



European
University
Institute

ROBERT
SCHUMAN
CENTRE FOR
ADVANCED
STUDIES

WORKING PAPERS

RSCAS 2014/07
Robert Schuman Centre for Advanced Studies

Truncation in the Matching Markets and Market
Inefficiency

Janine Balter, Michela Rancan, Olena Senyuta

European University Institute
Robert Schuman Centre for Advanced Studies

Truncation in the Matching Markets and Market Inefficiency

Janine Balter, Michela Rancan, Olena Senyuta

EUI Working Paper **RSCAS** 2014/07

This text may be downloaded only for personal research purposes. Additional reproduction for other purposes, whether in hard copies or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper, or other series, the year and the publisher.

ISSN 1028-3625

© Janine Balter, Michela Rancan, Olena Senyuta, 2014

Printed in Italy, January 2014

European University Institute

Badia Fiesolana

I – 50014 San Domenico di Fiesole (FI)

Italy

www.eui.eu/RSCAS/Publications/

www.eui.eu

cadmus.eui.eu

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies (RSCAS), created in 1992 and directed by Brigid Laffan since September 2013, aims to develop inter-disciplinary and comparative research and to promote work on the major issues facing the process of integration and European society.

The Centre is home to a large post-doctoral programme and hosts major research programmes and projects, and a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration and the expanding membership of the European Union.

Details of the research of the Centre can be found on:

<http://www.eui.eu/RSCAS/Research/>

Research publications take the form of Working Papers, Policy Papers, Distinguished Lectures and books. Most of these are also available on the RSCAS website:

<http://www.eui.eu/RSCAS/Publications/>

The EUI and the RSCAS are not responsible for the opinion expressed by the author(s).

Abstract

In this paper, we study the Ph.D academic job market. Based on the Gale and Shapley algorithm, we analyse whether a social planner can improve market efficiency by truncation, i.e., exogenously imposing a limit on the number of possible applications. Using simulations, we derive the optimal truncation level which balances the trade-off between being unmatched and gaining a better match in the aggregate. When graduates apply to their most preferred positions, we find that aggregate efficiency can be improved by limiting the number of applications. In particular, the limit can be considerable if the graduates' preferences over the positions are not very correlated. The derived limit is still the best one when graduates respond strategically (applying to universities which are at least individually most preferred at the expenses of those most preferred commonly) in a conservative sense: given the strategic behaviour of the graduates, the market efficiency can be further improved by choosing an even lower limit on the number of applications. Overall, this paper suggests a direction to improve the matching market for Ph.D. candidates by improving the quality of their matches and lowering the hiring costs for universities.

Keywords

Matching markets, Truncation, Gale-Shapley deferred acceptance algorithm, Preference misrepresentation

1. Introduction¹

In the last decades, an increased number of research doctorates have been awarded all over the world. The number of new doctorates increased by 40% between 1998 and 2008 (Cyranoski et al., 2011).² Nonetheless, the number of tenured positions in academia has decreased. As a consequence, the competition for new academic positions is significantly increased, and the number of applications received by universities in the recruitment period is abundant. The present paper, motivated by this evidence, presents a rationale for imposing a limit on the number of possible applications in the context of the one-to-one matching problem. We show that the introduction of this restriction can increase the match efficiency. This finding may seem counter-intuitive because a larger pool of candidates undoubtedly has the advantage of guaranteeing an ample choice set for both sides of the market and, in principle, a better matching for universities. However, the recruitment committees would have to look at a larger number of applications, which has some direct costs in terms of time and resources. In contrast, with the advent of the Internet and the web, the costs for applicants have become negligible. The problem has been exacerbated by the introduction of on-line platforms, such as EconJobMarket.org (EJM).³ Since fall 2007, EJM operates as an intermediary in the job market for academic economists, providing a secure central repository for the files of candidates. But in a very short time, the average number of applications that an employer receives per advertisement posted on EJM increased dramatically, from 134 in 2008/2009 to 242 in 2011/2012 (Bandyopadhyay et al., 2013). Thus, those on-line labor market intermediaries, originally created to improve the information flows between market participants, turned out to be the vehicle through which recruiters receive hundreds of applications.

Market participants are already concerned that there are too many ap-

¹This working paper has been written during our stay at the EUI. Balter and Senyuta are grateful to the Max Weber Programme. Rancan thanks the RSCAS for the hospitality. The authors have benefited from discussion with David Levine, Ramon Marimon, Jordi Massó, and Antonio Nicoló. All remaining errors are ours.

²A significant number of them were awarded in China, which has recently become the world's biggest producer of Ph.D.s, overtaking the U.S.

³Other intermediaries in the labor market for economists include Econ-jobs.com, Econ-jobs.com, thesupplycurve.com, and Walras.org. More broadly, for academic jobs, there is Academickeys.com, and in the general labor market a well-known web site is Monster.com.

plications submitted per job position (Bandyopadhyay et al., 2013). In order to help deal with the numerous applications submitted via EJM, a special web-based software, HeadHunter,⁴ was developed. The problem with EJM is that it helps to collect the applications, but not to analyze them. HeadHunter was developed to organize and coordinate the process of analyzing the applications in electronic form. This software has a direct connection to the EJM. When an institution places an advertisement for a position on EJM, all application materials for that position are transferred by HeadHunter to a separate database. HeadHunter allows recruiting committee to analyze the applications effectively, rate them, communicate and make decisions about interviews. It enables a recruiting committee to track the applicants from the moment they have applied for the position to the final stage of making offers. The time the recruiting committee spends on sorting documents and identifying what items in the applicant folders are missing is decreased. In turn, the committee has more time to read the contents of the applications.

In addition to the costs of processing the applications, it might be more difficult to identify the proper candidate to interview, and finally, to offer a position. Data on assistant professors in science and engineering from 1990 till 2001 at 14 United States universities suggest that almost 9% of them left their positions within the second year.⁵ The decision of a new faculty member to leave a position after such a short stay in a university is most likely the outcome of a bad matching.⁶ This undesirable outcome for universities results in further administrative costs and disruption of research and teaching programmes.

Several possibilities exist to address the problem of large number of applications per position both in decentralized and centralized systems. One way to deal with the issue is to introduce auctions or application fees (Bandyopadhyay et al., 2013; Damiano and Li, 2007). Such mechanisms could, at least partially, reveal information about applicants' preferences, and unnecessary applications would be avoided. Another possibility is to introduce a

⁴<https://editorialexpress.com/hh/2013.html>.

⁵Computations are performed using the data by Kaminski and Geisler (2012) for a cohort of 1594 assistant professors.

⁶In addition to bad matching other reasons or unforeseen circumstances could explain the departure of a new faculty member. However, the percentage is still surprisingly high. Also, the number of assistant professors unsatisfied with their jobs is at a high level (National Study of Postsecondary Faculty, 1999 and 2004).

signaling device to improve the matching outcomes. For example, this mechanism has been developed by the American Economic Association Ad Hoc Committee in the labor market for new economists.

Since 2006, Ph.D. candidates can send a signal about their special interest to up to two universities. Using job market data from the 2006-2009, Coles et al. (2010a) find that this device has been effective in increasing the probability of receiving an interview.⁷

In addition, in centralized admission processes, in contrast to decentralized markets, there is the possibility of limiting the number of applications an applicant can send. In this way, the number of applications which have to be reviewed per position can be effectively decreased. Furthermore, such a limit would force applicants to select carefully which positions to apply to, which means that the application itself would convey information on the applicants' preferences. Several examples are available in Europe. For instance, in the French academic job market, universities could rank only 5 candidates per position until the year of 2009. In Germany, the SfH supports German universities in the allocation of undergraduates to university places for Medicine and Pharmacy. The introduction of this system was necessary as the number of students have increased steadily. In the first stage of the process, students can submit a list of 6 universities.⁸

Motivated by these markets, we consider the problem of Ph.D job market from a central planner's perspective and we study the trade-off which arises when we exogenously limit the number of applications (truncation). In the present paper, we consider the effects of truncation on both sides of the market (Ph.Ds and universities) and we derive the optimal level of possible applications accounting for the strategic responses of Ph.Ds facing a limited number of universities to which they are allowed to apply.

The matching problem under consideration is similar to the marriage market which was analyzed in the seminal paper by Gale and Shapley (1962). They study a pairwise matching mechanism and derive the equilibrium as a stable matching which occurs when no couples would broke up and formed new matches which would make them better off. In our analysis, using the

⁷Nonetheless, there may still be students and universities unmatched, and thus a further device has been introduced: the scramble markets. This secondary market facilitates the exchange of information regarding unfilled positions and Ph.D. candidates looking for a position in the late spring of every year.

⁸Students are then matched according to their final grade at school.

Gale and Shapley matching mechanism, we derive the optimal truncation level in the number of applications. When all candidates submit an exogenously restricted number of applications, some of them will be unmatched, but, at the same time, other graduates will obtain a better match compared to the alternative of submitting a full list. Therefore, the optimal level of restriction of the number of applications is determined in order to improve the overall matching outcome and, at the same time, the risk of unmatched candidates is minimized.⁹

However, when the number of applications is exogenously limited and the preferences among Ph.Ds are correlated, graduates might use different strategies and misrepresent their preferences.¹⁰ Misrepresentation of preferences is equivalent to graduates not applying to their most preferred positions (including the inferior “safe option”), which reduces the risk of being unmatched. In general, the optimal response strategy will depend on the truncation level and the correlation in the graduates’ preferences. We analyze the inefficiency of the matching outcomes, and how it changes with different level of truncation. Using the simulation results, we are able to show that the aggregate level of matching inefficiency for Ph.D. graduates can be decreased if a limit on the number of applications is imposed. An optimal restriction on the number of applications p^* exists, and it increases as the graduates’ preferences become more correlated. In particular, social planner should not restrict the number of applications too much if graduates have very correlated preferences. Moreover, we find that the optimal level of truncation and the resulting strategic responses vary with the market features, such as number of available positions and preferences correlation. While Ph.D. graduates gain from the introduction of this restriction, universities experience greater inefficiency, although our framework does not take the direct and indirect costs of processing a large number of applications into account.¹¹ Altogether, universities may also benefit from imposing a limit

⁹In this paper, we emphasize the benefits of this truncation in a centralized setting; however, even in a partially decentralized system, truncation might be beneficial (i.e., when it has a signaling value, an optimal truncation level will emerge endogenously under particular conditions).

