

IZA DP No. 9386

**Procedures vs. Incentives:
The Case of the University Promotion System in Italy**

Carlo dal Maso
Enrico Rettore
Lorenzo Rocco

September 2015

Procedures vs. Incentives: The Case of the University Promotion System in Italy

Carlo dal Maso

IMT Lucca

Enrico Rettore

University of Padova, FBK-IRVAPP and IZA

Lorenzo Rocco

University of Padova

Discussion Paper No. 9386
September 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Procedures vs. Incentives: The Case of the University Promotion System in Italy

A common observation is that individuals strive to neutralize the effect of procedural rules designed to drive choices away from their private optimum. An example of this phenomenon is offered by the reaction of Italian academia to two reforms that modified the procedures of recruitment and promotion, by introducing random selection of the examiners not appointed by the recruiting school and reducing from two to one the number of candidates to be qualified. We model the negotiation occurring within evaluation committees and test the decision rule implied by the theoretical model on the sample composed of all selections to associate and full professorship initiated by the Italian schools of economics between 2004 and 2011. Particularly, we investigate whether these reforms decreased the relative weight of the examiner appointed by the recruiting school on committee's decision. Empirical results suggest that both reforms had little if no effect on examiners' weights.

JEL Classification: D71, M51, I2

Keywords: university recruitment, incentives, negotiation, formal procedures

Corresponding author:

Lorenzo Rocco
Department of Economics
University of Padova
Via del Santo 33
35121 Padova
Italy
E-mail: lorenzo.rocco@unipd.it

1 Introduction

When a formal procedure is introduced to constrain individual behavior, a strategic reaction should be expected. Often, people struggle to complying with the procedure without being forced to change their optimal choice, in so doing bypassing the spirit and the purpose of the norm.

An example of this phenomenon is provided by two reforms of recruitment and promotion of professors in Italian academia. The first reform, enacted in 2008, introduced the principle of randomization in the composition of the local evaluation committees. Before, one member of the committee was appointed by the school calling for the position and four members were professors elected by all peers nationwide. After the reform the four external examiners were randomly drawn from a list of peers. The second reform, introduced in 2005, suspended in 2008 and soon restored, reduced from two to one the number of candidates to be qualified in each local selection (called *concorso*). In both cases the declared purpose was that of enhancing the independence of the external examiners, reducing collusion within each committee and by that way making recruitment and promotion more meritocratic.

In this paper we evaluate the effect of such reforms. We first propose a model of candidates' qualification from which we derive the committee's optimal decision rule. The latter has a simple empirical counterpart that we estimate on all *concorsi* held in Italy between 2004 and 2011 and called by the schools of economics. Particularly, we estimate the relative weight of the internal examiner on committee decisions.

Our results suggest that both reforms did not significantly altered the preeminent role of the internal examiner on recruitment and promotion decisions, supporting the claim that Italian academia was able to neutralize the effect of the new procedures and continue its business as usual.

This result is not surprising given that the reforms modified only procedural aspects of recruitment but did not alter the fundamental incentives of academia and academics. Indeed, both reforms were introduced in a context of low-powered incentives, where funds were equally distributed among universities, professors salary depended on seniority only, while high quality teaching or strong scientific productivity played no role. Professors directly managed their institutions and their power and rents depended on colleagues' consensus. Hence, rather than being considered a resource, the most productive candidates in the *concorsi* were perceived as a threat, as they could have altered the *status quo* in the schools. Quite the opposite, the preferred candidates were those offering enough guarantees regarding their willingness to preserve and support the *status quo*.

Our analysis is related with Perotti (2002) and Durante et al. (2011). They look at

the selections held in Italy immediately after the decentralization of recruitment occurred in 1998. Perotti (2002) concludes that decentralization did not achieve a higher degree of meritocracy compared to the previous system. Being a candidate affiliated or well connected to the recruiting school had a dominant influence on the probability of being qualified and appointed, especially when the scientific productivity of the evaluation committee was below the median. Interestingly, Perotti (2002) analyzes the determinants of the number of votes that external examiners received and finds no relation with their scientific record. Durante et al. (2011) study familism/nepotism in academic promotions by looking at the pattern of family names in the recruiting university and among the candidates. They find a strong association between civic capital measured at the province level and nepotism and argue that while decentralization increased the possibility to behave opportunistically everywhere, misconduct increased mainly in the areas poor of civic capital. A recent contribution by Di Paola and Scoppa (2011) exploits the random composition of the evaluation committees (the same reform we analyze in this paper) to establish whether promotions are affected by gender bias. They follow the same approach adopted by Bagues and Esteve-Volart (2007) that investigate gender bias in national public exams to access four main corps of the Spanish Judiciary over 1987-2005, Zinovyeva and Bagues (2011) that estimate gender bias in academic promotions in Spain, and in Zinovyeva and Bagues (2012) that study the role of connections compared to productivity. Also Combes et al. (2008), by using the French data of the *agrégation nationale* between 1984 and 2003, assess the relative importance of scientific production compared to professional links between candidates and evaluators.

Our paper is organized as follows. Section 2 illustrates some features of the institutional background, Section 3 discusses the dataset and Section 4 introduces the model. Section 5 presents the empirical implementation. Results are reported in Section 6. Possible alternative explanations of our findings are discussed in Section 7. Conclusions and technical appendices follow.

2 Institutional Background

Italian *concorsi* have been widely criticized for not selecting the best candidates, reflecting the scarce attention of academia to meritocracy. The features of the public university funding prevailing between 1999 and 2010 help understanding what incentives universities, schools and academics faced in that period. There were neither a nationwide evaluation of university productivity or a firmly established link between funding and performance. Before 2009 universities were assigned government funds - the so called *Fondo di Finanziamento Ordinario* (FFO) - on a completely egalitarian basis, regardless

of their performance in research and teaching. Salaries increased each other year for everyone by the same proportion and career progression depended very little on publications and productivity (Perotti, 2002). Professors were subject to very few obligations regarding teaching and research. Overall, university provided decent rents to academics, in terms of a prestigious social position, a good salary, the possibility of carrying on external professional activities.¹

There is some consensus - even if based only on anecdotal evidence - that Italian academy was quite successful in preserving and maximizing its rent, an achievement generally incompatible with a meritocratic recruitment.² The decentralized promotion system introduced in 1998, by which a school with a vacancy was allowed to call a *concorso*, offered schools the possibility of recruiting the candidates fitting better with their needs and, by enlarging the opportunities of collusion, favored the purpose of maintaining the *status quo*. Academics affiliated to different schools jointly determined the outcome of several *concorsi*, by mutually agreeing on the principle that each school had the indisputable right of deciding who was to be hired, regardless of any consideration of candidate's scientific merits. As discussed by Perotti (2002) also examiners with good scientific records that participate to evaluation committees often accepted this principle, because they knew that any deviation from the collusive equilibrium could have been punished in the future *concorsi* that their own schools would have opened.

Collusion was also favored by the rules governing *concorsi*. Candidates were evaluated by a committee made up of five professors: one member of the committee was appointed by the recruiting school (internal examiner) and the other four came from different universities (external examiners). The evaluation committee selected a fixed number of candidates to be qualified (called *idonei*) and the recruiting school had the option, but not the obligation, of appointing one *idoneo* of its choice. The appointed *idoneo* was the winner of the *concorso*. The remaining *idonei* could be appointed by any other university, including those the external examiners came from, within two or three years. Between 1999 and 2008 external examiners were elected by the whole national body of full professors in the same scientific field, a procedure that allowed the election of external examiners to be "managed" in order to obtain committees accommodating

¹Nevertheless, the absence of a link between productivity and remuneration or career progression does not imply that the generality of Italian professors were inactive, nor that they did not care about their publication record. It means only that good research was not promoted by the university system and that it ultimately rested on the intrinsic motivation and good will of individuals and schools. The situation might improve in the next future. In 2011 the national agency of evaluation (ANVUR) became fully operational and the first nationwide assessment of research production (VQR 2004-2010) terminated in 2014. Since 2009 a small share of the FFO has been linked to some productivity indicators (this share was 7 percent of total funding in 2009 and increased up to 15 percent in 2013).

²Indeed, recruiting highly productive scholars, more interested in scientific achievements than on rents, would likely alter the *status quo*.

the will of the recruiting school (Perotti, 2002). Between 1999 and 2001 three candidates in each *concorso* could be qualified. Between 2002 and 2005, this number decreased to two (the *double qualification system*). Having more than one qualification left room for extensive “horse-trading” between internal and external examiners and between schools. The internal examiner could more easily obtain the qualification of his preferred candidate by promising of voting for the preferred candidate of an external examiner for the second *idoneità*.

