matches, and valuation $\nu_i(\Phi)$ may depend not only on the set of DMAs that advertiser i is matched to but also on the identities of other advertisers matched to these DMAs (and similarly for the other side of the market). For example, let us assume that there are three advertisers, a, b, c, and three DMAs x, y, and z. a is matched with x and y while both b and c are matched with z. Let us call this matching A: $A = \{ax, ay, bz, cz\}$. Now consider an alternative matching B, which is exactly the same as A except for the fact that b also deals with y, i.e., $B = \{ax, ay, by, bz, cz\}$. The valuation function of a may take different values under matchings A and B even though the matches involving a are the same in both. This is because in B an additional advertiser, b, is matched with y, an agency associated with a. It is easy to see that these features of the valuation function are necessary to capture complementarities.

In addition, for each matching $\Phi \in \Omega$, we define an $|I| \times |J|$ "transfers matrix" T^{Φ}, whose (i,j)-th entry,

[17] In the full model to be estimated, there will be M different matching markets, indexed by m. However, we will analyze each market separately and hence present the model for a single market.

[18] This assumption derives from the data itself. Alternative models allow independent advertisers to be matched with empty sets (Nosal, 2016).

[19] Here $\Phi \backslash K$ is the set of all matches that belong to Φ but not K. Formally, $\Phi \backslash K = \{(i,j) \in \Phi \,|\, (i,j) \notin K\}$.

t_{ij}^{Φ}, is the payment that advertiser i makes to DMA j when the matching is Φ. We assume that advertisers make payments only to agencies to which they are matched so that $t_{ij}^{\Phi}=0$ whenever $ij \notin \Phi$. Thus the set of valid transfers matrices for the matching Φ is $\mathcal{T}^{\Phi}=\left\{T^{\Phi}\in\mathbb{R}^{|I|\times|J|}\,\middle|\, j\notin\Phi_i \implies t_{ij}^{\Phi}=0\right\}$.

Advertisers and DMAs' profit functions are such that, for each $(\Phi,T^{\Phi})\in\left\{\Omega\times\mathbb{R}^{|I|\times|J|}\,\middle|\,T^{\Phi}\in\mathcal{T}^{\Phi}\right\}$:

$$\pi_i(\Phi,T^{\Phi})=\nu_i(\Phi)-\sum_{j\in\Phi_i} t_{ij}^{\Phi},\ \textit{and} \tag{1}$$

$$\pi^j(\Phi,T^{\Phi})=\nu^j(\Phi)+\sum_{i\in\Phi^j} t_{ij}^{\Phi}. \tag{2}$$

Since payments flow from one side to the other, the joint profit of advertisers and DMAs is not affected by the transfers:

$$\sum_{i\in I}\pi_i(\Phi,T^{\Phi})+\sum_{j\in J}\pi^j(\Phi,T^{\Phi})=\sum_{i\in I}\nu_i(\Phi)+\sum_{j\in J}\nu^j(\Phi)$$

Similarly, for advertiser i matched to DMA j,

$$\pi_i(\Phi,T^{\Phi})+\pi^j(\Phi,T^{\Phi})=\nu_i(\Phi)+\nu^j(\Phi)-\sum_{k\in\Phi_i\setminus\{j\}} t_{ij}^{\Phi}+\sum_{k\in\Phi^j\setminus\{i\}} t_{ij}^{\Phi}, \tag{3}$$

because the transfer from i to j cancels out.

4.2 Equilibrium Concept

Our equilibrium concept is *pairwise stability*. A matching is pairwise stable if it cannot be blocked by any pair of agents $(i,j)\in I\times J$ or by any individual agent $i\in I$ or $j\in J$. Assuming that forming a new match requires both parties to agree, whereas an existing match can be severed unilaterally by either party, matching can be blocked either by (1) an agent profitably breaking one or more of its existing matches, or (2) a pair of agents not matched to each other profitably forming a match between them (and possibly dissolving one or more of their existing matches with others). If neither of these is possible, the matching is pairwise stable. To complete the definition in our transferable utility setting, we must specify how the transfers will change in a blocking deviation. We postulate that the formation or dissolution of match ij will not affect any transfers other than the one between i and j, i.e., for all $kl\neq ij$, $t_{kl}^{\Phi-ij}=t_{kl}^{\Phi}$ and $t_{kl}^{\Phi+ij}=t_{kl}^{\Phi}$. Note that this assumption uniquely pins down $T^{\Phi-ij}$ given T^{Φ} (because $t_{ij}^{\Phi-ij}=0$, as $ij\notin\Phi-ij$).

Given valuation functions $(\nu_i)_{i\in I}$ and $(\nu^j)_{j\in J}$, a matching $\Phi\in\Omega$ with transfers matrix T^{Φ} is stable if and only if the following two conditions hold:

No agent blocks by dissolution: $\forall i\in I,\forall K\subset\{(i,j):\ ij\in\Phi\}:\pi_i(\Phi,T^{\Phi})\geq\pi_i(\Phi\setminus K,T^{\Phi\setminus K})$ and $\forall j\in J,\forall K\subset\{(i,j):\ ij\in\Phi\}:\pi^j(\Phi,T^{\Phi})\geq\pi^j(\Phi\setminus K,T^{\Phi\setminus K})$

No pair of agents blocks by rematching: $\forall ij\notin\Phi,\forall K\subseteq\{ik\in\Phi\}\cup\{kj\in\Phi\},\forall T^{\Phi+ij\setminus K}$ s.t. $\forall kl\neq$

$ij, t_{kl}^{\Phi+ij\setminus K} = t_{kl}^{\Phi\setminus K}$:
$\pi_i(\Phi + ij \setminus K, T^{\Phi+ij\setminus K}) > \pi_i(\Phi, T^{\Phi}) \implies \pi^j(\Phi + ij \setminus K, T^{\Phi+ij\setminus K}) < \pi^j(\Phi, T^{\Phi})$ and
$\pi^j(\Phi + ij \setminus K, T^{\Phi+ij\setminus K}) > \pi^j(\Phi, T^{\Phi}) \implies \pi_i(\Phi + ij \setminus K, T^{\Phi+ij\setminus K}) < \pi_i(\Phi, T^{\Phi})$

As transfers are not observable to the researcher, some manipulation on the definition of pairwise stability is needed to be able to estimate the model. This characterization will be employed for estimation following Fox (2018). With this aim in mind, let's consider a matching Φ with $ij, i'j' \in \Phi$ and at the same time $ij', i'j \notin \Phi$. Moreover, let's define $\bar{t}_{ij'}$ as the transfer that would make j' indifferent between the current matching and one where the matches $i'j'$ and ij are replaced with ij'. Isolating the transfer from the profit equivalence, we obtain:

$$\bar{t}_{ij'} = \nu^{j'}(\Phi) + t_{i'j'}^{\Phi} - \nu^{j'}(\Phi - ij - i'j' + ij').$$

Now, let $T^{\Phi-ij-i'j'+ij'}$ be obtained from T^{Φ} by setting the ij and $i'j'$ terms to zero and setting the ij' term to $\bar{t}_{ij'}$ (this clearly satisfies the condition for a valid transfer matrix in the definition of stability). Because $\pi^{j'}(\Phi - ij - i'j' + ij', T^{\Phi-ij-i'j'+ij'}) = \pi^{j'}(\Phi, T^{\Phi})$, as j' is indifferent between $\Phi - ij - i'j' + ij'$ and Φ, the matching Φ will be jointly blocked by j' and i unless $\pi_i(\Phi - ij - i'j' + ij', T^{\Phi-ij-i'j'+ij'}) \leq \pi_i(\Phi, T^{\Phi})$, i.e. if

$$\nu_i(\Phi) - t_{ij}^{\Phi} \geq \nu_i(\Phi - ij - i'j' + ij') - \bar{t}_{ij'}.$$

Let's now plug the definition of $\bar{t}_{ij'}$ into the latter inequality:

$$\nu_i(\Phi) + \nu^{j'}(\Phi) - t_{ij}^{\Phi} + t_{i'j'}^{\Phi} \geq \nu_i(\Phi - ij - i'j' + ij') + \nu^{j'}(\Phi - ij - i'j' + ij').$$

One way to read the inequality above is that the joint surplus of i and j' deriving from matching Φ in which they are not linked is greater than the joint surplus of i and j' in a matching $\Phi - ij - i'j' + ij$ in which they deal with each other and transfers cancel out. Next, we repeat the same reasoning for the remaining pair involved, namely advertiser i' and DMA j, using the alternative match $i'j$. The result is the inequality

$$\nu_{i'}(\Phi) + \nu^{j}(\Phi) + t_{ij} - t_{i'j'} \geq \nu_{i'}(\Phi - ij - i'j' + i'j) + \nu^{j'}(\Phi - ij - i'j' + i'j).$$

Finally, summing over the two inequalities, all the transfers cancel out and we have the following inequality:

$$\begin{aligned}
&\nu_i(\Phi) + \nu^{j}(\Phi) + \nu_{i'}(\Phi) + \nu^{j'}(\Phi) \\
&\quad \geq \nu_i(\Phi - ij - i'j' + ij') + \nu^{j}(\Phi - ij - i'j' + i'j) + \nu_{i'}(\Phi - ij - i'j' + i'j) + \nu^{j'}(\Phi - ij - i'j' + ij').
\end{aligned}$$

The above inequality will be referred to as the *sum of revenues inequality* and will be key for the estimation. One such inequality can be computed for each pair of matches $ij, i'j' \in \Phi$ such that $ij', i'j \notin \Phi$.