¹⁰Misrepresentation of preferences was already considered by Roth (1982), who states that no matching mechanism which can ensure no misrepresentation of preferences exists.

¹¹Such costs of making an offer are present in the academic job market. According to the data from the 1995-1996 survey, average costs for visiting the AEA conference by the de-

on the number of possible applications.

Overall, this paper suggests a direction to improve the matching market of Ph.D. candidates and contain the inefficiency which arises from the features of many matching markets such as incomplete information, large size and the high costs of initial screening. While our setting fits the matching problem between universities and Ph.D. graduates well, it might be generalized to other markets. Implementation and policy considerations might differ depending on the specificities of the context and will be the object of further studies. In addition, our paper shows a desirable property of matching markets with a limited number of applications. When universities are the proposing side of the market and candidates apply to all possible positions, employers minimize their inefficiency at the expense of the Ph.D. graduates; in contrast, when Ph.D. graduates propose, their inefficiency would be minimized at the expenses of the universities. Our results show that, by introducing the “right” limit on the number of applications, a more even distribution of inefficiency between the two sides of the market is achieved.

The outline of this paper is as follows. Section 2 gives an overview of the previous literature. Section 3 formally presents our setting. Section 4 shows the main results, and Section 5 concludes.

2. Related Literature

Roth (1982) was the first paper to ask to what extent the matching procedures provide participants with incentives to report their ranking of the available alternatives honestly.¹² With a set of theorems, the author is able to show that there exists no stable matching procedure for which truthful revelation of preferences is a dominant strategy for all participants (impossibility

partment were over \$1800, and the costs of campus interview were over \$650 per candidate (Stock et al., 2000). Recent estimates of the campus visit costs are up to \$2500, depending on where the campus is located (Claire Potter, “The Job Market Is A Lot Like The PBS News Hour, And Other Advice For Skype Interviews”, Tenured Radical (Blog), November 28, 2010, (11:16 pm), <http://chronicle.com/blognetwork/tenuredradical/2010/11/why-job-market-is-lot-like-pbs-newshour/>).

¹²The matching procedure for which any agent can do no better than state his true preferences is called *strategy-proof*.

theorem). Further, if there exists a stable matching procedure which delivers an optimal matching outcome for *one side* of the market, truthful revelation of preferences is the dominant strategy for *that side* of the market. And also, the *other side* of the market does not have incentives to misrepresent their most preferred alternative, but might benefit from misrepresenting the other alternatives. As an example, consider the men-proposing Gale-Shapley algorithm. This mechanism will always deliver stable optimal allocation for men. According to Roth (1982), it is a dominant strategy for every man to reveal his true preferences, and it is not, in general, a dominant strategy for every woman to state her true preferences, except for her most preferred alternative.

To summarize, the *impossibility theorem* states that there does not exist a stable matching procedure which never gives any participant incentives to misrepresent his/her choice $\#k$, where $k \neq 1$. This means that the Gale and Shapley matching algorithm is not strategy-proof for the accepting side of the market.

Despite the fact that honest representation of the preferences is not a dominant strategy, the literature finds it very complicated to provide market participants with advice on what the optimal misrepresentation strategy is. The paper by Roth and Rothblum (1999) considers the problem of which optimal strategies can be suggested to the market participants. The authors show that, in order to determine correctly the set of profitable deviations from stating preferences truthfully, participants have to know the preferences of both sides of the market perfectly. This amount of information is very rarely available to the participants. However, authors are also able to show that, in situations in which a participant has no information about the preferences of the others, it is still possible to find profitable deviation strategies (deviation from truthfully stating their preferences). These deviation strategies are very special, because they require a shortening of the list of the preferred alternatives without changing the order of those alternatives which remain on the shorter list, and are called *truncation strategies*.¹³ For the total number

¹³Let us consider the participant on the accepting side of the market. An important requirement for the optimal truncation strategy of a person is that his preferences are symmetric with respect to the proposing side of the market: each participant on the accepting side has strict preferences over the alternatives $i \in 1..N$, but he cannot say which of the participants on the proposing side of the market is likely to prefer him more, and also which of those alternatives $i \in 1..N$ is preferred by the participants on the

of alternatives N on the list, this truncation strategy requires the listing of only the top p alternatives. In the context of matching Ph.D. graduates with academic positions, this corresponds to demand of the Ph.D. graduates to send applications to p positions out of N listed job positions. Therefore, we will refer to *truncation strategies* as the strategies according to which the graduates are sending $p < N$ applications. As a result, each graduate can receive offers from, at most, p universities.

The intuition regarding why the truncation of the preference list might be beneficial is the following: submitting a shortened preference list decreases the probability of being matched with a less favored alternative. In fact, once the less favorable alternative is removed from the list, there is zero probability of being matched with it. But simultaneously, submitting a shorter preference list increases the risk of being unmatched. As a result, participants are likely to use truncation strategies which will increase their probability of being matched. Balancing these two effects will give the optimal truncation strategy. Roth and Rothblum (1999) propose the following approach to determine the gains from truncation strategies by the participant i :

- (1) compute the matching outcome for all possible preference profiles of other players (player i does not know the preferences of others, but can have some expectation about them);
- (2) compute in each matching outcome gains or losses compared to the situation with no truncation;
- (3) check for the expected gain, assuming that all realizations of preference profiles are, for example, equally likely from the point of view of player i .

This approach was used later in the paper by Coles and Shorrer (2013). In large markets, it is not possible to obtain matching outcomes for all the possible matching profile realizations, therefore authors assume that each participant has no information about the preferences of others, and calculate the expected utility gain from truncation by iterating Gale-Shapley algorithm 100,000 times and averaging the gains so that the “expected” gain is obtained. Moreover, the paper also finds *symmetric truncation equilibrium*: the more others truncate, the less an individual participant should truncate (optimal

accepting side of the market.

In general, uncorrelated preferences among the accepting side of the market, and correlated preferences among the proposing side of the market satisfy the requirement of symmetric preferences.

response), and in this way the equilibrium truncation point is determined numerically. To illustrate this for the market size $N = 30$ and uncorrelated preferences, the *symmetric truncation equilibrium* requires about a 50% drop of the preference list by every market participant on the accepting side of the market.

A new direction in the matching literature considers the idea that applying itself could be a signal of preferences. This approach has a similar motivation to ours: matching markets are too congested and market participants would benefit if they could credibly signal their preferences over the alternatives. Signaling of preferences was studied in few recent theoretical works (Coles et al., 2013b; Lee and Schwarz, 2007; Avery and Levin, 2010). In those models, the participants of the accepting side of the matching market have incentives to signal their preferences over the alternatives and this signaling increases the welfare of the accepting side of the market and can increase the total welfare. It is possible to draw some parallels between signaling of preferences and preference truncation by the participants on the accepting side of the market. If participants submit truncated lists of preferences, the proposing side of the market will infer that the alternatives which remain in the truncated list are probably of higher value for the accepting side participants, and those which are dropped, are valued less.¹⁴

As mentioned above, theoretically, searching for the optimal strategies in the matching markets is very complicated. Therefore, substantial research was developed in the direction of showing what the optimal deviations from stating preferences truthfully look like in the *experimental setup*, or using *real matching data*. Both experimental and empirical data suggest that, whenever the participants are restricted from submitting their full list of preferences, there is a manipulation of preferences, in the sense that the top p choices are not included in the list.

Calsamiglia et al. (2010) find that Gale-Shapley mechanism is not superior to the Boston mechanisms in cases of constraints. The reason is that individuals behave strategically, listing schools in which they are most likely to be accepted, and not necessarily the ones which are ranked highest. Klijn et al. (2013) consider the school choices with restricted number of submissions under Gale-Shapley and Boston mechanisms, looking also at risk aversion and

¹⁴Given that the equilibrium truncation strategy requires the participants to submit their p top ranked alternatives truthfully.

preference intensities. They find that high risk averse individuals play a protective strategy particularly in the Gale-Shapley constrained mechanism.¹⁵ Once the participants were required to truncate their preference list, the proportion of the cases in which the preferences were manipulated was higher than 75%. Chen and Sonmez (2006) conduct an experiment in which a certain number of students with heterogeneous preferences have to be allocated to schools. The experiment shows that not all students reveal their true preferences but act strategically even in a setting with no restrictions regarding the length of the preferences list. Other matching mechanisms have been proved to be more prone to manipulation. For example, Pathak and Sonmez (2013) compare different matching mechanisms based on the extension to which those are manipulable.

One of the papers which uses the data from real matching markets is Haeringer et al. (2010) who investigate the French academic job market of Ph.D.s in mathematics. In this market, candidates are shortlisted and interviewed, then the recruiting committees can rank only five candidates. Interestingly, they find that the matching quality in the non-competitive market depends on the degree of competitiveness across departments in a given year. Using the data from the centralized German system for the allocation of university places (former ZVS), Braun et al. (2010) show that agents do not state their true preferences but rather decide to submit a ranking which would lead to their individual best outcome. The authors were able to illustrate this on empirical data (via indirect measures for strategic behavior) for the application and acceptance process. This allocation of university places constitutes a priority matching game in which it is known that revealing true preferences is not a dominant strategy. Students are not allowed to state more than 6 preferences, and, since the ZVS reveals statistics about the acceptance process, the students can thus learn about their chances and behave accordingly.

In this paper, we build on the previous works that consider preference list truncation to improve match efficiency in Gale-Shapley algorithm (Roth and Rothblum, 1999; Coles and Shorrer, 2013). Our analysis contributes to the existing literature by quantifying this match improvement, with regard to different structures of preferences and the level of truncation. Our approach

¹⁵Participants had 3 alternatives, but were constrained to submit the ranking which includes only 2 alternatives.

is different from Coles and Shorrer (2013), we characterize and evaluate the optimal level of exogenous preference truncation level. In addition, we address the problem of strategic preference manipulation under this restricted preference reporting mechanism. Even though we are not able to solve for the equilibrium best response strategy, we describe the broad class of manipulation strategies, which lead to further improvement of the match quality under preference truncation. We also quantify the effect of applying manipulation strategies by participants.