This system underwent two reforms. The first regarded the composition of evaluation committees. Since 2007, external examiners were randomly selected from a (long enough) list of professors. The purpose of the reform was that of making more difficult to manipulate the composition of the evaluation committee and of increasing the independence of the external examiners, by this way reducing the weight of the internal examiner and the recruiting school on the final decision. The second reform regarded the number of *idonei* in each *concorso*. Since 2005 only one candidate could qualify for each position (the *single qualification system*). This decision lasted little and two qualifications were re-introduced in 2007. Eventually, following public condemnation, the single qualification system was re-established in 2008. Figure 1 depicts this dynamic.³

[FIGURE 1 ABOUT HERE]

These reforms modified two formal rules of the *concorsi* with the purpose of inducing committees to select better deserving candidates even if in contrast with the desires of the recruiting school. In fact, these reforms contradicted the logic of the decentralized system and recognized that schools had little incentive to hire and promote on a meritocratic basis. However, hardly are downstream interventions effective if not accompanied with more radical reforms able to modify fundamental incentives. Indeed, if schools agreed - even implicitly - on the principle that each of them is free to decide who is to be hired, then neither the way external examiners are selected, nor the number of *idonei* would matter for the final outcome of a *concorso*.

3 Theoretical model

We model the selection procedure as a negotiation between internal and external examiners. The model presented in this section refers to the case of double qualification. The single qualification setting may be interpreted as a special case and will be discussed in Appendix B.

³Since 2012 a new centralized recruitment system has been introduced and the first promotions and recruitment based on the new system have taken place in 2014.

Suppose that a continuum of aspirant professors apply to a *concorso*. Define the quality of each candidate as the set of his relevant characteristics and let it be summarized by a uni-dimensional measure, denoted h , with $h \geq 0$. Each candidate has a particular value of h .

Evaluation committees receive and examine candidates curricula and eventually decide who are the *idonei*. To simplify the analysis we assume that the committee is composed by one internal examiner and only one external examiner.⁴ We also assume that the internal examiner shares the same preferences of the recruiting school, so that any agency problem can be set aside.

Internal and external examiners negotiate over which applicants should qualify. Next the internal examiner, on behalf of the recruiting school, decides whether to appoint and what candidate to appoint.⁵

Each party conceives an ideal candidate and weights each applicant according to the distance from such ideal. Each party's ideal is a candidate that perfectly reflects party's characteristics, i.e. with a h score equal to party's h . By this characterization we capture the preference for preserving the *status quo* that was prevalent in the Italian university in those years.⁶

We formalize the negotiation between internal and external examiners and the subsequent decision of the internal examiner as a modified "collective model" (see Chiappori, 1988)⁷ with a guarantee of minimum payoff to the internal examiner. Thus, committee preferences are described by a weighted average of its members payoffs. In this welfare function the "Pareto weights" reflect the relative bargaining power of the two parties or their relative influence on the committee decision. The guarantee of minimum payoff accounts in a simple way for the option that the internal examiner has to refuse of appointing one of the *idonei* if they are both too far from his ideal. Upon refusal he receives a minimum payoff equal to the option value, exogenously fixed, of organizing another selection in the future and recruiting someone else possibly closer to his prefer-

⁴The latter assumption implies that the preferences of the external examiners can be summarized by those of a representative agent.

⁵This formalization is consistent with empirical data: typically committees seek an agreement among all examiners or at least of four out of five. Both anecdotal and empirical evidence exclude that the final decision in a *concorso* could be obtained by means of majority voting or other procedures that do not involve communication, coordination, mediation and compromise among examiners. Partly the search for an agreement is sponsored by the rules governing the recruitment procedures (e.g. most evaluation criteria are stated by the law and additional criteria need to be approved by the committee and bind all examiners evaluation), partly unanimity serves to discourage future complaints by the non qualified candidates and partly unanimity is the result of collusion and the mutual agreement on the principle of "right-to-decide" of each recruiting school.

⁶A school with an open vacancy that calls for selection procedure wants to preserve and reproduce its average quality, a strategy which guarantees the stability of the existing internal equilibria.

⁷The collective model is commonly used to represent the choice of a household composed of several members with different preferences.

ences. No such guarantee protects the external examiner. The inclusion of this device in the model serves to account for the relevant proportion of *concorsi* that end up with no appointment between 2004 and 2011, as we shall illustrate in the next section.

Formally, examiners utility function is defined as

$$\begin{aligned} U_i &= -\alpha(h_i - h_1)^2 - (1 - \alpha)(h_i - h_2)^2 \\ U_e &= -\beta(h_e - h_1)^2 - (1 - \beta)(h_e - h_2)^2 \end{aligned} \quad (1)$$

where subscript i and e identifies internal and external examiners respectively, U is the utility function, h_1 and h_2 are the qualities of the two qualified candidates, h_i and h_e are the qualities of the examiners and, finally, α and β are preference parameters over candidates of the internal and external examiners, respectively. We assume $\alpha > \frac{1}{2}$ so that *idoneo* 1 will weight more than *idoneo* 2 in the internal examiner utility function. Thus, at equilibrium *idoneo* 1 will be closer than *idoneo* 2 to the internal examiner ideal, and, in case of recruitment, he will be the chosen candidate (i.e. the winner).

Once two *idonei* have been identified, the internal examiner will appoint *idoneo* 1 if the utility U_i he derives from the selection procedure exceeds his minimum payoff \bar{U}_i . Otherwise, no one will be appointed and the *concorso* concludes with a *failure*. From the external examiner's viewpoint failure has pros and cons. On the pros side, in a *failure*, the external examiner obtains two *idonei* relatively close to his bliss point. On the cons side, although the two *idonei* may be appointed by other schools within two or three years, appointment is not guaranteed and it will certainly take time. Moreover the external examiner will be held responsible for the waste of resources suffered by the recruiting school and he will be at risk of being sanctioned in future selections procedures. We assume that the cost of a failure to the external examiner is exogenous and equal to δ .

Formally, the problem that internal and external examiners face is that of maximizing

$$W = \begin{cases} \gamma U_i + (1 - \gamma)U_c & \text{if } U_i \geq -\bar{U}_i \\ \gamma \bar{U}_i + (1 - \gamma)(U_c - \delta) & \text{if } U_i < -\bar{U}_i \end{cases} \quad (2)$$

with respect to h_1 and h_2 , where γ is the ‘‘Pareto weight’’ of the internal examiner. The first line corresponds to the *appointment regime*, where *idoneo* 1 will be eventually recruited, and the second to the *failure regime*, where two candidates are qualified but no one will be recruited.

Being the problem discontinuous, the two regimes have to be analyzed separately. Formally each regime involves a constrained maximization in its own. In Appendix A we show that, if δ is small enough, the solutions of problem (2) is either the unconstrained solution of the appointment regime or the unconstrained solution of the failure regime,

depending on the values of the parameters⁸.

Precisely, the optimum of (2), which represents the decision rule of the committee, is

$$\begin{cases} h_1^* = \frac{\alpha\gamma}{K}h_i + \frac{\beta(1-\gamma)}{K}h_e \\ h_2^* = \frac{(1-\alpha)\gamma}{1-K}h_i + \frac{(1-\beta)(1-\gamma)}{1-K}h_e \end{cases} \quad \text{if } (h_i - h_e)^2 \leq \Delta \quad (3)$$

and

$$\begin{cases} h_1^* = h_e \\ h_2^* = h_e \end{cases} \quad \text{if } (h_i - h_e)^2 > \Delta \quad (4)$$

where $K = \alpha\gamma + \beta(1 - \gamma)$ and Δ is a complex function of the parameters $(\alpha, \beta, \gamma, \delta, \bar{U}_i)$ defined in Appendix A.

Indeed, if the distance between the internal and external examiner's bliss points is not too large, the solution of the negotiation will satisfy the internal examiner and the appointment of *idoneo* 1 will conclude the procedure. In this case both *idonei* have characteristics located between the bliss points of the two examiners and they will be closer to the internal examiner bliss point the higher his weight (for any fixed preference parameters α and β). Instead, when such distance exceeds a well defined threshold, the outcome of the *concorso* will be a failure, no candidate will be appointed, but the two *idonei* h score will coincide with the bliss point of the external examiner.