5 Identification

We follow the identification and estimation strategy of Fox (2018). Each agent has both observable and unobservable attributes. Each agent ($i \in I$ or $j \in J$) is fully identified by a corresponding pair $(h, k) \in H \times K$ (or $(\hat{h}, \hat{k}) \in \hat{H} \times \hat{K}$). We let H and $\hat{H}$ denote the (finite) set of observables for the advertisers and DMAs, respectfully; and let K and $\hat{K}$ denote the infinite set of unobservables for the advertisers and DMAs, respectfully.

The unobservables are invisible not only to the researcher but also to other agents. Consequently, it makes sense to assume that an agent's valuation of matching will not depend on these attributes.

Assumption 1: Valuations ν_i and ν^j do not depend on the unobservable attributes of agents other than i (respectively, j).

For example, if DMAs' past revenues are not in the data, then Assumption 1 means that advertisers do not have valuations over DMAs' past revenues.

The valuations ν_i and ν^j do depend, however, on the unobservable attributes of i and j (the past revenue of the advertiser, for example, influences their own choice over the DMAs). Nonetheless, it is assumed that valuations can be decomposed into a deterministic part that depends only on the observables of the evaluating agent and a stochastic component that is determined by the unobservables. That is, when agent i has type $(h, k) \in H \times K$ and agent j has type $(\hat{h}, \hat{k}) \in \hat{H} \times \hat{K}$, we have

$$\nu_i(\Phi) = \bar{\nu}_h(\Phi, \theta) + \epsilon_k(\Phi);$$
$$\nu^j(\Phi) = \bar{\nu}^{\hat{h}}(\Phi, \hat{\theta}) + \epsilon^{\hat{k}}(\Phi),$$

where:

- $\bar{\nu}_h(\Phi, \theta)$ (resp., $\bar{\nu}^{\hat{h}}(\Phi, \hat{\theta})$) is the deterministic part of the valuation function. It is a parametric function of the observable attributes h (resp., $\hat{h}$) and a vector of parameters to be estimated, θ (resp., $\hat{\theta}$);
- $\epsilon_k(\Phi)$ (resp., $\epsilon^{\hat{k}}(\Phi)$) is the unobservable or stochastic valuation part, which depends on the unobservable attributes k (resp., $\hat{k}$). The distribution of the unobserved term $\epsilon^k(\Phi)$ depends on the matching Φ, but not on the unobservables of agents other than i (resp., j).

Assumption 2: Let Φ be a matching. If $i'j \in \Phi$ and $ij, i''j \notin \Phi$, then $\nu_i(\Phi) = \nu_i(\Phi - i'j)$ and $\nu_i(\Phi) = \nu_i(\Phi + i''j)$. Similarly, if $ij' \in \Phi$ and $ij, ij'' \notin \Phi$, then $\nu^j(\Phi) = \nu^j(\Phi - ij')$ and $\nu^j(\Phi) = \nu^j(\Phi + ij'')$.

This assumption states that if advertiser i is not matched with DMA j, then i is indifferent if there is a match between j and other advertisers i' and i'', and j is indifferent if there is a match between i and other agencies j' and j''. This assumption is necessary to constrain how a valuation changes across similar matchings. It is nevertheless a weaker assumption than others used in the literature, such as for instance, the assumptions in Chen and Song (2013), where the valuation of agent i is affected only by i's own matches, Φ_i. This assumption is necessary if the researcher is interested in studying complementarities (Baccara et al., 2012; Nosal, 2016). Additionally, we impose the following assumption:

Assumption 3: The deterministic components of the valuation functions are linear in the parameters θ (resp., $\hat{\theta}$): for any h and $\hat{h}$,

$$\bar{\nu}_h(\Phi,\theta) = X(h,\Phi)'\theta;$$
$$\bar{\nu}^{\hat{h}}(\Phi,\hat{\theta}) = \hat{X}(\hat{h},\Phi)'\hat{\theta},$$

where $X(h,\Phi)$ and $\hat{X}(\hat{h},\Phi)$ are vectors of observables and θ and $\hat{\theta}$ are the vectors of parameters to be estimated.

As mentioned, the key identification assumption is that the observed matching is part of an equilibrium. However, for the reasons discussed in the introduction, we allow the agents to not necessarily be aware of all conceivable blocking coalitions, and have instead different *consideration sets*. Formally, given a matching Φ, and any matched pair $(i,j) \in \Phi$, we let $C(i,j) \subseteq I \times J$ denote their consideration set, i.e. the set of pairs (i',j') that agents j and i are aware of, respectively. Then, we assume that the observed matching Φ is C-stable in the sense that, for any $(i,j) \in \Phi$, there exists no blocking $(i',j') \in C(i,j)$.[20]

As we mentioned above, and as we will describe in the next Section, the consideration sets will be estimated from the data. For given collection of consideration sets, however, it is clear that since an equilibrium Φ depends on agents' preferences, which consist of both the observables characteristics, $X(h,\Phi)'\theta$, and the 'error terms', ϵ_k, the observed matching Φ is itself stochastic. As the discrete choice literature indicates, the deterministic part of the valuation, $X(h,\Phi)'\theta$, can be interpreted as the "mean" valuation for agents described by h. The crucial assumption on the stochastic component that we make is that the probability that different matchings are observed is ordered according to the observable components of the agents' preferences. Formally:

Assumption 4 (Rank Order): Let Φ be an arbitrary matching, and let $ij, i'j' \in \Phi$, be such that $i'j, ij' \notin \Phi$, $(i',j') \in C(i,j)$ and $(i,j) \in C(i',j')$. Then, let $\Phi^* = \Phi - ij - i'j' + i'j + ij'$ denote the matching obtained by swapping i and j's with i' and j' matches, respectively, and let the observable types corresponding to i, j, i', and j' be, respectively, h, h', $\hat{h}$, and $\hat{h}'$. Then:

$$\begin{aligned}&Pr(\Phi) > Pr(\Phi^*) \iff \\ &\bar{\nu}_h(\Phi) + \bar{\nu}_{h'}(\Phi) + \bar{\nu}^{\hat{h}}(\Phi) + \bar{\nu}^{\hat{h}'}(\Phi) > \bar{\nu}_h(\Phi^*) + \bar{\nu}_{h'}(\Phi^*) + \bar{\nu}^{\hat{h}}(\Phi^*) + \bar{\nu}_{\hat{h}'}(\Phi^*).\end{aligned}$$

[20]Note that this notion generalizes the standard notion of stability from Section 4. In particular, the standard model obtains as a special case of C-stability, letting $C(i,j) = I \times J$ for all (i,j). Note also that, as the consideration sets gets smaller, the stability notion becomes less restrictive (in the extreme case where $C(i,j) = \{(i,j)\}$ for all $(i,j) \in \Phi$, then C-stability imposes no restrictions on Φ, and it is maximally restrictive under the standard concept, where all other agents are in consideration set.

(Note that, for ease of exposition, we have omitted the arguments θ and $\hat{\theta}$ from the $\bar{\nu}$ functions above.)

This assumption simply states that between two alternative matchings, one has more probability to be observed in the data if and only if the deterministic components of the joint surplus obtained by advertisers and DMAs under that matching are greater than under the alternative.

This assumption is necessary to make an estimation for large matching markets computationally tractable. The Rank Order assumption lies at the core of the maximum score identification strategy as it enables us to work directly with inequalities instead of solving the discrete choice integral. Also note that, since the stability condition becomes more restrictive as the consideration sets get larger, the bite of this identification assumption gets weaker as the consideration sets are smaller. The Rank Order assumption explains how the unobservables enter in the choice model and it is sufficient to point identify the parameters if at least one element of $X(h, \Phi)$ has $\mathbb{R}$ as support.

Identification of the maximum score estimator relies on inequality relations rather than equalities. The inequality appearing in the Rank Order assumption allows us to characterize pairwise stability by comparing the total deterministic valuation functions arising from two alternative matchings.

From all the possible inequalities, only those that include actual matchings observed in the data are taken into consideration. Suppose Φ is the matching observed and Φ^* is the alternative matching (where Φ and Φ^* are as in the statement of Assumption 4). As before, let the observable types corresponding to i, j, i', and j' be, respectively, h, h', $\hat{h}$, and $\hat{h}'$. The inequality that compares the joint valuations of the observed matching on the left and the alternative matching on the right is:

$$\begin{aligned} &\Big((X(h,\Phi) + X(h',\Phi))'\theta + (\hat{X}(\hat{h},\Phi) + \hat{X}(\hat{h}',\Phi))'\hat{\theta}\Big) > \\ &\Big((X(h,\Phi^*) + X(h',\Phi^*))'\theta + (\hat{X}(\hat{h},\Phi^*) + \hat{X}(\hat{h}',\Phi^*))'\hat{\theta}\Big). \end{aligned} \tag{4}$$

Rearranging terms and grouping parameters common for all regressors we obtain:

$$\begin{aligned} &(X(h,\Phi) + X(h',\Phi) - X(h,\Phi^*) - X(h',\Phi^*))'\,\theta + \\ &(\hat{X}(\hat{h},\Phi) + \hat{X}(\hat{h}',\Phi) - \hat{X}(\hat{h},\Phi^*) - \hat{X}(\hat{h}',\Phi^*))'\hat{\theta} > 0. \end{aligned} \tag{5}$$

All the selected inequalities then enter a matching maximum score function, which constitutes the objective function to be maximized to estimate the true parameters. Denote the left-hand side of equation (5) with $Z(\theta, \hat{\theta})$ and let G be the set of inequalities arising from the observed matching, $Z(\theta, \hat{\theta}) \in G$. Then the Maximum score is:

$$F(\theta, \hat{\theta}) = \sum_{G} \mathbb{1}\left[Z(\theta, \hat{\theta}) > 0\right] \tag{6}$$

It is clear that the objective function does nothing more than count true inequalities, i.e., inequalities where the observed matching is more profitable than the alternative.