3. Background

Let the matching game be defined by $\{G, U, q, w_g, w_u\}$, where G is the set of Ph.D. graduates and U is a set of universities. It holds $|G| = |U| = N$.¹⁶ Thus, we have both N Ph.D. graduates, g_1, g_2, \dots, g_N and N universities, u_1, u_2, \dots, u_N . q is the capacity vector which describes the number of positions each university has to fill. In our case, we consider $q = (1, \dots, 1)'$, $q \in 1 \times N$. Thus, every university has 1 position to allocate. The game can easily be generalized to $|q_l| \geq 1$, $l = 1, \dots, N$.

Let w_g be the matrix which contains the preferences of all Ph.D. graduates:

$$w_g := \begin{pmatrix} u_{g_1,1} & u_{g_1,2} & \dots & u_{g_1,N} \\ u_{g_2,1} & u_{g_2,2} & \dots & u_{g_2,N} \\ \dots & \dots & \dots & \dots \\ u_{g_N,1} & u_{g_N,2} & \dots & u_{g_N,N} \end{pmatrix} \quad (1)$$

Thus, the i -th row states the preferences of Ph.D. graduate g_i given by $u_{g_i} : (u_{g_i,1}, u_{g_i,2}, \dots, u_{g_i,N})$, which consists of the N universities $u_{g_i,j}$, $j = 1, \dots, N$, ordered according to his strict preferences, such that $w_{g_i}(u_{g_i,1}) > w_{g_i}(u_{g_i,2}) > \dots > w_{g_i}(u_{g_i,N})$, where $w_{g_i}(u_{g_i,j})$ is the utility the graduate g_i obtains from being matched with university position $u_{g_i,j}$.

For example, assume that $N = 3$ and that Ph.D. graduate g_1 has the preference list $u_{g_1} : (u_2, u_3, u_1)$, such that $w_{g_1}(u_2) > w_{g_1}(u_3) > w_{g_1}(u_1)$. It means that the utility for graduate g_1 is the highest when he is matched with university u_2 , and he will gain the third highest utility from being matched with university u_1 . We also assume that graduates always rank the possibility

¹⁶We restrict our simulations to this case for now, but will relax this assumption in future simulations.

of being unmatched as the worst possible outcome. We say that a Ph.D. is unmatched, when he is matched with university u_0 .

The preferences of the universities w_u over the set of Ph.D. students are defined accordingly:

$$w_u := \begin{pmatrix} g_{u_1,1} & g_{u_1,2} & \cdots & g_{u_1,N} \\ g_{u_2,1} & g_{u_2,2} & \cdots & g_{u_2,N} \\ \cdots & \cdots & \cdots & \cdots \\ g_{u_N,1} & g_{u_N,2} & \cdots & g_{u_N,N} \end{pmatrix}$$

The preferences of university u_j are given in the j -th row. The expression g_0 will define that the university could not fill its position (is unmatched).

To generate the matrix w_g of potentially correlated preferences among the graduates, we adopt a method initially suggested by Caldarelli and Capocci (2001). Thus, the preference of each Ph.D. graduate for a university u_j is composed of an individual opinion, which can differ from graduate to graduate, and a common opinion by all graduates. In this way, we allow preferences to be, in part, individual, for example, if a graduate prefers a certain university because it is close to his home town. The common part of preferences is induced by objective measures (for example, university rankings), which are visible to all graduates. Formally, each graduate g_i assigns a score $S_{g_i u}$ to each university, according to the following equation:

$$S_{g_i u} = \varepsilon_{g_i u} + V_g J_u \tag{2}$$

Both the entries of the vector $\varepsilon_{g_i u}$ and the entries of vector J_u are uniformly drawn from $[0, 1]$. Vector $\varepsilon_{g_i u}$ represents the individual opinion, and vector J_u the common view of the graduates on universities. Thus, J_u remains the same in all the scores for graduates. The constant parameter V_g , $V_g \in [0, \infty)$, measures the level of correlation of the graduates' preferences. The larger V_g , the more alike the preferences of graduates, where V_g defines no correlation.

In the same way, we generate the vector of preferences of each university u_j over the set of graduates:

$$S_{u_j g} = \varepsilon_{u_j g} + V_u J_g \tag{3}$$

where we allow the correlation parameters V_g and V_u to differ.

Note that we need at least cardinal-scaled preferences as our measure for inefficiency relies on the fact that we can interpret the distance between the

most preferred choice and the actual match. Thus, ordinal-scaled preferences would not be enough. Moreover, we will later exploit the special characteristic of this score - the ability to distinguish between individual and common opinion - by proposing a smart way of optimal response to truncation.

To illustrate the calculation of the score $S_{g_i u}$ with an example, consider the case of two university positions u_1 and u_2 , and two Ph.D. graduates, g_1 and g_2 . Let $J_u = (0.2, 0.3)'$, which means that the joint preference of the graduates towards university u_1 is 0.2, and 0.3 towards university u_2 . The individual preference of Ph.D. graduate g_1 is $\varepsilon_{g_1 u} = (0.5, 0.1)'$ and of Ph.D. graduate g_2 is $\varepsilon_{g_2 u} = (0.1, 0.4)'$. Then, $S_{g_1 u} = (0.5, 0.1)' + V_g(0.2, 0.3)'$ and $S_{g_2 u} = (0.1, 0.4)' + V_g(0.2, 0.3)'$. For $V_g = 2$, we would have $S_{g_1 u} = (0.9, 0.7)'$ and $S_{g_2 u} = (0.5, 1)'$, and the preference lists would be: $u_{g_1} : (u_1, u_2)$ and $u_{g_2} : (u_2, u_1)$. And the preference matrix of universities is:

$$w_u := \begin{pmatrix} u_1 & u_2 \\ u_2 & u_1 \end{pmatrix}$$

The outcome of the matching algorithm is the pairing of Ph.D. graduates and universities. We define this pairing by the mapping $m : G \rightarrow U$. For every Ph.D. graduate g_i , this mapping m will define the university u_j he is matched with. If Ph.D. graduate g_1 is matched with university position u_3 , we would have $m(g_1) = u_3$. If a Ph.D. graduate g_1 is unmatched, we have $m(g_1) = u_0$.

We can also define the matching from the university perspective. Then, we have a mapping $n : U \rightarrow G$. For this case, we have $n(u_3) = g_1$, when university u_3 is matched with g_1 . In the case that the university could not fill its position, we have $n(u_3) = g_0$.

It will be useful to define the set of *stable* matchings. The matching outcome $m(G) \subset U$ is *stable* if there is no *blocking pair*. A *blocking pair* is a graduate g_i and university u_j , not currently matched to one another, who would prefer to be matched with each other, rather than being matched with their current matches in $m(G) \subset U$ (Roth and Rothblum, 1999).¹⁷ In general, there can exist more than one stable matching outcome for the given preferences w_g and w_u . The set of stable matching outcomes is also called the *core*. Further, we define a university u_j as achievable for a Ph.D. g_i , if there exists a stable matching in which $m(g_i) = u_j$ (Coles and Shorrer, 2013).

¹⁷Note, that the ‘‘current match’’ could be also $m(g_i) = u_0$, and, in this case, graduate g_i prefers to be matched to u_j rather than be unmatched.

3.1. Gale-Shapley mechanism

The most popular mechanism which produces stable matches is the Gale-Shapley mechanism (also called the Deferred-Acceptance algorithm) (Gale and Shapley, 1962). The algorithm finds a stable match between two sets of elements (usually referred to as “men” and “women”), which have certain preferences over the elements in the opposite set.

Suppose that graduates and universities submit their preferences to a social planner. Matching outcomes m are obtained by using a “centralized” authority, where participants submit their preferences $\{w_g, w_u\}$, and the centralized authority applies the Deferred-Acceptance algorithm to determine the matches. The algorithm is stated as if the market was decentralized: universities are making offers and graduates are accepting positions according to their preferences, but they are not submitting those preferences to the centralized authority. This emphasizes that the Deferred-Acceptance algorithm resembles a decentralized market for obtaining matches.

The algorithm works as following:

At the beginning, all graduates and universities are not matched (“free”).

Round 1. From the amount of applications it has received, each university u_j proposes to its favorite graduate g_i . Each graduate g_i retains his most favorite university position (from the universities which made him an offer), if any, and rejects all other positions which have proposed to him (in the case that he receives more than one proposal).

Round 2. All rejected universities in the first round propose to their second best option. Then, each graduate chooses from the pool of new proposers and the position he keeps from Round 1. All other offers are rejected.

The algorithm stops when there are no more proposals left.

Gale and Shapley (1962) demonstrate formally that this mechanism leads to a stable matching outcome for any preference profiles.

3.2. Inefficiency of the matching outcome

The inefficiency of the matching outcome $m(G)$ is measured by a function of the distance of the realized match to the most preferred choice, where we also account for the specific position preference of the participants.¹⁸ The

¹⁸A similar specification has been used in the previous literature (e.g., Coles and Shorrer, 2013).

inefficiency for graduate g_i , $i = 1, \dots, N$, resulting from his matching outcome $m(g_i)$ is given by:

$$\vartheta(g_i) = \begin{cases} 1, & \text{if } m(g_i) = u_0 \\ \left(\frac{rk_g(m(g_i)) - 1}{N} \right)^{a_g}, & \text{otherwise} \end{cases} \quad (4)$$

with $\vartheta : G \rightarrow \mathbb{R}$.