Quite similarly, in the single-qualification case we have

$$h_1^* = \gamma h_i + (1 - \gamma)h_e \quad \text{if } (h_i - h_e)^2 \leq \Theta \quad (5)$$

and

$$h_1^* = h_e \quad \text{if } (h_i - h_e)^2 > \Theta \quad (6)$$

where Θ is a function of the parameters $(\gamma, \delta, \bar{U}_i)$. We refer to Appendix B for a formal derivation of the single-qualification case.

4 Data

Our data include all *concorsi* to associate and full professorship called by Italian schools of economics (*facoltà di economia*) and concluded between 2004 and 2011. We have discarded all competitions in non-core scientific areas (foreign languages, informatics, sociology) and focused on economics, management, law and quantitative disciplines (stats and maths). In Table 1 we report the number of *concorsi* by year of the call, distinguish-

⁸In other words the optimum of the problem never occurs at the discontinuity frontier of the objective function W .

ing between *concorsi* with elected committees (*Pre-reform*) and with randomly selected committees (*Post-reform*) and between *concorsi* with one or two qualifications.⁹

[TABLE 1 ABOUT HERE]

Quite surprisingly, more than 10 percent of all *concorsi* ended up without winners, i.e. the recruiting school eventually decided not to appoint any qualified candidate. We refer to these cases as *failures* (see Table 2). We have investigated what justification was provided by the recruiting schools in case of failure: in most cases the reason for non appointing was a major misalignment between *idonei*'s characteristics and the characteristics of the position the school posted in the call.¹⁰

[TABLE 2 ABOUT HERE]

Data include biographical information of all examiners and qualified candidates, the names of the *idonei* and of the winner (if any) for each competition, the dates of the call and of the end of the procedure. We do not have information on candidates who apply for the position but eventually did not qualify.

Perhaps the most critical choice in this study is how to measure candidates' and examiners' quality. To be fully consistent with the model we should consider an index that combines scientific productivity, teaching quality and all other characteristics that might be relevant in a selection procedure. Being data on all these dimensions unavailable - if not unobservable altogether - we focus on scientific productivity assuming that it is an unbiased predictor of candidates' and examiners' overall quality.

Scientific productivity is the only dimension that is measurable for all subjects involved in our analysis in an objective manner. We adopt the *h index* proposed by Hirsch, an index which takes into account both quantity and impact of each scholar scientific production. For instance, an h-index equal to n indicates that a researcher has published at least n papers cited by at least n other publications belonging to the same dataset. By construction, the index is robust to cases of scholars with few but highly cited papers as well as of authors with lot of papers with only a modest impact on the scientific community.

We have collected the h-index of all qualified candidates, examiners and the entire faculty of the recruiting school, from the "Publish or Perish" software (Harzing, 2007) which exploits the Google Scholar database of publications and citations. Eventually, we

⁹Note that in 2007 there are no calls. This happened because the introduction of new rules forced the Minister to postpone all outstanding competitions. The procedures for the *concorsi* called for in 2008 were suspended for almost two years and they ended up in 2010 and 2011.

¹⁰There are two cases where faculty motivated the decision with budget problems or a major reorganization of teachings underway. We have excluded these cases from our dataset.

have collected the h-index of candidates, examiners and faculty members for almost 500 *concorsi* between 2004 and 2011 making up a database of about 36000 entries.¹¹ The choice of measuring scientific productivity over the Google Scholar data, which includes not only published articles and books but also gray publication such as working papers and reports, depends on the fact that we need information for quite different scientific areas (e.g. law v.s. economics) with quite different publication traditions. Furthermore, Google Scholar is the most comprehensive search engine for social sciences (Harzing and Van der Wai, 2008).

We use as an unbiased predictor of h_1 and h_2 , candidates qualities, their own h-index as of the date of the call, h_e by the average h-index of the four external examiners in each evaluation committee and h_i by the average h-index of the faculty of the recruiting school in the same field of the position to be awarded. This way we make operational our assumption that the internal examiner acts on behalf of his school and completely shares his school's preferences.

In Tables 3 and 4 we report summary statistics of the h-index of candidates and examiners distinguishing between double and single qualification, by outcome of the *concorso* (failure or appointment), and by method of committee selection (elected, randomized)¹².

[TABLE 3 ABOUT HERE]

[TABLE 4 ABOUT HERE]

Descriptives based on the “Pooled Sample” (right-hand panel) immediately reveal that the *concorsi* ending up with a failure are those where the difference between the internal and the external examiners' h-index is relatively large, consistently with the predictions of our model. Figure 2 depicts the cumulative distribution of the absolute value of $(h_i - h_e)$ by type of outcome (after controlling for a full set of dummies for scientific field, year of the *concorso*, geographical area of the school, level of the *concorso*). In accordance with the implications of the theoretical model the (absolute) difference in quality between internal and external examiners is larger when a failure occurs.

[FIGURE 2 ABOUT HERE]

Another interesting evidence emerging from the double-qualification case is that the median h-index of both *idonei* is in between h_i and h_e in case of appointment and close

¹¹The h-index is based exclusively on the publications appeared on Google Scholar before the date of call of each *concorso*

¹²The sample of *concorsi* that we consider coincides with that used in the empirical analysis. Compared to the full sample of *concorsi* reported in Table 1, 27 *concorsi* are missing for lack of some relevant information.

to the external examiner bliss point in case of failure. Qualitatively similar results holds for *concorsi* with a single qualification, although in the appointment case median h-index of the unique *idoneo* is outside the interval $[h_i, h_e]$.

Comparing the pre-reform (elected committees) and the post-reform (randomized committees) samples, we note that the h-index of post-reform candidates and examiners are moderately larger and more dispersed. Unexpectedly however, not only the variance between-*concorsi* of external examiners' h-index increased little after the reform (as the IQR values of Tables 3 and 4 indicate), but also the variance within-*concorsi* remained remarkably stable, suggesting that, overall, randomization had a relatively small impact on external examiners, at least as regards their h-index.¹³ Actually, the more pronounced differences regarded internal examiners, i.e. the the members chosen by the recruiting schools, who were not directly influenced by the reform.

5 Empirical Implementation.

After rewriting the decision rule of the committee as

$$h_1^* - h_e = \begin{cases} \frac{\alpha\gamma}{K}(h_i - h_e) & \text{if } (h_i - h_e)^2 \leq \Delta \\ 0 & \text{if } (h_i - h_e)^2 > \Delta \end{cases} \quad (7)$$

$$h_2^* - h_e = \begin{cases} \frac{(1-\alpha)\gamma}{1-K}(h_i - h_e) & \text{if } (h_i - h_e)^2 \leq \Delta \\ 0 & \text{if } (h_i - h_e)^2 > \Delta \end{cases} \quad (8)$$

the *key implication* of the model becomes apparent: the distance between internal and external examiners ($h_i - h_e$) determines both the *discrete outcome* (agreement vs. failure) and the *continuous outcome* (the quality for the qualified candidates).

To take to the data the equation for the decision rule we set

$$h_i - h_e = (H_i - H_e) + X\theta + \varepsilon \quad (9)$$

$$h_1^* - h_e = (H_1 - H_e) + X\theta + \varepsilon \quad (10)$$

$$h_2^* - h_e = (H_2 - H_e) + X\theta + \varepsilon \quad (11)$$

where H is for Hirsch h-index. The component $X\theta + \varepsilon$ (X observable and ε unobservable) is a *concorso*-specific prediction error of the unobserved quality of candidates and

¹³The average standard deviation rose from 2.56 to 2,86 in the double qualification case and from 2.56 to 3.30 in the single qualification case.

examiners. Since this component is *concorso*-specific we impose the restriction that θ is common across equations (9)-(11). Furthermore, ε is a random variable with zero-mean and standard deviation σ accounting for unobservable characteristics of each *concorso* that we assume to be orthogonal to $(H_i - H_e)$, $(H_1 - H_e)$ and $(H_2 - H_e)$ conditional to X .¹⁴

Replacing equations (9)-(11) into equations (7) and (8) yields the empirical counterparts of the optimal decision rules in the case of agreement:

$$H_1 - H_e = \frac{\alpha\gamma}{K}(H_i - H_e) + \left(\frac{\alpha\gamma}{K} - 1\right)X\theta + \left(\frac{\alpha\gamma}{K} - 1\right)\varepsilon \quad (12)$$

and

$$H_2 - H_e = \frac{(1-\alpha)\gamma}{1-K}(H_i - H_e) + \left(\frac{(1-\alpha)\gamma}{1-K} - 1\right)X\theta + \left(\frac{(1-\alpha)\gamma}{1-K} - 1\right)\varepsilon \quad (13)$$

Agreement occurs when $-\sqrt{\Delta} \leq (H_i - H_e) + X\theta + \varepsilon \leq \sqrt{\Delta}$ and failure when $(H_i - H_e) + X\theta + \varepsilon < -\sqrt{\Delta}$ or $(H_i - H_e) + X\theta + \varepsilon > \sqrt{\Delta}$.