Which type of parameters can be identified with such a strategy? Since identification relies on inequality between different matchings, every variable which does not differ along matchings cancels out. As a result, parameters cannot be identified for fixed attributes of players.[21] Let's take once again the advertisers' and DMAs' past revenue as an example: the past revenue of a single firm will be exactly the same on the left and right side of the inequality (4), making the related parameter unidentifiable. What can be identified are parameters that multiply interactions between advertisers and DMA attributes: the product (not sum) of the number of advertiser's employees and the matched DMA's employees is different for each matching, making the associated parameter identifiable, even if of little value. Interactions are a simple subset of variables that change along with the matching.

Finally, since the inequality holds true after being multiplied on both sides by a constant factor, a scale normalization on the parameters is required. This is achieved by fixing to ± 1 a single parameter θ_0. What is important is the restriction on the absolute value $|\theta_0| = 1$, while the sign can be estimated from the data. To sum up, the identification is obtained by simulating a great number of exogenous shocks at the matching assignment level. These shocks on the assignment allow us to consider different configurations that show why the observed matching actually emerges. For a more in-depth discussion, see Fox (2018).

6 Estimation

The parameters of the advertisers and DMAs' profit functions are estimated by maximizing the maximum score objective function presented above.[22] Relative to the rapidly growing literature on the estimation of matching games, the two characterizing features of our estimation approach are to allow for many-to-many matching and for-profit functions that are not additively separable on matches, but which depend instead on the whole matching (as it is necessary in order to account for complementarities).

There is a simple intuition behind the parameters' estimation: estimates are those parameter values that, once plugged into the inequalities that define the stability of the observed match, lead to the greatest number of the inequalities being satisfied. Note that, even if the parameters were the true ones, not all the inequalities would be satisfied because of the stochastic terms ϵ^k. The intuition is that the true parameters describe the deterministic part of valuation that enters the inequality, but the observed matching is determined also by ϵ^k. Assumption 4, Rank Order, states that the observed matching makes the inequality likely to be satisfied, not necessarily satisfied with probability 1.

Starting from the observed matching Φ, any possible pair of matches is selected leading to a total of $s_m(s_m - 1)/2$ pairs, where s_m is the number of matches counted in the observed matching Φ in matching market m. In every pair of matches, up to four players are involved. Consider, for instance, a case with

[21] This is analogous to the situation in the discrete choice model's estimation, where any element of the utility that is invariant across choices cannot be estimated.

[22] Specifically, we implement an estimation algorithm in R starting from the MSE-R code by Theodore Chronis and Christina Tatli, `https://github.com/tatlchri/MSE-R` Their code builds upon the maximum score estimator for matching data developed by Jeremy Fox in a series of studies, (Fox, 2010, 2018; Santiago and Fox, 2009).

two different advertisers (i, i') matched with two different DMAs (j, j'). For every pair of matches, a new alternative matching can be formed by switching agents' matches. For pairs $ij, i'j' \in \Phi$, the alternative match is $\Phi^* = \Phi - ij - i'j' + ij' + i'j$. In the end, we obtain $s_m(s_m - 1)/2$ alternative matchings which are evaluated against the observed ones, or, equivalently, we evaluate $s_m(s_m - 1)/2$ inequalities for each market.

The counterfactual matching used in every inequality is created in such a way that it includes the exact same number of matches of the observed matching Φ, since they are exchanged but never destroyed or created. We select the entire set of inequalities.

For every theoretical inequality that is selected, we compute the vector of regressors $X(h, \Phi)$ and create the empirical inequality $Z(\theta)$. Maximizing the objective function, denoted by F, means evaluating it at each parameter θ, which boils down to plugging in the parameter in the first inequality and checking whether it is satisfied. If it is, the score is increased by 1, otherwise, it does not change. Hence, the sum of scores over each market is the value of F for a given θ:

$$F_M(\theta) = \frac{1}{M} \sum_{m \in M} \sum_{G} 1[Z(\theta) > 0] \tag{7}$$

The Maximum score estimator is: $\hat{\theta} = \arg\max_\theta F_M(\theta)$. Because the objective function is a sum of indicator functions, it is not smooth in parameters. For this reason, optimization algorithms that rely on differentiation are ill-suited for the task. Although a smooth maximum score function has been suggested to bypass such a problem (Horowitz, 1992), we keep the original non-smooth function and maximize it by relying on a differential evolution algorithm.[23]

Finally, since the distribution of errors is not known, it is necessary to infer the confidence intervals for the parameters directly from their empirical distribution. This is done through the subsampling procedure presented by Politis and Romano (1994). Subsampling is equivalent to the bootstrap procedure, apart from the fact that each sampling is done without replacement and is of size $b \ll g$, where g is the size of the original sample (i.e., the set G used to estimate the parameters). More precisely, the subsampling procedure needs the key assumption that b converges to infinity at a lower rate than g. Since the literature does not point to any suggested size for b, we use 25% of total inequalities for each subsample. These are drawn without replacement and used to estimate the parameter. The subsampling procedure is repeated 500 times. The empirical distribution of parameters derived from these 500 subsamples is then used to compute 95% confidence intervals. Subsampling is executed both at the market level and firm level, showing no significative difference in confidence interval ranges.

[23] A well-known limitation of this approach is that estimates are sensitive to the choice of the optimization algorithm, especially when optimization problems are complex (Knittel and Metaxoglou, 2014). We employ the DEoptim algorithm. The robustness of the optimization routine is checked by trying several control parameters. DE belongs to the class of genetic algorithms which use biology-inspired operations of crossover, mutation, and selection on a population, in order to minimize an objective function over the course of successive generations (see Mitchell, 1998). As with other evolutionary algorithms, DE solves optimization problems by evolving a population of candidate solutions using alteration and selection operators. For the choice of control parameters, we follow Storn and Price (1997) in setting the initial population number (50), crossover probability (0.3) and weighting factor (0.5), and Maximum iterations (1000). Boundaries are between -50 and 50. Increasing the number of iterations, lowering crossover probability, and increasing the weighting factor do not affect the estimates.

6.1 Construction of the Cosideration Sets using Support Vector Machine

One of the main challenges with this method is the unobserved consideration sets of each advertiser. On the one hand, the whole market is too large as a matching market. It is unlikely that each of the advertisers considers and compares more than 100 agencies for potential collaboration. Same for the agencies that would need to consider 300 - 400 advertisers each year. On the other hand, the definition of the matching market based on the advertiser industry is to some extent arbitrary since, as we have already noted, many agencies match with advertisers in different industries.

In order to address this issue, as well as to ease the computational complexity, we use the Support Vector Machine (SVM) method.[24] In particular, we make use of the network structure to infer possible links between agencies and advertisers. Based on the network characteristics in the current year, we solve a binary classification problem and predict what pairs match (1) or do not match (0) in the following year. In the literature this is called the *link prediction problem*: given that there is no link between two nodes in the current graph, what is the likelihood of a future link? It has been studied, for instance, for networks of co-authors (Liben-Nowell and Kleinberg (2003); Newman (2001)). If the choice sets found with this method are subsets of the true unobserved choice set, Fox (2007)'s pairwise maximum score estimator is consistent (Crawford, Griffith and Iaria (2021)).

We use the network structure in the current year to predict matches during the next year, in particular: the average degree of advertiser and agency; the average betweeness centrality of advertiser and agency; the shortest path between the advertiser and the agency; a dummy indicating the agency and the advertiser are part of a common cluster in the network; the number of common third-level neighbours. Table 4 provides variables' descriptions. The idea is that if both firms have more connections, they are more likely to match. Moreover, if an advertiser and an agency are in the same cluster, there is a direct or indirect connection between the two, hence they are more likely to match, and also if an advertiser and an agency have a direct or indirect connection in common, they are more likely to match. Crucially, we do not use the firm's characteristics that would also influence the decision to match.[25]

In order to obtain more stable estimates and achieve a higher accuracy in prediction, we employ a bagging algorithm.[26] In particular, we train an ensemble of 100 SVMs on 100 random samples of observed and unobserved matches.[27] We use years 2014 - 2015 for training, whereas years 2016 - 2020 are used for estimation. We find the bagged SVM achieve an accuracy of around 93% on the test set, with very satisfactory values of both sensitivity and specificity (both around 92%). At the same time, the ratio of true positives over the number of predicted positives is low, around 12%. This means the model produces

[24] First introduced in Boser, Guyon and Vapnik (1992), it is one of the most used machine learning algorithms for classification problems.

[25] Note that using specific firms' characteristics would make the method much less generalizable since they would change depending on the particular application.

[26] First introduced in Breiman (1996), bagging consists in sampling with replacement from the training set m times, fitting m regressors on each of the samples, and then aggregating their predictions through a majority rule.

[27] Note that, in principle, given the sparse nature of the graph, there is an extremely high number of unrealized matches in each year. Therefore, when bootstrapping, we choose a f-ratio of 2, which means we randomly select twice the number of unrealized matches with respect to the realized ones, in each sample.

many false positives with respect to the number of observed positives. This is an expected behaviour given the strongly unbalanced nature of the problem, and it is going to become helpful for the creation of the consideration sets.

Table 4: Covariates Considered by the Support Vector Machine Algorithm

Variable	**Description**
ShortestPath$_{i,j}$	Also called the geodesic distance, it is defined as the minimum number of edges connecting any two nodes in the network.
AverageDegree$_{i,j}$	The degree of a vertex is the number of edges incident to it. This variable is the average degree of firm i and matched agency j. The intuition is that pairs involving more connected firms and agencies are more likely to match.
AverageBetweenness$_{i,j}$	Betweenness measures the number of shortest paths passing through a vertex. This variable is defined as the average betweenness of firm i and matched agency j. The idea is that firms and agencies that lie on many network paths are more likely to match.
SameCluster$_{i,j}$	A cluster is defined as a connected component of the graph, i.e., a subgraph in which there exists a path linking any two nodes and which cannot be enlarged by adding other nodes. This indicator equals one if firm i and agency j belong to the same connected component, implying a direct or indirect connection between them.
N.CommonL3Neighbors$_{i,j}$	Level-3 neighbors of a vertex are nodes that are at most three edges away. If vertex i is a firm, its level-3 neighbors are agencies, and vice versa. This variable counts the number of level-3 neighbors common to firm i and agency j. The intuition is that shared direct or indirect connections increase the likelihood of a match.