The function $rk_g : G \rightarrow \mathbb{R}$ is a function which returns the rank. Rank is the number which refers to the "position" of the match $m(g_i) = u_j$ in the vector u_{g_i} . For example, if graduate g_i is matched with his third best choice u_{j_3} , $rk_g(m(g_i)) = 3$. Thus, the more his match deviates from his most preferred choice, the higher his inefficiency $\vartheta(g_i)$ is. Precisely, the inefficiency of the match increases monotonically by $(1/N)^{a_g}$. Being unmatched is treated as being matched with your $N + 1$ preferred option. The parameter a_g characterizes the curvature of the inefficiency function. We can refer to this inefficiency measure $\vartheta(g_i)$ as "dis-utility" of the match for the graduate g_i . Then, if $a_g > 1$, graduate g_i has increasing marginal "dis-utility" of the match. Figure 1 below plots this inefficiency function for parameter $a_g > 1$. As we can see, moving from 6th most preferred choice to 7th most preferred choice increases inefficiency of the match much more than moving from choice ranked 5 to choice ranked 6. Interpretation is the following: for $a_g > 1$ graduates value the possibility of obtaining a better match less than the loss of efficiency if they obtain a worse match. The situation for $a_g < 1$ is the opposite: graduates value the possibility of obtaining a better match more than the loss of efficiency if they obtain a worse match.

Similarly, the inefficiency of the universities is defined by:

$$\eta(u_i) = \begin{cases} 1, & \text{if } n(u_j) = g_0 \\ \left(\frac{rk_u(m(u_j)) - 1}{N} \right)^{a_u}, & \text{otherwise} \end{cases} \quad (5)$$

with $\eta : U \rightarrow \mathbb{R}$. The function $rk_u : U \rightarrow \mathbb{R}$ gives the rank of the student to which the university is matched. We allow a_g and a_u to differ from each other.

4. Results

4.1. Exogenously truncated preferences and match inefficiency - theoretical results

Using the notation, for $p \in (0, N]$, we define the preference list w_g^p which includes only g_i 's p most preferred university positions (ordered according

to the true preferences w_g). We call this p -truncation of the graduates' true preferences. Therefore, w_g^p is the part of the matrix w_g , which included only first p columns. We set level p as exogenously given, imposed by the central planner. In other words, every graduate is constrained to apply to a maximum of p positions. At this stage, we assume that the graduates apply to their top p preferred positions, and estimate the aggregate inefficiency from matching outcomes.

Later, we will discuss the possibility of strategic behavior in this framework and will show that the strategic responses of the graduates can only decrease the inefficiency level compared to the setup with no strategic responses. In other words, the optimal level of truncation remains robust to the strategic responses of the graduates. Moreover, this is the most conservative truncation level, in the sense that it provides optimal truncation level under the "worst-case" scenario.

Building on the literature of Roth and Rothblum (1999) and Coles and Shorrer (2013), truncation can only be profitable strategy for the "accepting" side (in our case, the graduates) of the market. The benefit derives from the increased probability of being matched with a more preferred university. However, submitting a shorter preference list increases the risk of being unmatched.

Below, we will demonstrate with a simple example how truncation of the individual preference list can make some participants better off. Consider a matching game with $N = 3$. The table below presents the preferences of graduates and universities:¹⁹

Preferences and matching outcomes (without truncation)

$$\begin{array}{l} w(g_1) = \{\mathbf{u}_1 \quad u_3 \quad u_2\} \\ w(g_2) = \{u_2 \quad u_1 \quad \mathbf{u}_3\} \\ w(g_3) = \{u_1 \quad u_3 \quad \mathbf{u}_2\} \end{array} \left| \begin{array}{l} w(u_1) = \{\mathbf{g}_1 \quad g_3 \quad g_2\} \\ w(u_2) = \{\mathbf{g}_3 \quad g_1 \quad g_2\} \\ w(u_3) = \{g_1 \quad \mathbf{g}_2 \quad g_3\} \end{array} \right.$$

Preferences and matching outcomes (with truncation)

$$\begin{array}{l} w(g_1) = \{\mathbf{u}_1 \quad u_3 \quad u_2\} \\ w(g_2) = \{\mathbf{u}_2 \quad u_1 \quad \quad\} \\ w(g_3) = \{u_1 \quad \mathbf{u}_3 \quad u_2\} \end{array} \left| \begin{array}{l} w(u_1) = \{\mathbf{g}_1 \quad g_3 \quad g_2\} \\ w(u_2) = \{g_3 \quad g_1 \quad \mathbf{g}_2\} \\ w(u_3) = \{g_1 \quad g_2 \quad \mathbf{g}_3\} \end{array} \right.$$

¹⁹Bold symbols in the matrix refer to the matching outcome of the Gale-Shapley algorithm.

Using the Gale-Shapley matching algorithm, we obtain matching outcome $\{m(g_1) = u_1; m(g_2) = u_3; m(g_3) = u_2\}$. Graduate g_1 is matched with his most preferred university. The other two graduates are matched with their least preferred universities. If graduate g_2 decides to truncate his preference list to include only his two most preferred positions, such that $w_{g_2} = \{u_2, u_1\}$, the resulting matching outcome is: $\{m(g_1) = u_1; m(g_2) = u_2; m(g_3) = u_3\}$. Note that not only does graduate g_2 benefit from the truncation (he is now matched with his most preferred university), but also graduate g_3 . Also, universities are paired with worse options than before the truncation.²⁰

In the next example, we change the preferences such that graduate g_3 now prefers university u_2 over university u_3 . The matching outcome is $\{m(g_1) = u_1; m(g_2) = u_3; m(g_3) = u_2\}$. If g_2 excludes u_3 from his preference list, the resulting matching is $\{m(g_1) = u_1; m(g_2) = u_0; m(g_3) = u_2\}$. Therefore, as a result of truncation graduate g_2 is unmatched.

Preferences and matching outcomes (without truncation)

$$\begin{array}{l} w(g_1) = \{\mathbf{u}_1 \quad u_3 \quad u_2\} \\ w(g_2) = \{u_2 \quad u_1 \quad \mathbf{u}_3\} \\ w(g_3) = \{u_1 \quad \mathbf{u}_2 \quad u_3\} \end{array} \left| \begin{array}{l} w(u_1) = \{\mathbf{g}_1 \quad g_3 \quad g_2\} \\ w(u_2) = \{\mathbf{g}_3 \quad g_1 \quad g_2\} \\ w(u_3) = \{g_1 \quad \mathbf{g}_2 \quad g_3\} \end{array} \right.$$

Preferences and matching outcomes (with truncation)

$$\begin{array}{l} w(g_1) = \{\mathbf{u}_1 \quad u_3 \quad u_2\} \\ w(g_2) = \{u_2 \quad u_1 \quad \quad\} \\ w(g_3) = \{u_1 \quad \mathbf{u}_2 \quad u_3\} \end{array} \left| \begin{array}{l} w(u_1) = \{\mathbf{g}_1 \quad g_3 \quad g_2\} \\ w(u_2) = \{\mathbf{g}_3 \quad g_1 \quad g_2\} \\ w(u_3) = \{g_1 \quad g_2 \quad g_3\} \end{array} \right.$$

Coles and Shorrer (2013) define symmetric equilibrium in truncation level, such that there exists a symmetric equilibrium level \tilde{p} that every man reports true preferences and every woman uses a symmetric truncation strategy $\sigma(\tilde{p})$, in other words, all submit a list of the length \tilde{p} (of true preferences). In contrast to Coles and Shorrer (2013), our approach will be to consider efficiency gains and losses from simultaneous preference list truncation by all participants on the “accepting” side of this market (the graduates). That is, we calculate the *ex-ante* (expected) inefficiency of the matching outcomes (calculated via the function ϑ and η). The overall market inefficiency is given

²⁰This is true because the initial matching outcome (without truncation) was optimal for the universities, in the sense that any other stable matching outcome will make at least one of the universities worse off.

by the sum over the inefficiencies of all market participants. We investigate the aggregate match inefficiency for different truncation levels p . We are able to show numerically that there exists a level of truncation p^* which delivers the lowest inefficiency for all the participants on the “accepting” side of the market as:

$$p^* = \arg \min_p \mathbf{E}_{w_g, w_u} \left(\sum_{i=1, \dots, N} \vartheta_{g_i}(m[w_{g_i}^p, w_{g-i}^p, w_u](g_i)) \right), \quad (6)$$

where $m[w_{g_i}^p, w_{g-i}^p, w_u](g_i)$ is the matching outcome for the graduate g_i given the other graduates’ submitted truncated lists w_{g-i}^p , and university preferences w_u .

Theorem 1. *There is no p^* such that $p^* > \tilde{p}$.*

Proof. Suppose there is a p^* that is minimizing the total inefficiency of the matching outcome of the graduates, and that $p^* > \tilde{p}$. Then, if all graduates are truncating in equilibrium at \tilde{p} , the optimal response of at least one graduate g_j would be to choose a longer preference list, and to send $p > \tilde{p}$ applications. But then \tilde{p} is not an equilibrium truncation level. By contradiction, it cannot be that $p^* > \tilde{p}$. ■

Below we show with numerical simulations that p^* is lower than the symmetric equilibrium truncation level described in Coles and Shorrer (2013). However, with this truncation level p^* , the participants of the “accepting” side of the market are free to choose to which positions to apply. This means that the optimal strategy of the graduates is not necessarily to include top p choices in the preference list. We will return to the issue of optimal responses of the graduates below.

4.2. Exogenously truncated preferences and match inefficiency - numerical simulations

For the numerical simulation, we use the Gale-Shapley algorithm, where universities propose to graduates. There are 5 model parameters, which can be changed. We assume that our market consists of 100 universities and 100 graduates. Each university has 1 position to be filled. The parameters V_g and V_u define the degree of correlation in the preferences. At the beginning, we assume $V_g = V_u = 0$, which means that both universities and graduates preferences are uncorrelated. The parameters for the inefficiency function are

set to $a_g = a_u = 1$. Consequently, they treat the step from being matched with their second choice to being matched with their third choice in the same way as, for example, the step from being matched with the least preferred university/graduate to being unmatched.

We consider all possible truncation levels, ranging from 1 (applying to most preferred position only) to 100 (applying to all positions, no truncation). For now, we assume that students reveal their preferences truthfully, in the sense that they apply to top p most preferred positions.