*Observable concorso characteristics include a dummy variable equal to one if the concorso is to appoint a full professor and equal to zero if it is to appoint an associate professor; dummies for the scientific field of the competition; dummies of geographic area (to capture the characteristics of the schools calling the selection); and year dummies. The inclusion of scientific field and geographic dummies captures fundamental differences among scientific fields in their propensity to publish and among schools as regards the criteria they adopt to appoint new scholars. To reduce the dimensionality of the regression problem we replace X in our equations with a *bivariate* Generalized Propensity Score including the estimated conditional expectation, $GPS1(X)$, and conditional variance, $GPS2(X)$, of $(H_i - H_e)$ on the control variables in X (see Imbens, 2000). The key mathematical property of the GPS is that if conditioning on the whole set of control variables warrants the identification of the causal parameter of interest, then conditioning on the GPS based on the very same set of control variables warrants identification as well.*

The effect of the reform is captured by including an interaction between $(H_i - H_e)$ and a dummy *reform* which takes one for *concorsi* with a randomized committee.¹⁵ To simplify estimation, we impose from the outset the restriction that the threshold Δ relevant for the outcome failure/agreement is *not* affected by the reform. At first glance this might seem inconsistent with the theoretical model, but in appendix C we suggest this is indeed

¹⁴Note that rewriting the model in terms of deviations from h_e implicitly imposes the constraint that the coefficients of h_i and h_e sum up to 1.

¹⁵The crucial identifying restriction we are imposing here is that conditional on the control variables the Pareto weight would have been stable overtime in the absence of reforms.

the case at least approximately.

Hence, the equations we estimate in case of agreement are:

$$H_1 - H_e = \pi_{11}(H_i - H_e) + \pi_{12}(H_i - H_e) \times reform + \quad (14)$$

$$+ (\pi_{11} - 1) (\theta_0 + \theta_1 GPS1(X) + \theta_2 GPS2(X)) + (\pi_{11} - 1) \varepsilon$$

and

$$H_2 - H_e = \pi_{21}(H_i - H_e) + \pi_{22}(H_i - H_e) \times reform + \quad (15)$$

$$+ (\pi_{21} - 1) (\theta_0 + \theta_1 GPS1(X) + \theta_2 GPS2(X)) + (\pi_{21} - 1) \varepsilon$$

where $\frac{\alpha\gamma}{K} = \pi_{11}$ and $\frac{(1-\alpha)\gamma}{1-K} = \pi_{21}$ in the *concorsi* with elected committees (*reform=0*) and $\frac{\alpha\gamma}{K} = \pi_{11} + \pi_{12}$ and $\frac{(1-\alpha)\gamma}{1-K} = \pi_{21} + \pi_{22}$ in *concorsi* with randomized committees (*reform=1*).

The key parameter is the Pareto weight of the internal examiner, γ , before and after the reform that introduced randomization. Our strategy identifies only $\frac{\alpha\gamma}{K}$ and $\frac{(1-\alpha)\gamma}{1-K}$ both before and after the reform. Note however that if $\frac{\alpha\gamma}{K} > \frac{(1-\alpha)\gamma}{1-K}$, a condition directly testable against the data, then $\alpha > K$. This fact bears two important implications:

1) $\pi_{21} < \gamma < \pi_{11}$ (and $\pi_{21} + \pi_{22} < \gamma < \pi_{11} + \pi_{12}$ after the reform) i.e. the estimated parameters π_{21} (resp. $\pi_{21} + \pi_{22}$) and π_{11} (resp. $\pi_{11} + \pi_{12}$) are lower and upper bounds, respectively, for the parameter of interest γ , since $\frac{1-\alpha}{1-K}\gamma < \gamma < \frac{\alpha}{K}\gamma$.

2) $\alpha > \beta$, i.e. the internal examiner weights more *idoneo* 1 compared to the external examiner.

Although without further assumptions on preference parameters α and β we can achieve only partial identification of γ , this will prove to be enough to draw conclusions.

The case of single qualification is alike. The empirical equation to estimate and the condition for agreement vs failure can be easily derived from equations (5) and (6)

$$H_1 - H_e = \pi_{s1}(H_i - H_e) + \pi_{s2}(H_i - H_e) \times reform + \quad (16)$$

$$+ (\pi_{s1} - 1) (\theta_0 + \theta_1 GPS1(X) + \theta_2 GPS2(X)) + (\pi_{s1} - 1) \varepsilon$$

where $\gamma = \pi_{s1}$ in the *concorsi* with elected committees and $\gamma = \pi_{s1} + \pi_{s2}$ in the *concorsi* with randomized committees. Note that in the single-qualification case exact identification of γ is obtained.

From now on consider only *idoneo* 1 - the cases for *idoneo* 2 and that of single-qualification being very similar.

5.1 Estimation procedure

To estimate the parameters of the model we proceed in two steps. First we estimate a modified Probit model for the probability to observe a *concorso* ending with an agreement. The standard Probit model is modified to take into account that a failure might occur either because $h_i - h_e = \theta_0 + (H_i - H_e) + \theta_1 GPS1(X) + \theta_2 GPS2(X) + \varepsilon < -\sqrt{\Delta}$ or because $h_i - h_e = \theta_0 + (H_i - H_e) + \theta_1 GPS1(X) + \theta_2 GPS2(X) + \varepsilon > \sqrt{\Delta}$. The resulting likelihood function is:

$$L^t = \prod_{\{failure\}} \left[\Phi \left(-\frac{1}{\sigma} (\Delta)^{1/2} - \frac{1}{\sigma} (H_i - H_e) \right) + \Phi \left(-\frac{1}{\sigma} (\Delta)^{1/2} + \frac{1}{\sigma} (H_i - H_e) \right) \right] \times \quad (17)$$

$$\times \prod_{\{agreement\}} \left[\Phi \left(\frac{1}{\sigma} (\Delta)^{1/2} - \frac{1}{\sigma} (H_i - H_e) \right) - \Phi \left(-\frac{1}{\sigma} (\Delta)^{1/2} - \frac{1}{\sigma} (H_i - H_e) \right) \right]$$

(for simplicity we omit from equation (17) the two components of the *GPS* and the constant). Maximization of (17) yields an estimate for $\frac{1}{\sigma}$ and for $\frac{1}{\sigma} (\Delta)^{1/2}$.

Second we estimate equations (14) and (15) on the sub-sample of *concorsi* ended with an agreement by means of a truncated regression estimator. The adjustment to take into account truncation is derived in Appendix D.¹⁶

Second step standard errors must be corrected to account for the use of estimated values for $\frac{1}{\sigma}$ and for $\frac{1}{\sigma} (\Delta)^{1/2}$ that enter in the adjustment for truncation. We implement the correction by running a Montecarlo simulation. From the distribution of ε , assumed to be normal with zero-mean and standard deviation equal to the value of σ estimated at the baseline, we draw 500 random vectors, denoted ε_j for $j = 1 \dots 500$. For each j we compute $I_j^* = \hat{\theta}_0 + (H_i - H_e) + \hat{\theta}_1 GPS1(X) + \hat{\theta}_2 GPS2(X) + \varepsilon_j$ and we generate a pseudo-outcome variable Y_j which takes 0 if $(I_j^*)^2 \leq \hat{\Delta}$ (predicted case of agreement) and 1 if $(I_j^*)^2 > \hat{\Delta}$ (predicted case of failure), where $\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2$ and $\hat{\Delta}$ are baseline estimates. Next, for any Y_j we re-run our two-stage procedure where the dependent variables of the second-stage truncated regression (for both *idoneo* 1 and *idoneo* 2) are generated combining baseline predicted values with ε_j . Both dependent and independent variables include the correction for truncation which is derived from first-stage estimates for each j . Thus, we produce 500 replications of all parameter estimates that we use to derive parameter standard errors.