Notes: Network-based covariates used by the Support Vector Machine algorithm.

We use predicted consideration sets in the following way. For each pair of matches considered in the estimation, we check if the counterfactuals (i.e., matches obtained by switching agents' matches) are predicted by the SVM algorithm. If so, the pair is selected to build an inequality for the estimation. Otherwise, the pair is not considered.

As a proof of concept for the usefulness of our method in a setting where data quality and availability are not issues, we present a simulation algorithm to create artificial data which resembles the structure of the observed network, and then apply the same estimation procedure presented above, with a simple definition for advertisers' consideration sets in Appendix A.2.

6.2 Profit Functions Specification

As discussed in the identification section, single attribute features that do not vary across advertiser-DMA pairs cannot be identified and are not included among the elements of X. All these features of the profit functions are thus captured by the stochastic term. The variables that we can include in X must all come from interactions between features of advertisers and agencies. We consider two specifications of advertisers' profit function. The first includes the variables *comp*, *compnet*, *exclusive*, and *pastmatch*. In the second specification, *comp* and *compnet* are replaced by their sum, *compfull*. The agencies' profit function, instead, depends on two variables: *affilmarkets* and *affilmatches*. We now turn to their definitions which are also presented in Table A.1 in Appendix.

The first variable, *comp*, is an intensity measure of competition among advertisers dealing with a common DMA. The variable is computed as follows: first, the number of advertising markets where advertiser i is active is selected. Second, we repeat the same procedure for every other advertiser connected to the DMA j. Then, we compute the share of common markets between i and every other advertiser connected to it through the DMA j. The shares are the number of common markets divided by the number of participated markets. Finally, all these shares are summed. Formally, *comp* is calculated as follows: denote by i the advertiser and index by $l = 1, 2, ..., n$ all other advertisers matched to DMA j, excluding i; by $\#market_i$ the number of markets where i is active; and by $\#market_{li}$ the number of markets shared by l and i. Then:

$$comp_{ij} = \left(\frac{\#market_{1i}}{\#market_i}\right) + ... + \left(\frac{\#market_{ni}}{\#market_i}\right).$$

This variable varies across all advertiser-DMA pairs, it is not symmetric for two advertisers matched to the same DMA and can be computed for both observed and counterfactual matches. These are features that will characterize also the other elements of the profit functions presented below. Before that, however, notice that from an economics perspective *comp* proxies well what an advertiser should consider of a potential DMA match when focusing on the extent to which this DMA is working with rival advertisers. *Comp* is indeed constructed as a sum to capture every possible opportunity to overlap with competitors. In this regard, notice that variables increase when an advertiser active on many markets matches with DMAs with many clients only to the extent that these clients are active on at least some common markets. Lastly, we shall argue why it is reasonable to assume that advertisers observe the set of DMAs' clients. It turns out that most DMAs openly advertise, for instance on their websites, their portfolio of clients. Not to mention the fact that the Redbooks data exist precisely because advertisers value the possibility of learning about the advertisers-agencies connections in the market.

The second variable in the advertisers' profit function is *compnet*. This is similar to *comp*, except for the fact that it is calculated at the level of an agency network, rather than DMA, excluding the DMA under consideration. This time all advertisers operating with DMAs under the same holding network as i, excluding all those advertisers working with the same DMA j as i. Note that *compnet* is complementary to *comp*. As previously stated, *compfull* is the sum of *comp* and *compnet*.

The parameters on *comp*, *compnet* and *compfull* are of particular interest to understand the digital ad market: for all three we are agnostic about the expected sign and significance. As discussed earlier, advertisers have reasons both to like and dislike DMA's sharing: industry specialization might push toward agency sharing, but "sleeping with the enemy" motives pull advertisers toward different agencies. Similarly, sharing an agency network means sharing data and bidding algorithms with rivals through the activities of the network's agency trading desk. Nevertheless, also at the level of agency networks the "sleeping with the enemy," motives might be strong enough to make advertisers prefer networks less populated by rival advertisers.

The third variable, *exclusive*, is a dummy that assumes value 1 if both the advertiser and the DMA match with exactly one counterpart (meaning together) and 0 otherwise. Since unmatched agents are not allowed, *exclusive* takes on the value of zero if and only if at least one party in the match has more than one match. The aim is to understand if agents prefer to have a monogamous business relationship when they can afford only one match. In other words, if they value exclusivity when they are forced to provide exclusivity. In the analysis, the number of matches for each player always remains the same and is therefore treated as a fixed, exogenous, observable attribute. This is equivalent to stating that every player decides how many matches he will undertake before starting the matching game and signals the number to the counterpart population.[28]

The fourth variable, *pastmatch*, is another dummy which takes value 1 if the advertiser-DMA match was formed also in the year previous to the considered one. The objective is to capture the persistence of matches over time.[29]

With regard to the DMAs' profit function, we consider two variables. They are denoted as *affilmarkets* and *affilmatches* and they are intended to capture the fact that larger advertises work with DMAs affiliated to networks or in other terms a "brand effect" of networks. The first variable, *affilmarkets*, measures whether DMAs associated with networks value advertisers who are active on multiple markets more than independent DMAs do. The variable is built by interacting a dummy variable indicating affiliation or independence (1 if affiliated, 0 if independent) with the number of markets participated in by the clients. The rationale is that affiliated intermediaries have more resources to manage complex marketing campaigns, covering many markets. The second variable, *affilmatches*, measures if affiliated intermediaries tend to match with advertisers needing multiple DMAs or not. The variable is similar to the previous one but instead of counting the number of markets, it counts the number of matches for each advertiser linked.

6.3 Estimation Results

The estimation results are presented in Table 5. The top panel reports the coefficients for the parameters in the advertisers' profit function, while the bottom panel reports those for the agencies. The coefficient of *exclusive* is scale-normalized to ± 1. The main coefficients of interest are those on *comp* and *compnet* in specification (1), and that on *compfull* in specification (2).

We find that the coefficient of *compnet* is positive and significant for both datasets, whereas *comp* is not significant. The insignificance of *comp* suggests that advertisers do not place particular importance on the presence of competitors under the agencies they work with, or that benefits counterbalance costs. On the other hand, the positive coefficient of *compnet* indicates a preference for being under the same corporate holding as competitors. This positive effect is likely also driven by the perceived benefits of shared data and bidding algorithms through agency trading desks. As previously discussed, such benefits may include lower advertising costs through coordinated bidding, more effective ad targeting, and the internalization of externalities.

Our analysis alone cannot distinguish between these motives. Nevertheless, the descriptive evidence in

[28] An extension of the same variable not included in the model would be the interaction between the advertiser number of matches and DMA number of matches, trying to understand if a player who prefers multiple matches would rather choose counterparts with many matches as well. Also note that *exclusive*, while included in the advertisers' profit function could, in principle, be included in the DMAs' one as well since it is symmetric for both sides of the market.

[29] Again, also *pastmatch* may have been introduced in the DMAs' profit function, since it is symmetric.

Table 5: Profit Functions' Parameter Estimates

	(1)	(2)
Advertiser		
comp	-0.012	
	$(-0.096, 0.648)$	
compnet	0.448^{*}	
	$(0.221, 0.82)$	
compfull		0.023^{*}
		$(0.012, 0.499)$
pastmatch	48.447^{*}	45.874^{*}
	$(41.463, 49.221)$	$(39.127, 48.089)$
exclusive	-1	-1
Agency		
affilmarkets	0.022^{*}	0.022^{*}
	$(0.015, 0.046)$	$(0.013, 0.044)$
affilmatches	-2.846	-2.683
	$(-4.084, 0.463)$	$(-4.066, 0.34)$
N ineq.	78,998	78,998

Notes: Coefficients are bounded between -50 *and* 50 *except for exclusive, whose coefficient was set to* ± 1. *Confidence intervals are based on 500 subsamples of inequalities each with relative size of 33%. They are asymmetric and the confidence level is set to 5%.*

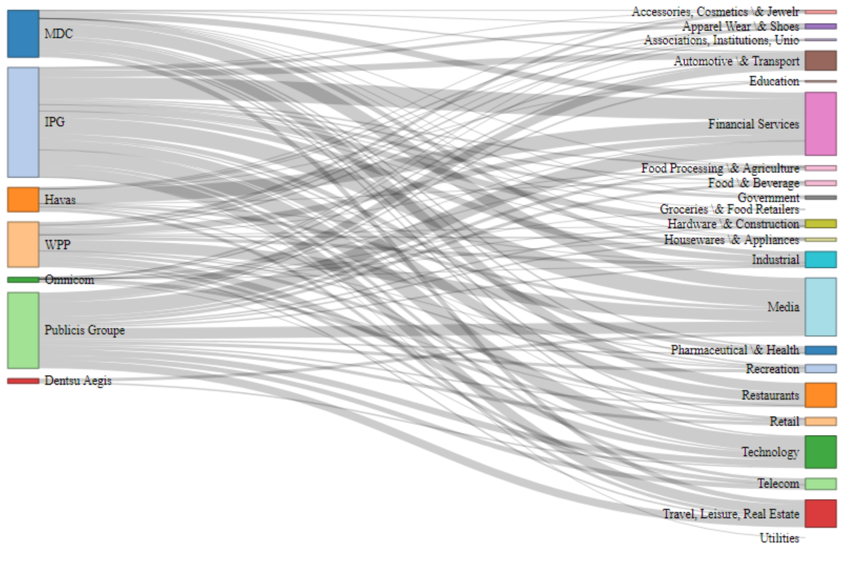
Figure 5 shows that the advertisers' preference for a common network is not merely due to the networks' specialization in particular markets. The figure represents, for each of the 7 agency networks (on the left), the industry to which their clients belong (on the right): none of the 7 networks appears to be particularly specialized in any specific industry. This visual evidence is also supported by the results in Table 6, where we report the Krugman index, the inequality in productive structure index, and the number of industries covered to measure the relative specialization of Networks in different industries.[30] Both indices point to a very low level of specialization, except for one network whose activities are circumscribed to three industries due to its limited presence in the digital segment. Network specialization, therefore, is not taking place in the whole digital advertising market, but only within industries.