As preferences are randomly chosen, we cannot simulate the matching procedure just once. In order to have meaningful results, we repeat each simulation 1,000 times and average the obtained results for match inefficiency.

The Gale-Shapley algorithm is modified in the following way: at the beginning, all graduates and universities are unmatched (“free”). Graduates retain only p most preferred universities in their preference list. This is the same as if graduates were applying to their p most preferred universities only. Universities can still rank all graduates, but can make offers only to the graduates who have applied to them. Therefore, phrase “apply to university position u_j ” used in this paper characterizes the situation when the graduate retains university u_j in his (truncated) preference list.

Round 1. Each university u_j offers to the most preferred graduate g_i , which has applied to it. Each graduate g_i accepts his most favorite university position (from the universities which have made him an offer), if any, and rejects all other positions which have proposed to him (in the case that he receives more than one proposal).

Round 2. All rejected universities in the first round propose to their second best option. Then, each graduate chooses from the pool of new proposers and the position he keeps from Round 1. All other offers are rejected.

The algorithm ends, when there are no more universities which can still make offers to some graduates.

Figure 2 shows the obtained inefficiency, separately for the proposing side (universities) and accepting side (graduates) for the different level of truncation p . When the graduates submit their full lists of preferences ($p = 100$), all positions are filled (Gale-Shapley algorithm matches all participants, in case of submitting full preference lists). Nevertheless, the inefficiency for graduates (blue bars) is at its highest level: this is because some graduates have a rather poor match. In fact, as soon as they start to apply to less than N positions, they benefit: the inefficiency decreases monotonically up to a truncation point of around 16 (meaning that they have applied to their

16 most preferred universities). However, as soon as they submit even less applications, inefficiency increases sharply. The reason for this increase can be seen from Figure 4 ($V = 0$, which mean that $V_g = V_u = 0$): severe truncation results in a high number of unmatched participants. In fact, the number of non-filled position rises up to 23 (for total of 100 positions) for $p = 1$.

$p = 1$ implies that students apply just to their most preferred university and many universities will not receive any applications. In turn, many students compete for the same position. Nevertheless, we also see from Figure 4 that the problem of unfilled positions is not that severe for the optimal truncation point of 16, just 2 graduates are unmatched.²¹

To summarize, graduates face a trade-off: they gain from limiting the number of applications by obtaining a better match. But with a lower number of applications (more truncation), they face a higher risk of being unmatched. The optimal number of applications, p^* , is substantially lower than $p = N$, where the latter refers to applying to all universities.

Regarding the universities, it is clear that they cannot benefit from graduates' truncation, as less graduates apply to them. Thus, there will always be some universities which are either unmatched or are matched to a less preferred graduate, when students truncate their list. For the proposing side, the best solution is always to have access to the full list of preferences (see Roth, 1982; Roth and Rothblum, 1999).

We then consider the case in which $M < N$, as a prevailing feature of many real-world contexts, meaning that there are more candidates than available positions. Importantly, we find that decreasing the number of available positions M the optimal truncation level p^* decreases. However, when $M < N$ the aggregate inefficiency is higher than the case $M = N$, no matter what the number of applications is allowed. Also, when the number of available positions is substantially lower than the pool of candidates $M \ll N$, changing the truncation level p has almost no effect on the aggregate inefficiency level. Therefore, the decrease in the inefficiency for Ph.D. graduates under p^* is low; and in particular when $M = 1$, implementing any truncation

²¹Notice the importance of the utility function defined above. If we assign less inefficiency to the unmatched outcomes (now, we set the inefficiency to 1), the optimal truncation point in terms of inefficiency moves to the left. When we assign more inefficiency to the unmatched outcomes, the optimum moves to the right. We will come back to this point relating to the form of the utility function later.

level does not affect the final outcome.

Next, we study the effect of preference correlation. High correlation means that students agree with respect to which are the best universities. Consequently, under high correlation (with large V), many students will apply to few universities and many universities will prefer few students, respectively. Intuitively, exogenous truncation of the preference list will be more costly now, because graduates are more likely to retain the same universities in their truncated preference lists, competition for the positions will be higher, and graduates are more likely to remain unmatched.

Figure 3 shows the findings for different V ($V_g = V_u = V$) and a . Let us first consider the case where $a = a_g = a_u = 1$ (middle figure). As expected, inefficiency is the lowest under no preference correlation ($V = 0$). Indeed, in this case, graduates and universities are likely to differ substantially in their preferences. Thus, many students will receive an offer from their most preferred university. By increasing V , match inefficiency becomes higher. Regarding the optimal truncation point, we see that, for mild correlation $V = 0.5$, the social planner should choose a truncation point around 35. For even higher V , the optimal truncation point moves further to the right, where the inefficiency remains nearly constant from $p = 70$ ($V = 1.5$) and is at its lowest for $p = 100$ (submit full ranking), when $V = 5$. The reason is that graduates are likely to end up unmatched under extreme correlation. This can be seen also from Figure 4, which states the number of unfilled positions depending on p . When V is high, the number of unfilled positions is positive already starting from $p = 99$ and ends with around 90 unfilled positions, when $p = 1$. We also see that the unfilled positions arise nearly exclusively from missing applications and not from the fact that all graduates reject the offer from the university to which they have applied. For $V = 0$ (no correlation), we have the opposite case: in addition to the fact that there are only a few unfilled positions, those unfilled positions are mainly the result of rejection by the applicant, and not the result of no applications.

To summarize, for fixed truncation point p and the increase in the correlation of preferences, the number of unfilled positions considerably increases, as an increasing number of graduates apply to the same few universities and many universities do not even receive an application. As a result, for highly correlated preferences, it is no longer optimal for the social planner to enforce a truncation in the number of applications.

The findings are also likely to change, when just the preferences of the universities are highly correlated (as all universities would like to hire highly

quality graduates, which can be revealed by objective measures, such as grades), but Ph.D graduates have less correlated preferences (as they also have some individual non observable preferences, i.e., have geographical preferences). For the case $V_g < V_u$, we find that, with increasing V_u , the benefits from truncating are lower.

So far we have considered the case with $a = 1$, which corresponds to the inefficiency level plotted in the middle of Figure 3. We will now consider the aggregate inefficiency of the realized matches for different levels of a . Figure 3 shows the findings. Consider first two graphs at the top. Graduates with $a < 1$ do not care so much whether they could end up unmatched. In fact, they value an increasing of the rank of their match far more. They will not benefit from less truncation, although this increases the probability of being matched. For correlated preferences $V \neq 0$, we have an optimal truncation point (interior). In fact, inefficiency remains at the same level, from a certain truncation point p . Note that the lowest inefficiency is not achieved for $p = 1$, as even with $a < 1$ graduates still value a match (even if not as high as having a good match). Graduates with $a > 1$ (the two graphs in the last row) value more the fact that they are matched, no matter whether the university they are matched with is highly preferred or not. Thus, it is usually the case that the lowest inefficiency is achieved under no truncation ($p = 100$), especially if preferences are correlated. However, for uncorrelated preferences, it is possible to find interior optimal truncation point p^* .

4.3. Strategic preference manipulation under exogenous truncation and match inefficiency

As was noted above, the deferred-acceptance algorithm is not strategy-proof for the accepting side of the market. Therefore, when a maximum number of application p is exogenously imposed to Ph.D. graduates such that $p < N$, they might respond by applying to universities which are not the most preferred. In case $V_g = 0$ and $V_u = 0$, this strategy will never be optimal (Roth and Rothblum, 1999). Yet, when graduates' preferences are correlated, meaning $V_g \neq 0$, a graduate i might increase his probability of being matched well (or being matched at all) by misrepresenting his true preferences. We restrict our attention to the case when $V_u = 0$, meaning that universities preferences over graduates are uncorrelated. Given our specification that a graduate's preferences identifies by his individual preferences ε_{g_iu} and a

common component J_u

$$S_{g_iu} = \varepsilon_{g_iu} + V_g \times J_u,$$

the only optimal strategy would be to “uncorrelate” his own preferences from the common preferences of all graduates up to a certain degree. This can be modeled by putting less weight on J_u (which is weighted by graduates with $V_g/(1 + V_g)$) and putting more weight on his individual preferences ε_{g_iu} . Formally,

$$S_{g_iu} = (1 + \alpha)\varepsilon_{g_iu} + (V_g - \alpha) \times J_u, \quad (7)$$

with $\alpha \in [0, V_g]$. In this way, the other graduates weight the individual part by $1/(1 + V_g)$ and the common part by $V_g/(1 + V_g)$, whereas our optimizing graduate i gives the weight $(1 + \alpha)/(1 + V_g)$ to his idiosyncratic (individual) part and $(V_g - \alpha)/(1 + V_g)$ to the common part. This is a strategy which decreases graduate i 's matching inefficiency because, although he is not telling the truth, he benefits from choosing the universities which are at least individually preferred and putting them onto his application list. Imagine, for example, a graduate who thinks he has no chance of receiving an offer from Harvard (a university we can assume to be preferred by many students). If he excludes Harvard from his list of p most preferred university positions (truncated list of preferences), he is likely to include instead the university position to which is at least individually highly preferred. This strategy is smart, as randomly “lying” could result in being matched with a university which is not preferred by our student at all. Our imposed strategy guarantees that the probability that a university, which is neither preferred individually nor commonly, ends up on his application list is quite low.

Intuitively, given a certain p and V_g , the optimal α_i will depend on: i) the realized preferences of individual i , ii) the realized preferences of all other graduates $G \setminus \{i\}$, and iii) the realized preferences of N universities. This means that, for each market participants' realization of preferences, there exists an α_i such that minimizes (reduces) his match inefficiency. Similarly, for all other graduates. For this reason, it is not possible to derive a unique α^* . Nonetheless, we can evaluate the overall inefficiency when individuals are not truth-telling under different truncation levels.