As a benchmark we also report standard OLS estimates of equations (14) and (15) on the subsample of *concorsi* where agreement was achieved, neglecting the problem

¹⁶Note that the coefficients of $GPS1(X)$ and $GPS2(X)$ in equations in (14) and (15) are non linear functions of the coefficients of $(H_i - H_e)$ and of the coefficients of $GPS1(X)$ and $GPS2(X)$ in the agreement equation, something we can test against the data.

of truncation. This is because in this specific instance the bias induced by the upper truncation and the symmetric bias induced by the lower truncation might cancel out.

6 Empirical results

Results of first and second stage are reported in Table 5 and 6, for the case of double and single qualification respectively.¹⁷ As for the second stage, we report estimates of π_{11} , π_{12} and π_{21} , π_{22} respectively for *idoneo* 1 and *idoneo* 2, which are the bounds of γ , i.e. the Pareto weight of the internal examiner before and after the reform. For the case of single qualification we report the corresponding parameters π_{s1} and π_{s2} . There is a neat evidence that adjusting (two-step estimator) and not adjusting (OLS estimator) for truncation plays a minor role: point estimates are similar, although, as expected, OLS estimates are more precise. In the following we comment on the OLS.

Under the hypothesis that 1) evaluation committees aim to reach unanimous decisions to discourage subsequent appeals, and 2) external examiners perfectly coordinate (as assumed in the model), the Pareto weight of the internal examiner should be 0.2 (reflecting the fact that four external faced a single internal examiner). Let us take this value as a benchmark against which contrasting our results.

In the double-qualification case with elected committees (*reform* = 0) the bounds we find depict a narrow range for the Pareto weight of the internal examiner, going from 0.527 to 0.605. In the single qualification case, where exact identification is achieved, γ is slightly smaller (0.455), but still points to a preeminent role of the internal examiner as compared to the benchmark, even when reaching a decision is made more difficult by the impossibility of trading candidates within a given *concorso*. These estimates support the hypothesis that schools tend to replicate the *status quo* by recruiting candidates with a scientific standing similar to that of incumbent faculty.

In the double qualification case, the introduction of randomization in the composition of the committees (*reform* = 1) possibly widened the range of Pareto weights, although estimates are rather imprecise. We cannot thus exclude that γ decreased, but even in this case the internal examiner would have kept a substantial role. In the single qualification case, the post-reform Pareto weight of the internal examiner is not significantly different from its pre-reform level, though the point estimate is negative.¹⁸

¹⁷Note that in the first stage we do not impose the restriction that the censoring thresholds are symmetric around zero, since the intercept of the equation is left unconstrained.

¹⁸Note also that there is no clear evidence against the hypothesis that the coefficients of $GPS1(X)$ and $GPS2(X)$ are indeed the product of $(\pi_{c1} - 1)$, for $c = 1, 2, s$ and θ_1 and θ_2 respectively (see the bottom panels of Tables 5 and 6).

7 Discussion

From a repeated game perspective, academia played an equilibrium that granted each university the right of deciding the candidate to be promoted or recruited, at its sole discretion. External examiners accepted rather passively the choice of the recruiting school, fearing retaliation in the *concorsi* called by their own school in future. Given schools' preference for the *status quo*, at equilibrium each school was allowed to recruit a candidate of quality comparable to that of the incumbent faculty. Our analysis indicates that government attempts to alter this equilibrium by introducing new procedural rule failed as academia devised successful strategies to bypass the new formal constraints.

In this section we discuss three potential concerns regarding our results and interpretation.

First, a hiring pattern observationally equivalent to that obtained by schools recruiting scholars with a scientific quality similar to that of the incumbent faculty could be the result of a very different process. Suppose that candidates self-sort in the *concorsi* called by the schools closest to them in terms of scientific quality. If so, even if schools preferred to appoint the candidate with the highest standing to improve their quality, self-sorting would constraint their choice and schools would be forced to hire candidates that *de facto* replicate the *status quo*. If this were the case, however, in *concorsi* with double qualification, where we observe two applicants, we should systematically have that the winner's quality is higher than the second *idoneo* quality. In fact, only in one third of the cases H_1 exceeds H_2 as shown in Figure (3), where we report the distribution of $(H_1 - H_2)$ both before and after the reform.

Second, the lack of effect of the reforms that we find might be due to the countervailing influence of other policies. The only significant change occurred in the period under consideration was the rule enacted in 2009 that imposed of replacing at most 20 percent of all retirements in academia, a provision that significantly reduced the number of callable positions. However, this policy should have reinforced rather than offset the effect of the reform we analyze in this paper, given that it should have strengthen the competition among universities, being fewer positions available. Accordingly, the Pareto weight of the internal examiner should have decreased in contrast with the evidence we found.

Third, throughout the analysis we have maintained that all the characteristics of the examiners relevant for their preferences over candidates can be summarized by a unique index of scientific productivity, the h-index as computed on the Google Scholar dataset of publications. Even leaving aside the possible bias of the h-index as a measure of scientific productivity, there might be other individual characteristics, that influence the choice of the examiners, such as affinity in the field of research between candidates and examiners,

a scientific potential of candidates not yet captured by the h-index, the role played by the affiliation of the candidates and the role played by academic and extra-academic ties. Nonetheless, summary statistics fit well with model implications and make us confident that the possible specification errors we commit are not such that of invalidating our main conclusions.

8 Conclusions

We have analyzed the effect of two reforms of the recruitment system in Italian universities whose purpose was that of increasing meritocracy by making external examiners more independent and by reducing the opportunities of collusion within committees. We have argued that the purpose of these reforms contrasted with the fundamental incentive of large proportions of schools and incumbent scholars of maintaining the *status quo* and continuing to benefit of sizable rents.

To study the effect of the reforms, we have derived a theoretical model that represents the process of negotiation within evaluation committees in a way that accounts for the evidence that a non negligible share of *concorsi* ended up without appointment (failures). The theoretical model implies a simple decision rule that turns out to be a linear function of internal and external examiners qualities and allows to (partially) identify the Pareto weight of the internal examiner. The empirical counterpart of the decision rule has been estimated on the data from all *concorsi* called by the schools of economics in Italy and concluded between 2004 and 2011.

According to our estimations, neither the introduction of randomization of external examiners nor the reduction in the number of qualifications from two to one did significantly change the relative weight of the internal examiner on the evaluation committee.

The implication of our analysis is that procedural rules are little effective in modifying agents behaviors when they contrast with agents fundamental incentives. Changing procedural rules risks to load further administrative burden and costs without any substantial benefit. In the case of the recruitment system in Italian academia, reforms aiming to reward schools' scientific standing would be much more effective in making the adoption of meritocracy in schools' own interest.

Appendix A

In this appendix we characterize the solution of the negotiation problem between internal and external examiners.

We distinguish between two regimes, the appointment and the failure regime. In both cases we can have either an internal or a corner solution. Internal and corner solutions are linked to the structural parameters of the problem by different functional forms. As we are able to observe only what is the prevailing regime and not whether the solution is internal or corner, we need to determine under what conditions there is a one-to-one relation between observed regime and type of solution.

Assume $h_i > h_e$. The symmetric case is analogous.

Consider first the agreement regime.

The internal solution is

$$h_1^* = \frac{\alpha\gamma}{K}h_i + \frac{\beta(1-\gamma)}{K}h_e$$

$$h_2^* = \frac{(1-\alpha)\gamma}{1-K}h_i + \frac{(1-\beta)(1-\gamma)}{1-K}h_e$$

with

$$K = \gamma\alpha + (1-\gamma)\beta$$

The internal examiner utility at that solution is

$$U_i = -(h_i - h_c)^2\Gamma_i$$

the external examiner utility is

$$U_c = -(h_i - h_c)^2\Gamma_c$$

and the value of the problem is

$$W^A = -(h_i - h_c)^2 [\gamma\Gamma_i + (1-\gamma)\Gamma_e]$$

where

$$\Gamma_i = (1-\gamma)^2 \left[\frac{\alpha\beta^2}{K^2} + \frac{(1-\alpha)(1-\beta)^2}{(1-K)^2} \right]$$

$$\Gamma_e = \gamma^2 \left[\frac{\alpha^2\beta}{K^2} + \frac{(1-\alpha)^2(1-\beta)}{(1-K)^2} \right]$$

and

$$P = [\gamma\Gamma_i + (1-\gamma)\Gamma_e] = \gamma(1-\gamma) \left[\frac{\alpha\beta}{K} + \frac{(1-\alpha)(1-\beta)}{1-K} \right]$$

It is easy to check that $0 < \Gamma_i < 1$, $0 < \Gamma_e < 1$ and $0 < P < \gamma$.