The coefficient of *compfull* in specification (2) is positive and significant, with a magnitude similar to that of *compnet*. This is likely driven directly by the latter, as *compfull* is defined as the sum of *comp* and *compnet*. Again, this specification shows that there are benefits experienced by the advertisers from belonging to the same agency network as competitors in the advertising space.

Regarding the other coefficients of the advertisers' profit function, for *exclusive* we can only discuss its

[30]Krugman index is calculated as $\sum_{i=1}^{I}(|(b_i - \bar{b}_i)|)$ and takes a value between 0 (indicating that the two distributions are the same) and $2(I-1)/I$. The higher the index, the more dissimilar are the two distributions, with a value of 2 representing a situation where the two distributions have nothing in common. The Inequality in Production index is calculated as $\sum_{i=1}^{I}(b_i - \bar{b}_i)^2$ and takes a value between 0 and $(I-1)/I$, where I is the number of industries, b_i is the share of network client belonging to industry i and $\bar{b}$ the share of advertiser belonging to industry i.

Figure 5: Network clients divided by industry in 2015



Notes: the figure represents the (lack of) specialization of the 7 agency networks (on the left) in the advertisers' industries (on the right). Each line corresponds to one advertiser.

Table 6: Network Specialization at Industry Level

Networks	Industries Covered	Krugman Index	IP Index
Omnicom	11	0.878	0.067
Dentsu Aegis	3	1.232	0.221
WPP	16	0.333	0.007
Havas	12	0.692	0.047
MDC	17	0.511	0.031
Publicis Groupe	19	0.458	0.016
IPG	19	0.852	0.005

Notes: The Krugman index is calculated as $\sum_{i=1}^{I}(|(b_i - \bar{b}_i)|)$, *and takes value between 0 and* $2(I-1)/I$ *while Inequality in Production index is calculated as* $\sum_{i=1}^{I}(b_i - \bar{b}_i)^2$ *and takes value between 0 and* $(I-1)/I$. *Where* I *is the number of industries,* b_i *is the share of network clients belonging to industry* i, *and* $\bar{b}$ *is the share of advertisers belonging to industry* i. *Both indexes show that networks are not specialized in particular industries.*

sign. It is estimated to be negative, and this squares well with the presence of many-to-many matches in the data. *Pastmatch*, by contrast, is positive and significant, with large coefficient values across all specifications. This reflects the stability of the network over time: having previously partnered with a given agency is the strong predictor of maintaining that relationship in the future. This is probably due to the costs associated with initiating a collaboration with a different agency.

Turning to the agencies' profit function, we find that the coefficient on *affilmarkets* is positive and significant, whereas on *affilmatches* is negative but not significant. A positive *affilmarkets* coefficient suggests that affiliated agencies prefer advertisers operating across multiple markets, as such advertisers likely require management of larger or more complex marketing campaigns. This finding may also point to the existence of a network “brand effect”.

7 Evaluating Omnicom-IPG Merger

In the lively policy debate about how the tech giants are shaping the current economic and social landscape, competition policy has received nearly unprecedented attention. The two pillars of competition policy, merger review and antitrust, have both been invoked to respectively prevent further concentration in the tech sector and monitor, or sanction, anti-competitive behaviors. Since digital advertising is the source of financing for most of the tech sector, it is interesting to point out an application of our analysis to the problem of merger evaluation. This type of perspective analysis is conducted for both horizontal and vertical mergers when the firms involved are large enough to meet certain requirements that imply the compulsory notification of the intended merger to the public authority in charge (i.e., either the DOJ or the FTC in the US). Horizontal mergers involving products or services that are directly sold to consumers are assessed through methods that are by now well established and that, when the data allows, involve full-blown merger simulations through the estimation of structural models of demand and supply.

Our estimation approach provides a natural counterpart for the problem of assessing the effects on clients of an upstream horizontal merger between agencies. To the best of our knowledge, the idea of using a many-to-many matching estimator to evaluate the effects of mergers is novel in the literature and has never been done so far by competition authorities. The closest approach is that presented in the recent literature on the estimation of multilateral bargaining games and especially the works of Crawford and Yurukoglu (2012), Gowrisankaran, Nevo and Town (2015) and Ho and Lee (2019). While the methods in these studies could be applied to the evaluation of mergers between agencies, that would require observing the transfers. This is often difficult, even for competition authorities. Hence, we see our method as a viable alternative that allows overcoming this important shortcoming in the data. Therefore, while a full-blown merger simulation passing the high standards of a competition case is beyond the scope of this study, the remaining part of this section lays down the key elements entailed by such an approach.

7.1 Empirical Implementation

The advertising agency sector is characterized by increasing concentration. The most prominent example is the recently cleared merger between Omnicom and IPG, two of the largest global advertising networks. Each controls a broad portfolio of digital marketing agencies and serves hundreds of advertisers across nearly every major U.S. industry. Their consolidation materially reduces the number of independent agency networks, raising concerns about possible coordination, exclusionary practices, and diminished competition in the upstream provision of digital advertising services. At the same time, the transaction highlights potential efficiency gains through greater opportunities for data pooling, improvements in algorithmic bidding, and synergies across network operations.

We use our model to evaluate the Omnicom and IPG merger. To do so, we use the parameters estimated in the previous section in specification (1) to compute the profits under two different scenarios. The first scenario is the one without the M&A event, called the pre-merger scenario, while the counterfactual scenario is identical to the previous one apart from the fact that the Omnicom and IPG are treated as a unique network. In particular, all advertisers working with DMAs under Omnicom or IPG are now considered as working with DMAs under the same holding network. The analysis is conducted in partial equilibrium: in both the factual and counterfactual scenarios, the rest of the matching remains the same. Given that, since the advertiser-DMA matches did not change, *comp* variables remained the same. Instead, *compnet* is the variable that changes significantly since each of the advertisers under Omnicom or IPG is now connected to many more advertisers. The main channel through which the advertisers are affected is the network effect represented by the positive coefficient on the *compnet* variable: new affiliated DMAs bring fresh data about clients and markets, as well as enhanced possibilities for bid coordination, which help to improve the performance of all partner DMAs, benefiting all clients. Since the *compnet* estimates are positive, the combined effect on the advertisers from the acquisition is positive as well.

We focus on the normalized absolute difference in profits under the two scenarios, which is:

$$\frac{\Delta\hat{\theta}X_i}{max(|\Delta\hat{\theta}X_i|)}, \tag{8}$$

where $\Delta\hat{\theta}X_i$ is the difference between factual and counterfactual over every match.

Figure 6 summarizes the findings by showing the distribution of the gains from the acquisitions.[31]

Table 7: Normalized combined Valuation Gains

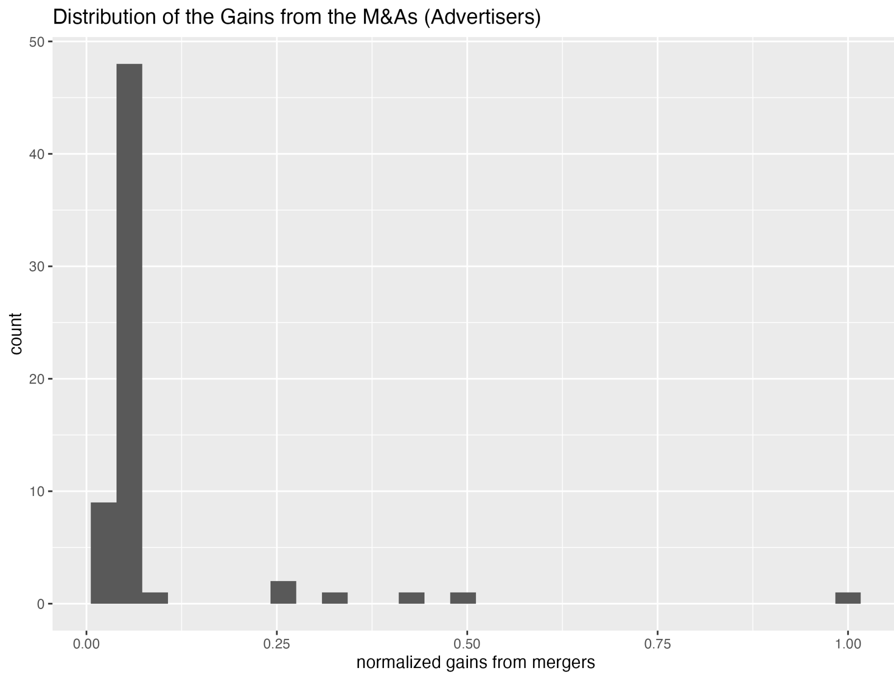
	IPG	Omnicom
Mean	0.045	0.458
Min	0.023	0.250
25%	0.041	0.273
Median	0.044	0.369
75%	0.047	0.480
Max	0.091	1.000

Notes: This table uses the point estimates obtained in the main estimation procedure to calculate the profits before and after the merger. We calculate $\frac{\Delta\hat{\theta}X_i}{max(|\Delta\hat{\theta}X_i|)}$. *IPG and Omnicom columns report statistics for the change in profits of advertisers and DMAs under IPG and Omnicom, respectfully.*

We find that our approach highlights that advertisers can win from the M&A activity in the DMA sector due to network effects. Moreover, results reveal substantial heterogeneity in advertiser gains across agency networks, with Omnicom generating markedly larger valuation improvements than IPG. While advertisers affiliated with IPG experience modest but uniformly positive gains—reflected in a mean of 0.045 and a relatively tight interquartile range—advertisers under Omnicom benefit from significantly higher and more

[31] Based on the changes in the *compnet variable.*

Figure 6: Distribution of advertisers' gains from the IPG and Omnicom merger



Notes: The figure displays the distribution of the normalized gains from the merger.

dispersed gains. In particular, the median gain under Omnicom (0.369) is an order of magnitude larger than that under IPG (0.044), and the upper tail of the distribution reaches the normalization bound, suggesting that some advertisers experience very large improvements following the ownership change. This contrast indicates that network-level complementarities are considerably stronger for Omnicom than for IPG, consistent with differences in network structure, client composition, or the effectiveness of internal coordination mechanisms.