Based on these considerations, we proceed in the following way. We compare market inefficiency for different patterns of preference manipulation

by graduates. We will consider three cases where *all* graduates choose to submit a preference lists according to one of the expressions below:

$$(a) \quad S_{g_i u} = (1 + \alpha_0^*)\varepsilon_{g_i u} + (V_g - \alpha_0^*) \times J_u, \text{ for } \forall i = 1, \dots, N \quad (8)$$

$$(b) \quad S_{g_i u} = (1 + \alpha_r^*)\varepsilon_{g_i u} + (V_g - \alpha_r^*) \times J_u, \text{ for } \forall i = 1, \dots, N \quad (9)$$

$$(c) \quad S_{g_i u} = (1 + \alpha_{V_g}^*)\varepsilon_{g_i u} + (V_g - \alpha_{V_g}^*) \times J_u, \text{ for } \forall i = 1, \dots, N \quad (10)$$

where $\alpha_0^* = 0$, α_r^* is drawn randomly from the set $\{0, 0.1, 0.2, \dots, V_g\}$ and $\alpha_{V_g}^* = V_g$. We simulate matching outcomes for different truncation level p , and compute aggregate inefficiency for all graduates.

Case (a) with α_0^* corresponds to the situation where graduates do not act strategically, and apply to their most preferred p positions (act according to their true preference list); case (c) corresponds to the situation in which graduates apply to the p positions according to their individual preference component $\varepsilon_{g_i u}$ only, and case (b) is the intermediate case, where each graduate randomly chooses α_r^* , which affects the selection of the positions to which he applies.

Figure 6 shows the simulation results for different levels of preference correlation ($V_g = 0, V_g = 1$ and $V_g = 5$) and the three cases: (a), (b) and (c).

Consider the case with $V_g = 0$. In this situation, there is no space for strategic manipulation of the list of positions where graduates apply - graduates already have uncorrelated preferences. In this situation, optimal truncation level p^* is between 10 and 20.

Now, for the case where $V_g = 1$, there is some space for strategic manipulation. We can see that if graduates do not manipulate their preference list and apply to their p most preferred position, aggregate match inefficiency is the highest (blue line). On the other hand, if graduates apply to the top p positions according to their individual preference component $\varepsilon_{g_i u}$, aggregate match inefficiency decreases considerable (red line). This is an intuitive result, because if every graduate applies according to his individual preference component, there is no pattern in the applications received by the universities (applications are uniformly distributed among universities), and graduates face the lowest level of competition for the position. In this situation, we obtain the highest level of filled positions.

However, the match inefficiency for $V_g = 1$ is still higher than for $V_g = 0$, even if graduates apply according to their individual preference component.

This happens because true preferences under $V_g = 1$ are determined also by the common component: making application lists less correlated with other graduates' lists ensures that the graduate is more likely to be matched, but this match is still less efficient than the match with true most preferred position.

Results for the case with $V_g = 5$ are similar.

(1) To summarize, aggregate inefficiency level for the case in which graduates apply to their top p most preferred position provides the upper bound on the inefficiency level which would be achieved under any strategic manipulation of preferences by the graduates.²²

(2) In addition, the optimal truncation level p^* decreases if graduates manipulate their preferences. For example, for $V_g = 1$, under assumption that graduates do not manipulate their preferences at all (α_0^*), optimal truncation is $p^* = 60$, but under the assumption that graduates act only according to their individual component of the preferences ($\alpha_{V_g}^*$), truncation level should be set to $p^* = 20$. Therefore, optimal truncation level p^* determined by the aggregate match inefficiency under the assumption of no preference manipulation provides the most conservative truncation level. In fact, whenever the social planner thinks that graduates would, to some extent, manipulate their preferences, it is safe to choose even a lower truncation level $p < p^*$.

5. Conclusion

In the last decades, many real-world markets are crowded as never before, thus challenging the hiring mechanisms. This paper provides a rationale for imposing a limit on the number of possible applications in the matching markets for new Pd.D. graduates. Our results show that, if the social planner introduces the optimal level of truncation the quality of matches increases for Pd.D. graduates. From the point of view of universities, the quality of matches decreases but this loss is offset by the reduction in the hiring costs. Moreover, the proposed mechanism guarantees a more even distribution of inefficiency between the two sides of the market. While the optimal level of truncation is derived under specific parameters, by considering different

²²By “any strategic manipulation”, we mean any strategic manipulation of a certain type considered in this paper: remove most commonly preferred positions from the preference list and substitute them with positions which are individually most preferred.

Truncation in the Matching Markets and Market Inefficiency

features of the market our analysis can provide useful insights also in other contexts.

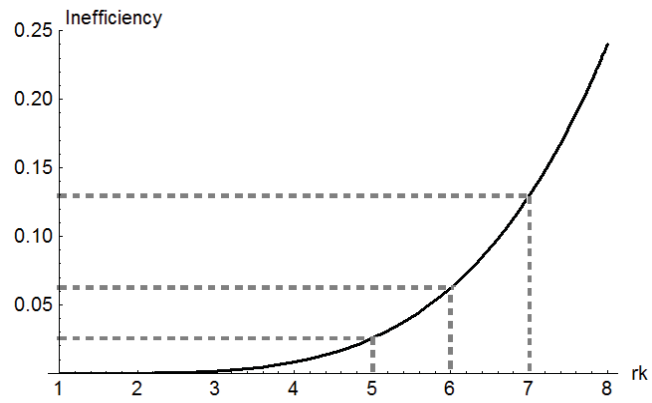


Figure 1: Plot of inefficiency measure ϑ , for $a > 1$. Horizontal axis shows rk , rank of the match; and vertical axis shows the value of inefficiency (dis-utility) for the match

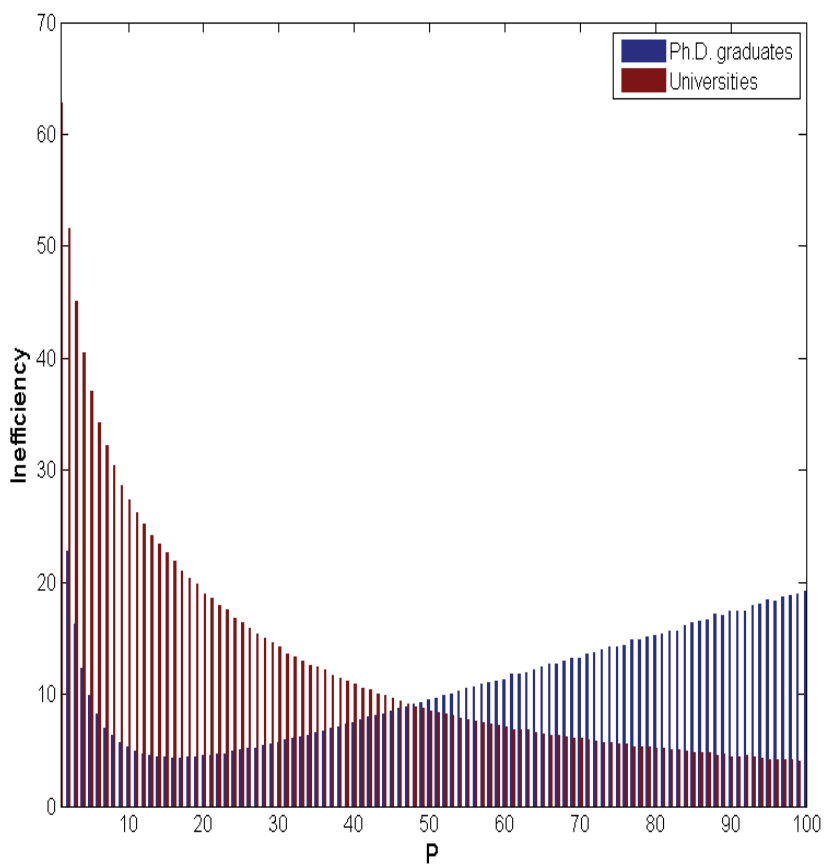


Figure 2: Matching Inefficiencies for Different Level of Truncation. This figure shows the aggregated (over all participants) inefficiency of the matching outcomes. We run 1000 simulations, and parameters are set as the following: $N = 100$, $V_u = V_g = 0$ and $a = 1$.