The condition which separates the internal from the corner solution is $-(h_i - h_c)^2\Gamma_i \geq$

$-\bar{U}_i$ i.e. we observe an internal solution if

$$(h_i - h_c)^2 \leq \frac{\bar{U}_i}{\Gamma_i}$$

and a corner solution otherwise.

The corner solution satisfies the first order conditions of the Lagrangian

$$L = W - \lambda(-\bar{U}_i - U_i)$$

with $\lambda \geq 0$.

At the corner solution, the marginal rate of substitution between h_1 and h_2 needs to be equalized between the internal and the external examiners and, furthermore, the corner solution needs to satisfy the constraint $U_i(h_1, h_2) = -\bar{U}_i$. These two conditions are enough to derive the constrained maximizer (h_1^c, h_2^c) . The superscript c indicates the constrained solution. Moreover exploiting the positivity of the Lagrangian multiplier we obtain

$$h_1^c > \frac{\alpha\gamma}{K}h_i + \frac{\beta(1-\gamma)}{K}h_e$$

and

$$h_2^c > \frac{(1-\alpha)\gamma}{1-K}h_i + \frac{(1-\beta)(1-\gamma)}{1-K}h_e$$

Thus, for the parameter combinations such that the constraint is binding, the constrained solution is certainly preferable to the internal examiner compared to what he would have obtained if the constraint were not protecting himself. Indeed, the internal examiner, obtains $U_i^{Ac} = -\bar{U}_i$ regardless of his relative bargaining power and the values of h_i and h_e . Correspondingly, the utility of the external examiner, denoted U_e^{Ac} is certainly lower compared to what he could obtain were the constraint not active. The value of the problem is $W^{Ac} = \gamma(-\bar{U}_i) + (1-\gamma)U_e^{Ac}$ and note that it is certainly smaller than the value of the unconstrained problem for the same parameters profile.

Next, consider the failure regime.

In this case (h_1, h_2) are such that $U_i \leq -\bar{U}_i$. The internal examiner is protected and obtains $-\bar{U}_i$, but no qualified candidate will be appointed. This implies that the external examiner can freely decide about the candidates, however at the additional cost δ that represents the utility loss due to the fact that no one of his preferred candidates will be immediately appointed.

In this regime the internal solution corresponds to the bliss point of the external examiners, i.e.

$$h_1 = h_2 = h_e$$

At the internal solution $U_i < -\bar{U}_i$ (strictly smaller) and particularly $U_i(h_e, h_e) = -(h_i - h_e)^2$. Therefore, the condition for an internal solution can be expressed as

$$(h_i - h_e)^2 > \bar{U}_i$$

The corresponding value of the problem is

$$W^F = \gamma(-\bar{U}_i) + (1 - \gamma)(-\delta)$$

For parameters configuration such that the constraint is binding, i.e. when $(h_i - h_e)^2 \leq \bar{U}_i$, the solution will correspond to the point on the constraint $U_i = -\bar{U}_i$ that is closer to the external examiner bliss point. Note that since internal examiner's preferences do not play any role in this choice, the constrained solutions of the failure and the appointment regimes will not coincide in general. The value of the problem is

$$W^{Fc} = \gamma(-\bar{U}_i) + (1 - \gamma)(U_e^{Fc} - \delta)$$

where U_e^{Fc} is external examiner's payoff at the constraint solution. The value of W^{Fc} will be certainly smaller than the value of the unconstrained problem for the same parameters profile.

Finally, comparing the two regimes, we have:

- For $\bar{U}_i < (h_i - h_e)^2 < \frac{\bar{U}_i}{\Gamma_i}$, it is $W^A > W^F$ if

$$(h_i - h_e)^2 < \Delta = \frac{\gamma\bar{U}_i + (1 - \gamma)\delta}{P} \quad (18)$$

The quantity Δ belongs to the interval $[\bar{U}_i, \frac{\bar{U}_i}{\Gamma_i}]$ if

$$\bar{U}_i \geq \frac{\Gamma_i}{\Gamma_e} \delta \quad (19)$$

and otherwise it is larger than $\frac{\bar{U}_i}{\Gamma_i}$. In all cases either the appointment or the failure internal solution is optimal.

- for $(h_i - h_e)^2 > \frac{\bar{U}_i}{\Gamma_i}$ we compare the appointment regime's constrained solution with the failure regime's internal solution. If condition (19) holds, the latter certainly dominates. Indeed, for all $(h_i - h_e)^2 > \Delta$ we have that $W^F > W^A$ and thus it must also be that $W^F > W^{Ac}$.
- Similarly, for $(h_i - h_e)^2 < \bar{U}_i$, the solution of the problem is the agreement region's

internal solution. Indeed for all $(h_i - h_e)^2 < \Delta$ we have that $W^A > W^F$ and thus is must also be that $W^A > W^{Fc}$.

Summing up, if condition (19) holds, or, in words, if the cost of failure is sufficiently small, either the internal solution of the appointment regime or the internal solution of the failure regime prevails. Otherwise it is possible to observe, for some $(h_i - h_e)^2 > \frac{\bar{U}_i}{\Gamma_i}$ the appointment regime's constrained solution. Intuitively, when the cost of the failure is large, it cannot be compensated by the additional utility enjoyed by the external examiner in the failure regime, corresponding to the freedom of choosing his preferred candidates.

Appendix B - one qualification

When only one candidate is to be qualified, the model simplifies as follows.

$$U_i = -(h_i - h_1)^2$$

$$U_e = -(h_e - h_1)^2$$

and

$$W = \begin{cases} \gamma U_i + (1 - \gamma)U_e & U_i \geq -\bar{U}_i \\ \gamma(-\bar{U}_i) + (1 - \gamma)(U_e - \delta) & U_i < -\bar{U}_i \end{cases}$$

Also in this case we treat separately the appointment and the failure regime and we assume $h_i > h_e$.

In the appointment regime, which occurs when $(h_i - h_1)^2 \leq \bar{U}_i$, the internal solution is

$$h_1 = \gamma h_i + (1 - \gamma)h_e$$

and

$$U_i = -(1 - \gamma)^2(h_i - h_e)^2$$

$$U_e = -\gamma^2(h_i - h_e)^2$$

$$W^A = -\gamma(1 - \gamma)(h_i - h_e)^2$$

This regime prevails if

$$(h_i - h_e)^2 < \frac{\bar{U}_i}{(1 - \gamma)^2}$$

Otherwise, for $(h_i - h_e)^2 \geq \frac{\bar{U}_i}{(1 - \gamma)^2}$ the constrained solution occurs, which satisfies the Lagrangian $L = \gamma U_i + (1 - \gamma)U_e - \lambda(-\bar{U}_i - U_i)$. At the constrained solution h_1^c satisfies $(h_i - h_1^c)^2 = \bar{U}_i$ and it is larger than it would have been in the unconstrained

problem. This guarantees the internal examiner's utility to remain at $U_i = -\bar{U}_i$ at the expense of the external examiner utility denoted again U_e^c . The value of the problem is $W^{Ac} = \gamma(-\bar{U}_i) + (1 - \gamma)U_e^c$.

The failure regime occurs when $(h_i - h_1)^2 \geq \bar{U}_i$. The internal solution corresponds to the bliss point of the external examiner

$$h_1 = h_e$$

The value of the problem is

$$W^F = \gamma(-\bar{U}_i) + (1 - \gamma)(-\delta)$$

At the bliss point the constraint is slack and equals to $(h_i - h_e)^2 > \bar{U}_i$.