Importantly, the fact that gains under IPG are small but pervasive, whereas gains under Omnicom are both larger and more heterogeneous, suggests distinct channels through which advertisers benefit from consolidation. In the case of IPG, network effects appear to operate primarily through incremental improvements that are broadly shared across advertisers. By contrast, Omnicom's structure seems to generate stronger complementarities for a subset of advertisers, potentially those with greater overlap in markets or higher reliance on network-level resources. Overall, these findings support the view that advertisers can benefit from M&A activity in the DMA sector due to network effects, but also highlight that the magnitude and distribution of such gains are highly network-specific rather than uniform across holding companies.

Independent industry and regulatory evidence provides external validation for the mechanisms highlighted by our model. In particular, the UK Competition and Markets Authority's (CMA) review of the Omnicom–IPG merger explicitly recognizes the role of scale and network size in shaping competitive dynamics in the advertising services industry. While the CMA ultimately cleared the transaction on the basis that the merged entity would continue to face significant competitive constraints from other global holding companies such as WPP and Publicis, its decision relies on extensive submissions from advertisers, media owners, and rival agencies describing how increased scale affects negotiating leverage, service offerings, and market outcomes.[32] The CMA's assessment thus implicitly acknowledges that network-scale effects are economically meaningful and can translate into benefits for advertisers, consistent with our finding that advertisers affiliated with larger agency networks may experience gains following consolidation.

Complementary evidence comes from industry reporting and firm-level deal rationales, which emphasize both enhanced media-buying power and deeper integration of capabilities as key consequences of the merger. Trade press coverage highlights concerns among competitors and market participants that the combined Omnicom–IPG entity will command unprecedented buying scale, enabling it to negotiate more favorable terms or bundled offerings with platforms and publishers—effects that may spill over to advertisers working within the merged network.[33] At the same time, Omnicom's public statements stress that the transaction is intended to unify creative, media, data, and technology assets under a single organizational structure, with the goal of delivering more effective and efficient solutions for clients.[34] This emphasis on integration rather than pure cost reduction closely mirrors the complementarities captured by our model and helps explain why merger-related gains may be heterogeneous across advertisers, depending on their ability to leverage the expanded network and integrated capabilities.

7.2 Discussion of the DMA Acquisition by a Network

The same framework can also be used to evaluate the acquisition (or divestiture) of a DMA by an agency network. For example, in 2017 Dentsu Aegis (one of the seven major networks) acquired the digital agency Merkle for approximately $1.5 billion. In this case, the key change is the DMA's affiliation with a network, while all other characteristics remain unchanged. As a result, the variables capturing affiliation-related characteristics—namely *compnet*, *affilmatches*, and *affilmarkets*—are directly affected by the transaction.

Advertisers and DMAs are influenced by such M&A activity through both direct and indirect channels. Directly affected players include the acquired (or divested) DMA and its clients, whose affiliation status changes. Indirectly affected players include all DMAs already affiliated with the acquiring (or divesting) network and their respective clients, whose competitive environment and network characteristics are altered as a consequence of the transaction.

While the sources of valuation changes for directly affected players are relatively straightforward, the indirect effects are less immediate. The primary channel through which affiliated DMAs are indirectly affected operates via the network effect captured by the positive coefficient on the *compnet* variable. Newly affiliated DMAs contribute additional information about clients and markets, as well as expanded opportunities for

[32] "Anticipated Acquisition by Omnicom Group Inc. of The Interpublic Group of Companies Inc.," CMA.
[33] UK Agencies Warn of Omnicom's Post-Merger "Buying Power-House."
[34] "Omnicom Completes Acquisition of Interpublic, Forming the World's Leading Marketing and Sales Company, Built for Intelligent Growth in the Next Era."

bid coordination, which improve the performance of partner DMAs within the network and, in turn, benefit their clients. This network-driven synergy provides a key explanation for why independent DMAs may find it optimal to join an agency network, and is consistent with the industry and regulatory evidence discussed in relation to the Omnicom–IPG merger.

8 Conclusion

This paper develops and estimates a structural model of many-to-many matching between advertisers and digital marketing agencies, designed to capture the rich network externalities that arise in modern digital advertising markets. Our estimates reveal a sharp distinction between agency-level and network-level effects: while advertisers do not systematically prefer sharing the same agency as their competitors, they do value being affiliated with the same agency network. This pattern is consistent with the institutional organization of the industry, in which creative and strategic decisions are largely decentralized at the agency level, whereas data aggregation, algorithmic bidding, and other scale-intensive activities are coordinated at the network level through agency trading desks. As a result, competing advertisers benefit from network-level complementarities even when agency-level conflicts remain salient.

These findings have direct implications for the ongoing consolidation of the digital marketing industry. In particular, we use our estimated model to assess the potential effects of the Omnicom–IPG merger, a transaction that substantially reduces the number of independent global agency networks. Holding existing advertiser–agency matches fixed, we evaluate how the merger reshapes valuations through changes in network composition alone. The results suggest that advertisers can experience positive valuation effects from such consolidation, driven by expanded network-level spillovers, but that these gains are highly heterogeneous and depend on the structure and client base of the merging networks. In the Omnicom–IPG case, network-level complementarities are considerably stronger on the Omnicom side, leading to larger and more dispersed advertiser gains than those associated with IPG.

More broadly, this exercise illustrates how our framework can serve as a practical tool for merger evaluation in vertical and intermediary markets where transfers are unobserved and traditional price-based approaches are infeasible. By explicitly modeling matching decisions, competitive overlap, and organizational structure, the framework provides a quantitative way to assess how consolidation among intermediaries affects downstream firms through non-price channels.

Finally, the combination of structural modeling with machine learning techniques offers promising avenues for future research. The use of support vector machines to reduce the dimensionality of the matching problem allows the model to scale to large, complex markets while preserving economically meaningful heterogeneity. This integration makes it feasible to study richer counterfactuals, such as alternative organizational designs within networks, endogenous re-matching following mergers, or the role of data-sharing and exclusivity in shaping market outcomes.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak. 2017. “The welfare effects of coordinated assignment: Evidence from the New York City high school match.” *American Economic Review*, 107(12): 3635–89.

Agarwal, Ashish, and Tridas Mukhopadhyay. 2016. “The impact of competing ads on click performance in sponsored search.” *Information Systems Research*, 27(3): 538–557.

Agarwal, Nikhil. 2017. “Policy Analysis in Matching Markets.” *American Economic Review*, 107(5): 246–50.

Athey, Susan, and Denis Nekipelov. 2010. “A structural model of sponsored search advertising auctions.” Vol. 15.

Azevedo, Eduardo M, and John William Hatfield. 2018. “Existence of equilibrium in large matching markets with complementarities.” *Available at SSRN 3268884*.

Baardman, Lennart, Elaheh Fata, Abhishek Pani, and Georgia Perakis. 2019. “Learning Optimal Online Advertising Portfolios with Periodic Budgets.” *Available at SSRN 3346642*.

Baccara, Mariagiovanna, Ayşe İmrohoroğlu, Alistair J Wilson, and Leeat Yariv. 2012. “A field study on matching with network externalities.” *American Economic Review*, 102(5): 1773–1804.

Balseiro, Santiago R, and Ozan Candogan. 2017. “Optimal contracts for intermediaries in online advertising.” *Operations Research*, 65(4): 878–896.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. “Automobile prices in market equilibrium.” *Econometrica: Journal of the Econometric Society*, 841–890.

Berto Villas-Boas, Sofia. 2007. “Vertical relationships between manufacturers and retailers: Inference with limited data.” *The Review of Economic Studies*, 74(2): 625–652.

Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. 1992. “A Training Algorithm for Optimal Margin Classifiers.” 144–152. ACM.

Breiman, L. 1996. “Bagging Predictors.” *Machine Learning*, 24: 123–140.

Chen, Jiawei, and Kejun Song. 2013. “Two-sided matching in the loan market.” *International Journal of Industrial Organization*, 31(2): 145–152.

Chiappori, Pierre-André, and Bernard Salanié. 2016. “The econometrics of matching models.” *Journal of Economic Literature*, 54(3): 832–61.

Choi, Hana, Carl F Mela, Santiago Balseiro, and Adam Leary. 2017. "Online display advertising markets: A literature review and future directions." *Columbia Business School Research Paper*, , (18-1).

Choo, Eugene, and Aloysius Siow. 2006. "Who marries whom and why." *Journal of political Economy*, 114(1): 175–201.