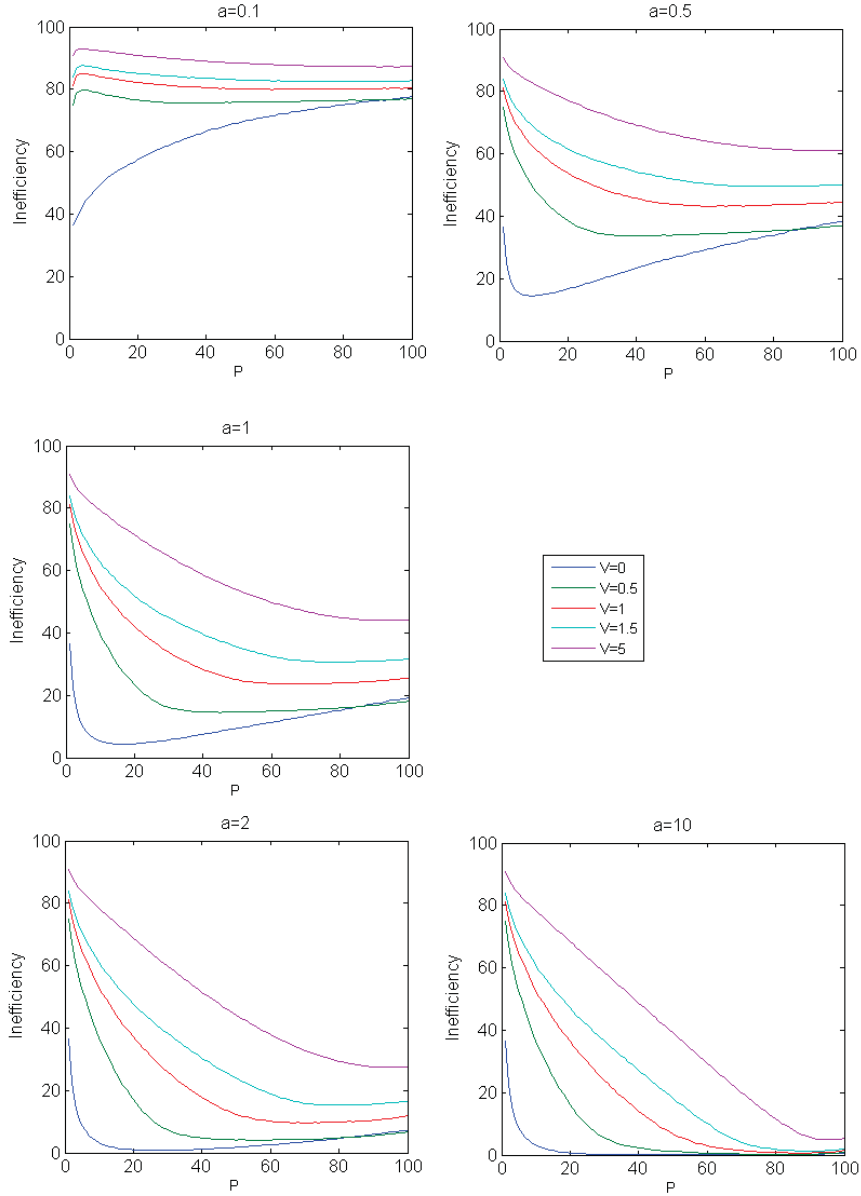


Figure 3: Matching Inefficiencies for Different Level Truncation, Preference Correlation and the Risk Attitude. This figure shows the aggregated (over all participants) inefficiency of the matching outcomes for the Ph.D. graduates. We run 1000 simulations, and parameters are set as the following: $N = 100$.

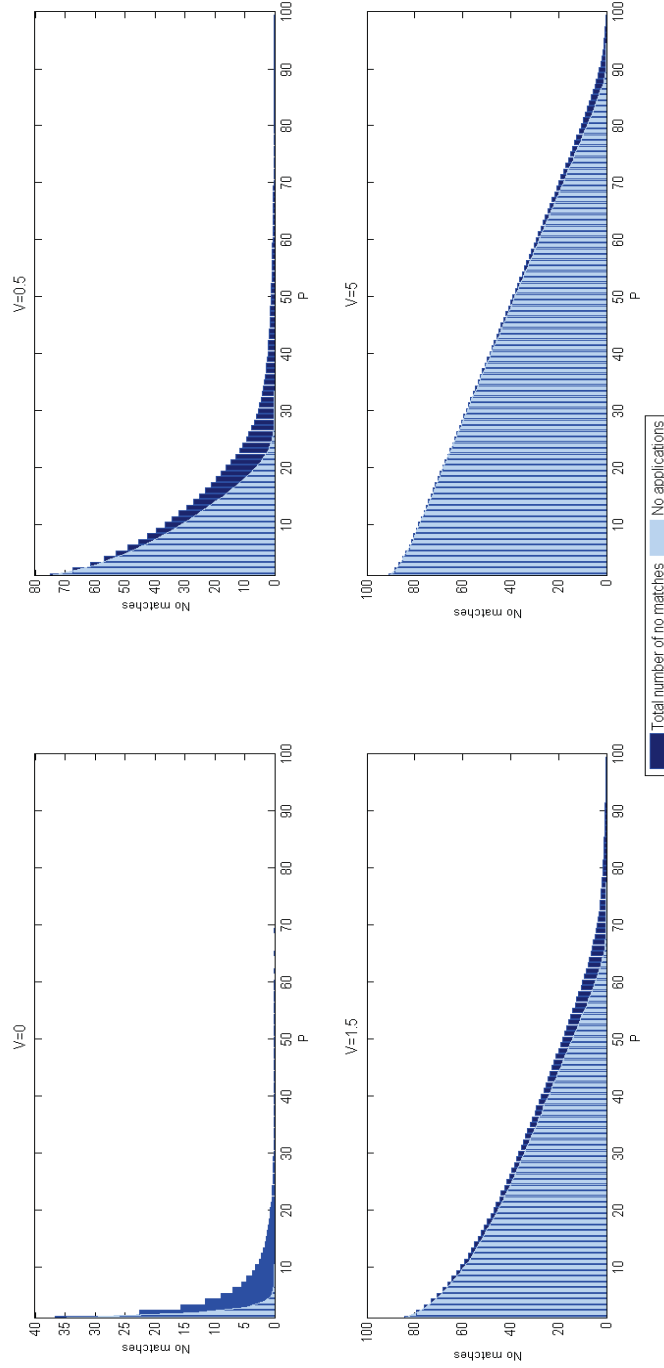


Figure 4: Total Number of No Matches, and the Number of No Matches Due to No Applications Received By the Universities. This figure shows the total number of non-matched universities (and graduates), for different level of truncation and preference correlation. We run 1000 simulations, and parameters are set as the following: $N = 100$, $\alpha = 1$.

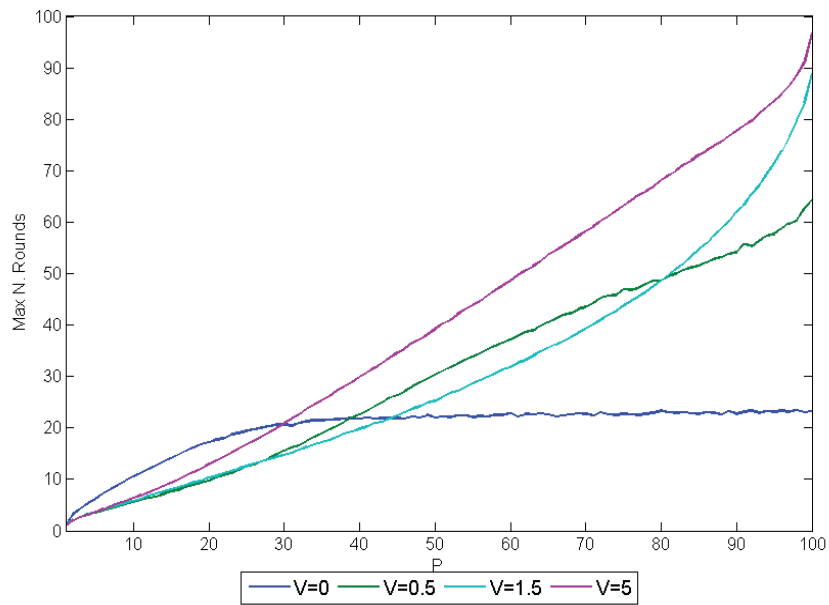


Figure 5: Number of Rounds. This figure shows the maximum number of rounds it takes to finalize the matching algorithm, for different truncation level and preference correlation. We run 1000 simulations, and parameters are set as the following: $N = 100$, $a = 1$.

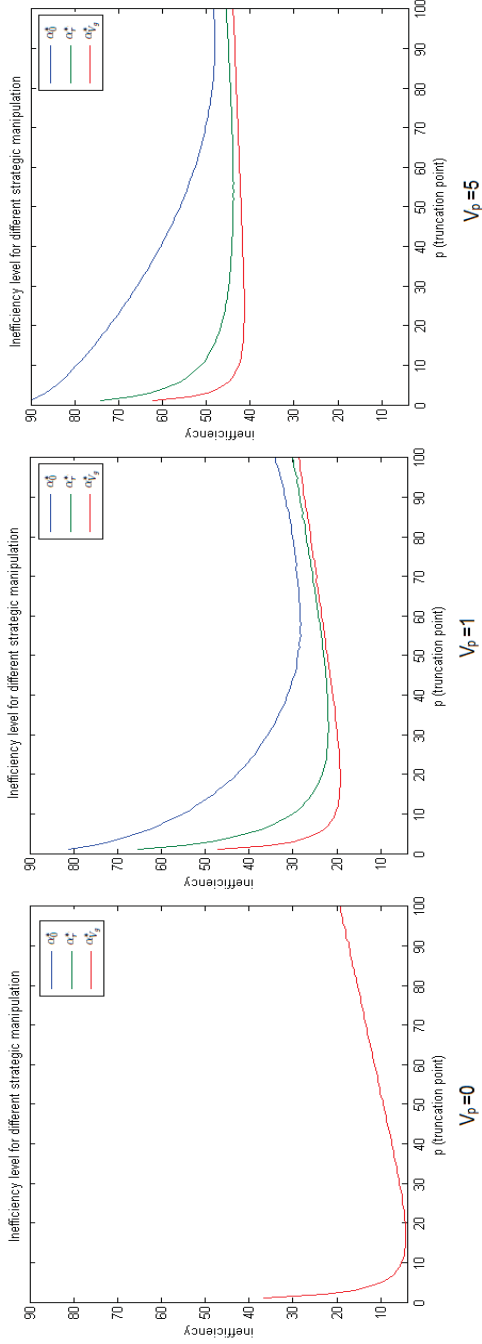


Figure 6: Matching Inefficiency. These figures show the matching inefficiency level for graduates (y -axis), for different truncation levels (x -axis), and degree of strategic manipulation α (blue line α_0^* , green line α_r^* , and red line α_g^*). We run 1000 simulations, and parameters are set as the following: $N = 100$, $V_g = 0$, $V_g = 1$, $V_g = 5$.