When $(h_i - h_e)^2 \leq \bar{U}_i$ the internal solution is not viable. At the constrained solution the external examiner does not obtain his bliss point. The selected candidate h_1^c is such that $(h_i - h_1^c)^2 = \bar{U}_i$ and the external examiner's utility is U_e^c .¹⁹ The value of the problem is

$$W^{Fc} = \gamma(-\bar{U}_i) + (1 - \gamma)(U_e^c - \delta)$$

Comparing the two regimes:

For $\bar{U}_i < (h_i - h_e)^2 < \frac{\bar{U}_i}{(1-\gamma)^2}$ we have $W^A > W^F$ when

$$(h_i - h_e)^2 < \Theta = \frac{\gamma\bar{U}_i + (1 - \gamma)\delta}{\gamma(1 - \gamma)} \quad (20)$$

The quantity Θ is internal to $\left[\bar{U}_i, \frac{\bar{U}_i}{(1-\gamma)^2}\right]$ for

$$\bar{U}_i > \frac{(1 - \gamma)^2}{\gamma^2} \delta \quad (21)$$

When condition (21) holds, either the internal solution of the appointment regime or the internal solution of the failure regime solve the maximization problem. Indeed, for $(h_i - h_e)^2 > \frac{\bar{U}_i}{(1-\gamma)^2}$, we have that $W^{Ac} < W^F$ (since W^{Ac} would be dominated by W^A if the constrained were not active) and for $(h_i - h_e)^2 < \bar{U}_i$, we have that $W^A > W^{Fc}$ (since W^{Fc} would be dominated by W^F if the constrained were not active).

Finally, note that the case with a single qualification can be derived from the case

¹⁹Compared to Appendix A, external examiner's utility at the constrained solution is the same in both regimes. This fact occurs because no substitution among candidates is possible in the problem with only one qualified candidate.

with two qualifications by setting $\alpha = \beta = 1$. Indeed, in this case, $\Gamma_i = (1 - \gamma)^2$, $\Gamma_e = \gamma^2$ and $P = \gamma(1 - \gamma)$.

Appendix C - elasticity of the threshold.

The small size of the samples at hand has forced us to pool together *concorsi* with elected and randomized committees. As a result, the empirical model estimated over the pooled sample assumes that the threshold separating agreement from failure is the same in both types of procedures. Since the threshold depends on γ , this restriction is questionable, unless the threshold is little sensitive to variations of γ .

In this appendix we compute the elasticity of $\sqrt{\Delta}$, (see equation 18), with respect to γ , denoted $\varepsilon_{\sqrt{\Delta}\gamma}$. Such elasticity is a rather complicate function of γ and of all other parameters. To simplify its expression, consider the following reparametrization (up to a small loss of generality). Let be $\alpha = \frac{1}{2} + \frac{m}{2}$, $\beta = \frac{1}{2} - \frac{m}{2}$ so that $m = \alpha - \beta$.

A linearization of $\varepsilon_{\sqrt{\Delta}\gamma}$ around $\gamma = \frac{1}{2}$, the midpoint of the domain of γ , yields

$$\varepsilon_{\sqrt{\Delta}\gamma}(\gamma) \simeq \left(\frac{1}{2} - \frac{\delta}{\bar{U}_i + \delta} \right) + 2 \left[(1 - m^2) + 2 \frac{\delta}{\bar{U}_i + \delta} \left(\frac{1}{2} - \frac{\delta}{\bar{U}_i + \delta} \right) \right] \left(\gamma - \frac{1}{2} \right) \quad (22)$$

Since at $\gamma = \frac{1}{2}$, it is $\bar{U}_i > \delta$ (by condition 19), both the intercept and the slope coefficient are positive.

In particular, at $\gamma = \frac{1}{2}$, the elasticity simplifies to

$$\varepsilon_{\sqrt{\Delta}\gamma} \left(\gamma = \frac{1}{2} \right) = \frac{1}{2} - \frac{\delta}{(\bar{U}_i + \delta)}$$

and it is bounded above by $\frac{1}{2}$.

Moreover, the elasticity is monotonically decreasing in δ and largest when $\delta = 0$. In this case, the least favorable, expression (22) becomes

$$\varepsilon_{\sqrt{\Delta}\gamma}(\gamma) \simeq \frac{1}{2} + 2(1 - m^2) \left(\gamma - \frac{1}{2} \right)$$

Hence, at $\gamma = \frac{3}{4}$ it is $\varepsilon_{\sqrt{\Delta}\gamma}(\frac{3}{4}) = 1 - \frac{1}{2}m^2$ and at $\gamma = \frac{1}{4}$ it is $\varepsilon_{\sqrt{\Delta}\gamma}(\frac{1}{4}) = \frac{1}{2}m^2$. We conclude that for a large range of values of γ , $\sqrt{\Delta}$ is inelastic.

The elasticity of $\sqrt{\Theta}$ - the corresponding threshold in the single qualification case (see equation 20) - behaves similarly.

Appendix D - truncation adjustment.

In this section we derive the adjustment to account for truncation in equations (14) and (15). Define:

$$\begin{aligned} I^* &= h_i - h_e = \theta_0 + (H_i - H_e) + \theta_1 GPS1(X) + \theta_2 GPS2(X) + \varepsilon \\ &= \Omega + \varepsilon \end{aligned}$$

Its expectation conditional on agreement is

$$E\left(I^* \mid -\sqrt{\Delta} \leq I^* \leq \sqrt{\Delta}\right) = \Omega + \Xi$$

where

$$\begin{aligned} \Xi &= E\left(\varepsilon \mid -\sqrt{\Delta} \leq I^* \leq \sqrt{\Delta}\right) = \\ &= \sigma \left[\phi\left(\frac{-\sqrt{\Delta}}{\sigma} - \frac{\Omega}{\sigma}\right) - \phi\left(\frac{\sqrt{\Delta}}{\sigma} - \frac{\Omega}{\sigma}\right) \right] / \left[\Phi\left(\frac{-\sqrt{\Delta}}{\sigma} - \frac{\Omega}{\sigma}\right) - \Phi\left(\frac{\sqrt{\Delta}}{\sigma} - \frac{\Omega}{\sigma}\right) \right] \end{aligned}$$

is the conditional expectation of $\varepsilon \sim N(0, \sigma^2)$, ϕ and Φ are the density and the cumulative distribution of a standard normal respectively.

In the pre-reform period, equation (14) is

$$\begin{aligned} H_1 - H_e &= \pi_{11}(H_i - H_e) + \\ &+ (\pi_{11} - 1)(\theta_0 + \theta_1 GPS1(X) + \theta_2 GPS2(X)) + (\pi_{11} - 1)\varepsilon \end{aligned}$$

and its expectation conditional on agreement is

$$\begin{aligned} E\left(H_1 - H_e \mid -\sqrt{\Delta} \leq I^* \leq \sqrt{\Delta}\right) &= \pi_{11}(H_i - H_e) + \\ &+ (\pi_{11} - 1)(\theta_0 + \theta_1 GPS1(X) + \theta_2 GPS2(X)) + (\pi_{11} - 1)\Xi \end{aligned} \quad (23)$$

Parameters of equation (23) can be estimated by means of the following adjusted regression

$$\widetilde{(H_1 - H_e)} = \pi_{10} + \pi_{11}\widetilde{(H_i - H_e)} + \pi_{13}GPS1(X) + \pi_{14}GPS2(X) + \xi \quad (24)$$

where

$$\widetilde{(H_1 - H_e)} = (H_1 - H_e) + \Xi \quad (25)$$

$$\widetilde{(H_i - H_e)} = (H_i - H_e) + \Xi \quad (26)$$

and ξ is a IID error term. Similarly we proceed after the reform, and for equation (15) and equation (16).

References

- Bagues, M.F. and Esteve-Volart, B. (2010) “Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment”, *Review of Economic Studies*, Vol. 77(4):1301-28.
- Chiappori, P., (1988) “Rational household labor supply”, *Econometrica*, 56(1):63-90.
- Combes, P., Linnemer, L., and Visser, M. (2008) “Publish or peer-rich? The role of skills and networks in hiring economics professors”, *Labour Economics* 15:423-441.
- De Paola, M. and Scoppa, V. (2011) “Gender Discrimination and Evaluators’ Gender: Evidence from the Italian Academy”, DES working paper 06/11
- Durante, R., Labartino, G. and Perotti, R. (2011) “Academic Dynasties: Decentralization and Familism in the Italian Academia”, NBER working paper 17572
- Harzing, A.W. (2007) “Publish or Perish”, available from <http://www.harzing.com/pop.htm>
- Harzing, A. and R. Van der Wai (2008) “Google scholar as a new source for citation analysis”, *Ethics in Science and Environmental Politics (ESEP)* 8 (1):61–73
- Hirsch, J. (2005). “An index to quantify an individual’s scientific research output”, *Proceedings of the National Academy of Sciences of the United States of America* 102 (46), 16569.4
- Imbens G. (2000), “The Role of the Propensity Score in Estimating Dose-Response Functions”, *Biometrika*, Vol. 87(3): 706-710.
- Perotti, R. (2002) “The Italian University System: Rules vs. Incentives”, presented at the First Conference on Monitoring, ISAE, Rome
- Zinovyeva, N. and Bagues, M. F. (2011) “Does gender matter for academic promotion? Evidence from a randomized natural experiment” IZA Discussion Paper 5537
- Zinovyeva, N. and Bagues, M. F. (2012) “It’s Not What You Know, but Who You Know: the Role of Connections in Academic Promotion”, paper presented at the EEA-ESEM Conference, Malaga, 2012