Condorelli, Daniele, Andrea Galeotti, and Ludovic Renou. 2016. "Bilateral trading in networks." *The Review of Economic Studies*, 84(1): 82–105.

Crawford, Gregory S., and Ali Yurukoglu. 2012. "The Welfare Effects of Bundling in Multichannel Television Markets." *American Economic Review*, 102(2): 643–85.

Crawford, Gregory S, Rachel Griffith, and Alessandro Iaria. 2021. "A survey of preference estimation with unobserved choice set heterogeneity." *Journal of Econometrics*, 222(1): 4–43.

Day, D. 2014. "Matching with Conflicts: An Application to the Advertising Industry." mimeo.

Decarolis, Francesco, and Gabriele Rovigatti. 2021. "From mad men to maths men: Concentration and buyer power in online advertising." *American Economic Review*, 111(10): 3299–3327.

Decarolis, Francesco, Maris Goldmanis, and Antonio Penta. 2020. "Marketing agencies and collusive bidding in online ad auctions." *Management Science*, forthcoming.

Decarolis, Francesco, Maris Goldmanis, Antonio Penta, and Ksenia Shakhgildyan. 2023. "Bid coordination in sponsored search auctions: Detection methodology and empirical analysis." *The Journal of Industrial Economics*, 71(2): 570–592.

Donna, Javier, Pedro Pereira, Tiago Pires, and Andre Trindade. 2018. "Measuring the Welfare of Intermediaries in Vertical Markets." *Available at SSRN 3285205.*

Edelmann, Benjamin, Michael Ostrovsky, and Michael Schwarz. 2007. "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords." *American Economic Review*, 97(1): 242–259.

Ellickson, Paul B, Stephanie Houghton, and Christopher Timmins. 2013. "Estimating network economies in retail chains: a revealed preference approach." *The RAND Journal of Economics*, 44(2): 169–193.

Fox, Jeremy T. 2007. "Semiparametric estimation of multinomial discrete-choice models using a subset of choices." *The RAND Journal of Economics*, 38(4): 1002–1019.

Fox, Jeremy T. 2010. "Identification in matching games." *Quantitative Economics*, 1: 203–254.

Fox, Jeremy T. 2018. "Estimating matching games with transfers." *Quantitative Economics*, 9(1): 1–38.

Fox, Jeremy T, and Patrick Bajari. 2013. "Measuring the efficiency of an FCC spectrum auction." *American Economic Journal: Microeconomics*, 5(1): 100–146.

Goldman, Matt, and Justin M Rao. 2014. "Position Auctions in Practice." *Available at SSRN 2524688.*

Gomes, Renato, and Kane Sweeney. 2014. "Bayes-Nash Equilibria of the Generalized Second Price Auction." *Games and Economic Behavior*, 86: 421–437.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town. 2015. "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry." *American Economic Review*, 105(1): 172–203.

Hausmann, Ricardo, and César A Hidalgo. 2011. "The network structure of economic output." *Journal of Economic Growth*, 16(4): 309–342.

Ho, Kate, and Robin S Lee. 2019. "Equilibrium provider networks: Bargaining and exclusion in health care markets." *American Economic Review*, 109(2): 473–522.

Horowitz, Joel L. 1992. "A smoothed maximum score estimator for the binary response model." *Econometrica: journal of the Econometric Society*, 505–531.

Jackson, Matthew O, and Asher Wolinsky. 1996. "A strategic model of social and economic networks." *Journal of economic theory*, 71(1): 44–74.

Knittel, Christopher R, and Konstantinos Metaxoglou. 2014. "Estimation of random-coefficient demand models: two empiricists' perspective." *Review of Economics and Statistics*, 96(1): 34–59.

Liben-Nowell, David, and Jon Kleinberg. 2003. "The link prediction problem for social networks." 556–559.

Manski, Charles F. 1975. "Maximum score estimation of the stochastic utility model of choice." *Journal of econometrics*, 3(3): 205–228.

McAfee, R. Preston. 2011. "The Design of Advertising Exchanges." *Review of Industrial Organization*, 39(3): 169–185.

McFadden, Daniel, et al. 1973. "Conditional logit analysis of qualitative choice behavior."

Newman, Mark EJ. 2001. "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." *Physical review E*, 64(1): 016132.

Nosal, Kathleen. 2016. "Two-Sided Matching in Physician-Insurer Networks: Evidence from Medicare Advantage." *Available at SSRN 2727272.*

Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii. 2015. "Moment inequalities and their application." *Econometrica*, 83(1): 315–334.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "GloVe: Global Vectors for Word Representation." 1532–1543. Doha, Qatar:Association for Computational Linguistics.

Politis, Dimitris N, and Joseph P Romano. 1994. "Large sample confidence regions based on subsamples under minimal assumptions." *The Annals of Statistics*, 2031–2050.

Santiago, David, and Jeremy T. Fox. 2009. "A Toolkit for Matching Maximum Score Estimation and Point and Set Identified Subsampling Inference." *miemo.*

Shakhgildyan, Ksenia. 2025. "Nonparametric Identification and Estimation of the Generalized Second-Price Auction." *Games and Economic Behavior*, 150: 480–500.

Silk, Alvin J, and Charles King III. 2013. "How Concentrated is the US Advertising and Marketing Services Industry? Myth versus Reality." *Journal of Current Issues & Research in Advertising*, 34(1): 166–193.

Spulber, Daniel F. 1996. "Market microstructure and intermediation." *Journal of Economic perspectives*, 10(3): 135–152.

Storn, Rainer, and Kenneth Price. 1997. "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces." *Journal of global optimization*, 11(4): 341–359.

Train, Kenneth, and Melvyn Weeks. 2005. "Discrete choice models in preference space and willingness-to-pay space." In *Applications of simulation methods in environmental and resource economics.* 1–16. Springer.

Uzzi, Brian. 1996. "The sources and consequences of embeddedness for the economic performance of organizations: The network effect." *American sociological review*, 674–698.

Varian, Hal R. 2007. "Position auctions." *international Journal of industrial Organization*, 25(6): 1163–1178.

Villas-Boas, J. Miguel. 1994. "Sleeping with the Enemy: Should Competitors Share the Same Advertising Agency?" *Marketing Science*, 13(2): 190–202.

A Appendix

A.1 Variable Descriptions

Table 8: Variable Descriptions

Variable	**Formula**	**Description**
$\text{comp}_{i,j}$	$\sum_{k \in C_j \setminus \{i\}} \sum_{m \in M} \frac{\mathbb{1}\{i,k \in m\}}{\sum_{m \in M} \mathbb{1}\{i \in m\}}$	Measures the intensity of competition faced by advertiser i through agency j. For each advertiser k matched with the same agency j, we compute the share of markets in which i and k are both active, normalized by the total number of markets in which i operates. The variable is the sum of these shares across all co-clients of agency j.
$\text{compnet}_{i,j}$	$\sum_{k \in N_j \setminus C_j} \sum_{m \in M} \frac{\mathbb{1}\{i,k \in m\}}{\sum_{m \in M} \mathbb{1}\{i \in m\}}$	Analogous to $\text{comp}_{i,j}$, but computed at the agency network level. Competition is measured with respect to advertisers matched to agencies belonging to the same holding network as j, excluding those matched through agency j itself.
$\text{exclusive}_{i,j}$	$\mathbb{1}\big(\lvert A_i \rvert = 1 \ \wedge\ \lvert A_j \rvert = 1\big)$	Dummy variable equal to one if advertiser i and agency j are exclusively matched to each other, and zero otherwise.
$\text{pastmatch}_{i,j}$	$\mathbb{1}\big((i,j) \text{ matched in } t-1\big)$	Dummy equal to one if advertiser i and agency j were matched in the previous year, capturing persistence in matching relationships.
$\text{affilmarkets}_{j,i}$	$\sum_{m \in M} \mathbb{1}\{i \in m\} \cdot \mathbb{1}(j \text{ affiliated})$	Captures whether affiliated agencies value advertisers operating in many markets more than independent agencies.
$\text{affilmatches}_{j,i}$	$\lvert A_i \rvert \cdot \mathbb{1}(j \text{ affiliated})$	Measures whether affiliated agencies tend to match with advertisers that rely on multiple agencies.
Definitions. M denotes the set of advertising markets (keyword clusters). A_i denotes the set of agencies matched with advertiser i, and A_j the set of advertisers matched with agency j. $C_j \subseteq A_j$ denotes the set of advertisers matched with agency j. N_j denotes the set of advertisers matched with agencies belonging to the same holding network as agency j.		

A.2 Monte Carlo Simulations

A.2.1 Simulation setting

Define the set of advertisers I and the set of DMAs J. Any matching is a subset of $I \times J$. In our simplified simulation model, each DMA and each advertiser is characterized by two attributes, which we call X_1 and X_2 for simplicity. We assume both variables follow a lognormal distribution with equal parameters, and they are iid across DMAs and advertisers.[35] These variables can be thought as individual characteristics proportional to the probability of forming a match with any other agent (e.g. they could represent the ability of an agency to promote advertisers, or the diversification of an advertiser's interests, which leads to the necessity of more DMAs for his promotion). We can then define the set of *admissible matches* A as the set of matches in which the utility that both parties obtain from the match is higher than their outside option utility. That is:

$$A = \big\{(i,j) \mid x_{1i}x_{1j}\beta_1 + x_{2i}x_{2j}\beta_2 \geq Q_p(X_1^I X_1^J)\beta_1 + Q_p(X_2^I X_2^J)\beta_2 + \epsilon_{ij},\ \forall i \in I,\ j \in J\big\} \tag{9}$$

Note that we are assuming the utility at the match level is given by a weighted sum of the value of the variables at the match level, which is given by the product of their values at the agents' level. The weights (or coefficients) are the parameters of our interest: β_1 and β_2. The outside option utility instead depends on $Q_p(X_1^I X_1^J)$ and $Q_p(X_2^I X_2^J)$, which are the p-th percentiles of the sample distribution of $X_1^I \cdot X_1^J$

[35] There is nothing special about this choice: it was made in order to obtain a positive right-skewed distribution.

and $X_2^I \cdot X_2^J$.[36] The outside option utility also depends on ϵ_{ij}, a random term that captures noise in the agents' estimation of their outside option utility. We assume ϵ follows a normal distribution with mean 0 and standard deviation equal to that of X_1 and X_2 times a constant c. Intuitively, the higher c, the less the agents are able to estimate their outside option utility, the more matches are "incorrectly formed" near the threshold.