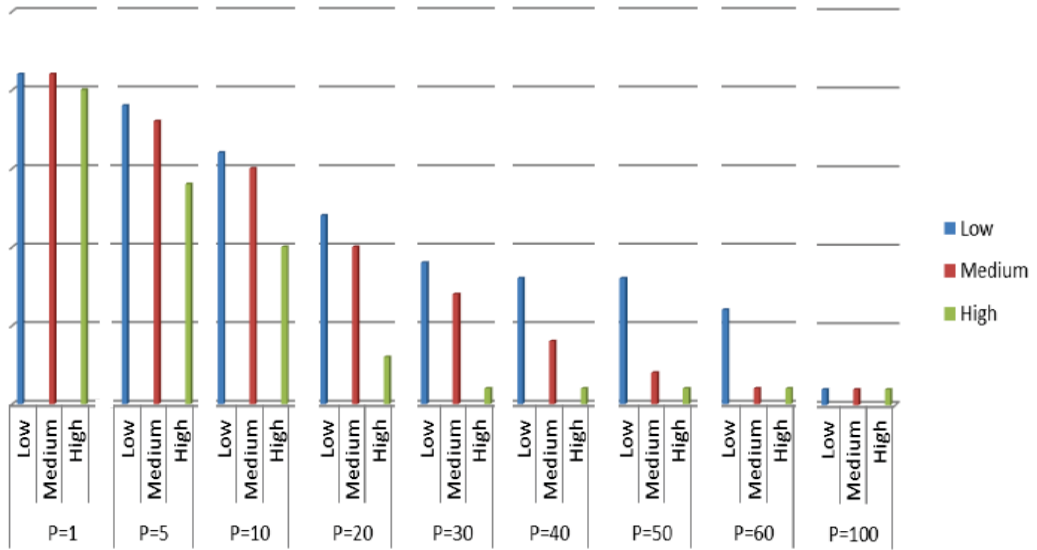


Figure 7: Optimal level of alpha. This figure shows the optimal level of alpha (y-axis), for low-type, medium-type, high-type and for some truncation levels ($P = 1, P = 5, P = 10, P = 20, P = 30, P = 40, P = 50, P = 60, P = 100$). We run 100 simulations, and parameters are set as the following: $N = 100, V_g = V_u = 1$.

Appendix A. Matching inefficiency and other comparative statics

The results of the present paper are based on the aggregate inefficiency both for Ph.D. graduates and universities, and inefficiency has been defined respectively in Equation 4 and Equation 5. This specification is consistent with the previous literature. For example, Coles and Shorrer (2013) use the same functional form, albeit in terms of efficiency. In a similar vein, Boudreau and Knoblauch (2010) use aggregate satisfaction defined as the sum of men's (and women's) rankings of their assigned mates. Yet, some further considerations are necessary. First, we choose a functional form which allow us to consider both the utility for matched and unmatched agents. While, in the first case, it is a measure of distance of the realized match to the most preferred choice, in the second case, the inefficiency is consistently defined as equal 1. In this way, unmatched outcomes are highly penalized even though, depending on the agent and on the context, they may be penalized more or less.²³ Also, the proposed measure gives an important role to the agents' rank. We believe that distance, $rk_g(m(g_i)) - 1$ and $rk_u(m(u_i)) - 1$, approximates quite well the heterogeneity in the inefficiency obtained by the possible matching outcomes. In reality, a Ph.D. graduate may care a lot about the top 25 universities, but be quite indifferent to middle ranked universities. Someone else may perceive a matching outcome with a lower ranked university as being very inefficient.²⁴ For this reason, we consider different shapes of the inefficiency function, captured by the parameter a_g and a_u .

Second, in addition to the aggregate inefficiency computed both for Ph.D. graduates and universities, other elements could be considered in a welfare comparison. For example, if the social planner is concerned about fairness, he might consider the dispersion of the inefficiency for the Ph.D. graduates (universities). In this case, our results vary in function of a_g and V_g (a_u and V_u). In addition, it is possible to look at who gains more by truncation.

²³Unmatched outcomes are different for Ph.D. graduates and universities. Universities will delay the hiring of an assistant professor to the next year and overcome the teaching needs with someone else. Ph.D. graduates might opt for post-doc positions and try the job market again the year after, or they might have good outside options (in another market). However, this varies substantially year by year depending on external factors (i.e., economic conditions).

²⁴An analogous reasoning applies to universities.

We demonstrate that very good universities in particular and highly-qualified graduates still benefit from high truncation, even under high correlation level. Otherwise, the social planner might care about the costs of implementing the matching mechanism, such as the number of rounds required in order to reach the stable matching outcome. Importantly, we find that limiting the number of applications would be substantially beneficial for decreasing the number of rounds as an a stable matching would be reached in a shorter time (see Figure 5).

Appendix B. Strategic preference manipulation with “types”

It is possible to extend the analysis in the paper to account for the universities’ preference correlation ($V_u > 0$). First, we assume that universities observe objective measures of the quality of graduates (grades, number of publications, quality of the journals where these articles are published, conference talks, and so on), and, for this reason universities will agree, to some extent, on their preferences over Ph.D. graduates. Thus, we will distinguish between three type of graduates. A graduate i can be a *very good graduate*, high type g_i^H , an *average graduate*, medium type g_i^M , or a *not very competitive graduate*, low type g_i^L , where we model the types using universities’ preferences. A very good graduate is modelled by taking the preferences of the universities

$$S_{u_jg} = \varepsilon_{u_jg} + V_u \times J_g, \quad \text{for, } \forall j = 1, \dots, N$$

and setting the joint preference factor, J_g , to 0.9. Setting $J_g = 0.5$ defines an average graduate and $J_g = 0.1$ defines a graduate who is not preferred much by universities.

In the simulations below, we consider one graduate, who has a certain type, and investigate his matching outcome under different optimization strategies. In the simulations we first consider one deviating candidate, which will be of a specific type (g_i^H , g_i^M or g_i^L). We determine his inefficiency via Equation (4) for different truncation levels p and different level of preference correlation among graduates, V_g and different α . The case $\alpha = 0$ refers to the truth-telling, where the graduate applies to p his most preferred universities (see formula (7)). The preferences of universities are given by $S_{u_jg} = \varepsilon_{u_jg} + V_u \times J_g$, for $\forall j = 1, \dots, N$ with $V_u = 1$.

The main results are the following (see Figure 7). First, the restriction on the number of applications p influences the level α chosen by the deviating

graduate. Misrepresenting the list of most preferred positions is not a worthy strategy when graduates can apply almost everywhere. But when graduates are very much restricted in the length of the application list (p is small), the benefit of misrepresentation increases. Second, we find that the optimal α decreases with the quality of the graduate, identified by the different types in Figure 7. More precisely, a not very competitive graduate will choose a high level of α , an average graduate will prefer an intermediate level of α , and a very good graduate will choose lower α . In other words, while a good candidate does not benefit from misrepresenting of the list of most preferred positions, a not very good candidate can decrease his inefficiency substantially by applying to not very popular universities (where the most of other candidates would not apply). Third, the level of correlation among graduates, V_g is important to determine the optimal strategy.

References

- Avery, C., Levin, J., 2010. Early admissions at selective colleges. *The American Economic Review* 100, 2125–2156.
- Bandyopadhyay, S., Iskhakov, F., Johnson, T., Lee, S., McArthur, D., Rust, J., Watson, J., Watson, J., 2013. Can the job market for economists be improved?, in: Nir Vulkan, A.E.R., Neeman, Z. (Eds.), *The Handbook of Market Design*. Oxford University Press. *The Handbook of Market Design*.
- Boudreau, J.W., Knoblauch, V., 2010. Marriage matching and intercorrelation of preferences. *Journal of Public Economic Theory* 12, 587–602.
- Braun, S., Dwenger, N., Kubler, D., 2010. Telling the truth may not pay off: an empirical study of centralized university admissions in Germany. *The B.E. Journal of Economic Analysis and Policy* 10, 1–38.
- Caldarelli, G., Capocci, A., 2001. Beauty and distance in the stable marriage problem. *Physica A: Statistical Mechanics and its Applications* 300, 325–331.
- Calsamiglia, C., Haeringer, G., Klijn, F., 2010. Constrained school choice: an experimental study. *American Economic Review* 100, 1860–74.
- Chen, Y., Sonmez, T., 2006. School choice: an experimental study. *Journal of Economic Theory* 127, 202–231.
- Coles, P., Cawley, J., Levine, P.B., Niederle, M., Roth, A.E., Siegfried, J.J., 2010. The job market for new economists: a market design perspective. *Journal of Economic Perspectives* 24, 187–206.
- Coles, P., Kushnir, A., Niederle, M., 2013. Preference signaling in matching markets. *American Economic Journal: Microeconomics* 5, 99–134.
- Coles, P.A., Shorrer, R.I., 2013. Optimal Truncation in Matching Markets. FEEM Working Paper 49.2013.
- Cyranoski, D., Gilbert, N., Ledford, H., Nayar, A., Yahia, M., 2011. Education: the PhD factory. *Nature* , 276–279.
- Damiano, E., Li, H., 2007. Price discrimination and efficient matching. *Economic Theory* 30, 243–263.

- Gale, D., Shapley, L.S., 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69, 9–15.
- Haeringer, G., Iehlé, V., Rodriguez-Planas, N., 2010. Knocking out the supply and sorting in centralized job markets. Mimeo .
- Kaminski, D., Geisler, C., 2012. Survival analysis of faculty retention in science and engineering by gender. *Science* 335, 864–866.
- Klijn, F., Pais, J., Vorsatz, M., 2013. Preference intensities and risk aversion in school choice: a laboratory experiment. *Experimental Economics* 16, 1–22.
- Lee, R.S., Schwarz, M.A., 2007. Signalling preferences in interviewing markets, in: Cramton, P., Muller, R., Tardos, E., Tennenholtz, M. (Eds.), *Computational Social Systems and the Internet*, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, Dagstuhl, Germany.
- Pathak, P.A., Sonmez, T., 2013. School admissions reform in chicago and england: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review* 103, 80–106.
- Roth, A.E., 1982. The economics of matching: Stability and incentives. *Mathematics of Operations Research* 7, 617–628.
- Roth, A.E., Rothblum, U.G., 1999. Truncation strategies in matching markets-in search of advice for participants. *Econometrica* 67, 21–43.
- Stock, W.A., Alston, R.M., Milkman, M., 2000. The academic labor market for economists: 1995–96. *Atlantic Economic Journal* 28, 164–185.

Author contacts:

Janin Balter

Max Weber Programme, European University Institute

Via dei Roccettini, 9

I-50014 San Domenico di Fiesole (FI)

Italy

Email: janine.balter@eui.eu

Michela Rancan

Via dei Protoveneti 7,

35044 Montagnana PD

Email: michela.rancan@eui.eu

Olena Senyuta

Max Weber Programme, European University Institute

Via dei Roccettini, 9

I-50014 San Domenico di Fiesole (FI)

Italy

Email: osenyuta@gmail.com