The set of *considered matches* C is defined as the set of matches in which each agent is in the other agent's consideration set. In this simplified setting, we assume consideration sets are built "like cities": advertisers and DMAs are each assigned to one city and they can only consider forming matches with other agents within the same city. Assuming we have n cities in total, the probability of an agent being assigned to any city is $1/n$ for all cities, and each assignment is independent. This means within each city the set of considered matches is the complete bipartite graph between the set of DMAs and advertisers belonging to such city, while across different cities no match can be considered.

Finally, define the set of *realised matches* R as the set of matches that are both admissible and considered, that is, $R = A \cap C$.

To fully replicate our methodology, and train the bagged SVM model to perform inference on the consideration sets, one needs to simulate at least three periods of time: the SVM is trained to predict link existence at $t = 2$ based on information available at $t = 1$, and then it can perform inference at $t = 3$ on the unseen network. Hence, we need a way to model evolution of the network in time. Again, we choose to keep things as simple as possible. In particular, we keep the set of considered matches C fixed, and only evolve over time the set of admissible matches A. The values of X_1 and X_2 and their respective parameters are kept constant in time, but the random term ϵ in the admissibility requirement 9 is re-drawn at different time periods, from the same distribution. Hence the set of admissible matches at time t is

$$A_t = \left\{(i, j) \mid x_{1i}x_{1j}\beta_1 + x_{2i}x_{2j}\beta_2 \geq Q_p(X_1^I X_1^J)\beta_1 + Q_p(X_2^I X_2^J)\beta_2 + \epsilon_{ijt},\ \forall i \in I,\ j \in J\right\} \tag{10}$$

where $t = 1, 2, 3$, and $\epsilon_{ijt} \sim \mathcal{N}\left(0, c \cdot \mathrm{sd}(X_1)\right)\ \forall i, j, t$. From A_t, as before we can derive $R_t = A_t \cap C$. Note that we are assuming there is no time dependency between periods.

A.2.2 Simulated network

In this section, we report the hyperparameters used for simulation, and we show visually the structure of the network. Note that, when referring to μ and σ in Table 9, we are talking about the mean and variance of the normal random variable used to define X_1 and X_2 (which are lognormal), that is:

$$Y \sim \mathcal{N}(\mu, \sigma), \quad X_1 \stackrel{d}{=} X_2 \sim e^Y \tag{11}$$

Figure 7 shows the whole simulated network with the reported hyperparameters, while Figure 8 only

[36] In simulation, we use $p = 75\%$.

displays one of the three "cities", or consideration sets.

parameter	value
N advertisers	400
N agencies	100
N cities	3
μ	2
σ	1
β_1	0.6
β_2	0.4
p	0.75
c	4

Table 9: Hyperparameters used for simulation

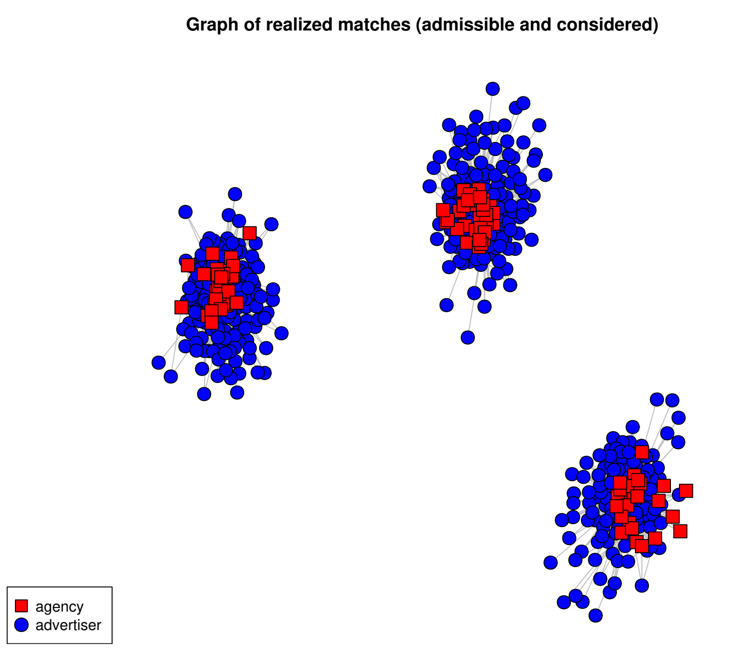


Figure 7: Graph of realized matches with default settings

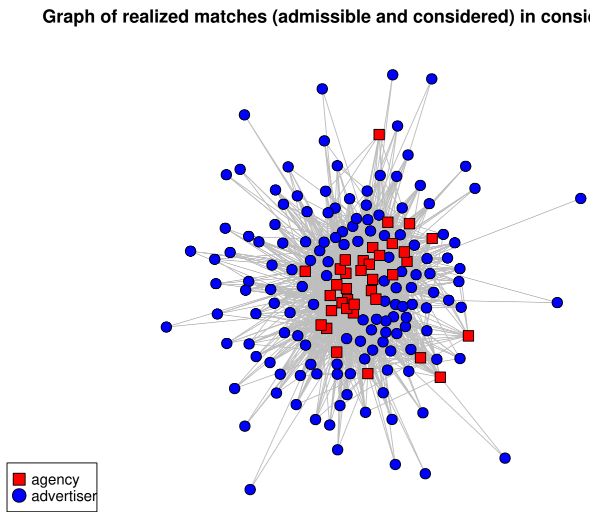


Figure 8: Graph of realized matches with default settings within a single city

A.2.3 Simulation Results

We apply the same bagged SVM algorithm used on real data on simulated data, with very minor differences.[37] We train the ensemble model using the network structure at $t = 1$ to predict the existence of matches at $t = 2$, and then with the same model predict match existence at $t = 3$.

The predictive performance of the bagged SVM on simulated data is similar to that on real data: balanced accuracy is around 90%, as sensitivity and specificity are respectively around 92% and 88% (clearly, results may strongly vary with different hyperparamter choices in the simulation algorithm).

As before, for each pair of observed matches at $t = 3$, we check if the pair of counterfactual matches is predicted by the model, and if so, the pair is used to build an inequality. We call the set of resulting empirical inequalities $\bar{Z}$. We also consider two additional sets of inequalities. Firstly, we define the set of inequalities $\tilde{Z}$, which is built by keeping all pairs of observed matches where both counterfactual matches are part of C, that is, such matches were considered by the agents, but they were not formed as the admissibility constraint in 10 was not satisfied at $t = 3$. This is the set of inequalities on which we would perform inference if we had perfect information on consideration sets. Secondly, define Z, where all pairs of observed matches are kept.[38] This is the set of empirical inequalities on which we would perform inference if we had no knowledge or way to estimate consideration sets of agents, as we pose no restriction on the counterfactual matches included,

[37]There are two relevant changes. First, the N.CommonL3Neighbors variable was dropped. In fact, in our simulated setting third-degree neighbors tend to capture the whole cluster (or "city") in which the agent is present, as simulated networks tend to be more connected than observed ones. Secondly, we only used 10 models for the ensemble (instead of 100) due to efficiency reasons.

[38]Note that this is not equivalent to $R \times R$ as one may intuitively think, as we can only keep pairs of matches where both of the counterfactual matches are not observed, while very often crossing two observed matches leads to one or more other observed matches.

which means we are essentially assuming that all counterfactual matches were considered by agents. Clearly, $\bar{Z} \subseteq Z$ and $\tilde{Z} \subseteq Z$.

Table 10: Estimates of β_1 and β_2 on $\tilde{Z}$, $\bar{Z}$ and Z for simulation with $\beta_1 = 0.6$ and $\beta_2 = 0.4$.

	$\tilde{Z}$	$\bar{Z}$	Z
X_1	1	1	1
	-	-	-
X_2	0.645*	0.633*	1
	(0.604, 0.684)	(0.582, 0.692)	(1, 1)
N ineq.	331,838	209,675	7,224,697

Notes: Coefficients are bounded between -1 and 1; the coefficient for X_1 is fixed at 1. Confidence intervals are based on 500 subsamples of inequalities with relative size of 33%. They are asymmetric and the confidence level is set to 5%.

The results of the parameter estimation on all three sets are collected in Table 10, where the network was simulated using $\beta_1 = 0.6$ and $\beta_2 = 0.4$. As before, we need to fix the coefficient of a variable to ± 1, in this case X_1, hence the correct estimate for the coefficient of X_2 would be the ratio between the true coefficients, that is $\frac{\beta_2}{\beta_1} \approx 0.667$. Note that performing estimation on the entire set of inequalities Z does not allow to pin down the coefficients, as the estimate of X_2 tends to diverge outside the bounds. Instead, exploiting the information on the consideration sets given by the bagged SVM, and estimating coefficients on $\bar{Z}$ leads to a similar performance to that obtained when performing estimation with full information on the consideration sets composition: in both cases the true coefficient is captured in the confidence interval, and the spread of the intervals is similar. This shows the usefulness of the consideration sets' estimation through the bagged SVM to identify the correct coefficients.