Tables & Figures

Table 1: *Concorsi* by year of the call

	Pre-reform		Post-reform		Total
	Single	Double	Single	Double	
2004	.	126	.	.	126
2005	24	115	.	.	139
2006	46	.	.	.	46
2007
2008	.	.	1	148	149
2009	.	.	12	20	32
Total	70	241	13	168	492

Table 2: Failures by type of *concorso*

	Pre-reform	Post-reform	Total
Double: Appointment	219	143	362
Double: Failure	22	25	47
Single: Appointment	56	10	66
Single: Failure	14	3	17
Total	311	181	492

Table 3: Summary statistics - Double qualification

	Elected Committees			Randomized Committees			Pooled Sample					
	Mean	Median	IQR	Obs.	Mean	Median	IQR	Obs.	Mean	Median	IQR	Obs.
Failure												
h index candidate	3.6	1.5	7.0	34	4.3	2.0	5.0	38	4.0	2.0	6.5	72
h index internal examiner H_i	2.2	1.5	1.8	34	4.0	4.3	4.1	38	3.1	2.2	3.7	72
h index external examiner H_e	3.9	1.3	4.3	34	4.2	2.8	6.0	38	4.0	1.6	4.6	72
$ H_i - H_e $	3.1	1.0	3.8	34	2.4	1.3	2.8	38	2.7	1.2	3.4	72
Appointment - Idoneo 1												
h index candidate	2.3	1.0	3.5	216	3.6	2.0	6.0	136	2.8	2.0	4.0	352
h index internal examiner H_i	2.4	1.8	3.1	216	3.7	2.7	5.0	136	2.9	2.3	3.4	352
h index external examiner H_e	2.5	1.5	3.0	216	3.1	2.0	3.5	136	2.7	1.8	3.5	352
$ H_i - H_e $	1.9	1.3	2.2	216	2.6	1.9	3.2	136	2.2	1.5	2.4	352
Appointment - Idoneo 2												
h index candidate	2.4	1.0	3.0	216	3.8	3.0	6.0	136	2.9	2.0	5.0	352
h index internal examiner H_i	2.4	1.8	3.1	216	3.7	2.7	5.0	136	2.9	2.3	3.4	352
h index external examiner H_e	2.5	1.5	3.0	216	3.1	2.0	3.5	136	2.7	1.8	3.5	352
$ H_i - H_e $	1.9	1.3	2.2	216	2.6	1.9	3.2	136	2.2	1.5	2.4	352
Total												
h index candidate	2.4	1.0	4.0	466	3.8	2.0	6.0	310	3.0	2.0	5.0	776
h index internal examiner H_i	2.4	1.7	3.0	466	3.7	2.7	4.8	310	2.9	2.3	3.6	776
h index external examiner H_e	2.6	1.5	3.0	466	3.2	2.0	3.5	310	2.8	1.8	3.5	776
$ H_i - H_e $	1.9	1.3	2.2	466	2.6	1.8	3.2	310	2.2	1.4	2.5	776
Observations	466				310				776			

Note: IQR stands for interquartile range.

Table 4: Summary Statistics - Single qualification

	Elected Committees				Randomized Committees				Pooled Sample			
	Mean	Median	IQR	Obs.	Mean	Median	IQR	Obs.	Mean	Median	IQR	Obs.
Failure												
h index candidate	2.4	1.5	2.0	10	9.5	9.5	7.0	2	3.6	2.0	5.5	12
h index internal examiner H_i	2.3	2.3	2.5	10	7.1	7.1	0.8	2	3.1	2.6	3.2	12
h index external examiner H_e	1.5	1.1	2.0	10	17.3	17.3	16.5	2	4.1	1.9	2.6	12
$ H_i - H_e $	1.0	1.0	1.0	10	10.1	10.1	17.3	2	2.6	1.1	1.5	12
Appointment												
h index candidate	2.4	1.0	3.0	55	3.2	2.5	4.0	10	2.5	1.0	3.0	65
h index internal examiner H_i	2.3	2.1	2.3	55	5.3	6.2	5.3	10	2.8	2.3	2.3	65
h index external examiner H_e	3.0	2.0	2.5	55	2.2	1.8	1.5	10	2.8	2.0	2.3	65
$ H_i - H_e $	1.8	0.9	1.3	55	3.4	2.5	4.9	10	2.0	0.9	1.3	65
Total												
h index candidate	2.4	1.0	3.0	65	4.3	3.0	6.0	12	2.7	1.0	4.0	77
h index internal examiner H_i	2.3	2.1	2.3	65	5.6	6.9	5.0	12	2.8	2.3	2.5	77
h index external examiner H_e	2.7	2.0	2.5	65	4.7	2.1	3.3	12	3.0	2.0	2.5	77
$ H_i - H_e $	1.6	0.9	1.2	65	4.5	2.6	4.9	12	2.1	0.9	1.3	77
Observations	65				12				77			

Note: IQR stands for interquartile range.

Table 5: Double qualification

Parameter	First Stage	Two Stage Procedure		Single Stage Procedure (OLS)	
		Second Stage Idoneo 2	Second Stage Idoneo 1 (Winner)	Idoneo 2	Idoneo 1 (Winner)
$\frac{\Delta_H}{\sigma}$	1.345*** (0.391)				
$\frac{\Delta_L}{\sigma}$	-2.677*** (0.413)				
$\frac{1}{\sigma}$	0.074** (0.036)				
$\frac{\theta_1}{\sigma}$	-0.211 (0.142)				
$\frac{\theta_2}{\sigma}$	-0.079*** (0.022)				
$\pi_{.1}$		0.535*** (0.179)	0.595*** (0.152)	0.527*** (0.092)	0.605*** (0.080)
$\pi_{.2}$		-0.071 (0.169)	0.142 (0.153)	-0.174 (0.122)	0.027 (0.121)
$(\pi_{.1}-1)\theta_1$		-0.206 (0.394)	0.100 (0.305)	-0.296* (0.161)	0.073 (0.162)
$(\pi_{.1}-1)\theta_2$		0.091** (0.045)	0.034 (0.037)	0.030 (0.025)	-0.003 (0.023)
σ	13.519** (6.563)				
Δ_H	18.190* (10.479)				
Δ_L	-36.192** (17.646)				
θ_1	-2.846 (1.902)	0.442 (0.920)	-0.246 (0.733)		
test θ_1 [p-val]		[0.063]	[0.128]		
θ_2	-1.067* (0.604)	-0.194* (0.093)	-0.083 (0.733)		
test θ_2 [p-val]		[0.109]	[0.081]		

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The upper panel reports estimates as obtained from modified probit (column 1), truncated regressions (columns 2 and 3), and OLS on the sub-sample of *concorsi* ended with an appointment (column 3). The lower panel reports the implied structural parameters. Standard errors of columns 1 to 3 are obtained by parametric bootstrapping the two stage procedure.

Table 6: Single qualification

Parameter	Two-Stage Procedure		Single-Stage Procedure (OLS)
	First Stage	Second Stage - Winner	Winner
$\frac{\Delta_H}{\sigma}$	0.950** (0.395)		
$\frac{\Delta_L}{\sigma}$	-3.780** (1.675)		
$\frac{1}{\sigma}$	0.178 (0.109)		
$\frac{\theta_1}{\sigma}$	-0.427* (0.230)		
$\frac{\theta_2}{\sigma}$	-0.084** (0.036)		
π_{s1}		0.458*** (0.162)	0.455*** (0.136)
π_{s2}		-0.165 (0.248)	-0.151 (0.302)
$(\pi_{s1} - 1)\theta_1$		0.039 (0.231)	-0.088 (0.203)
$(\pi_{s1} - 1)\theta_2$		-0.025 (0.031)	-0.059 (0.044)
σ	5.614 (3.437)		
Δ_H	5.332 (3.600)		
Δ_L	-21.221** (10.597)		
θ_1	-2.396* (1.226)	-0.072 (0.417)	
test θ_1 [p-val]		[0.064]	
θ_2	-0.470* (0.274)	0.047 (0.059)	
test θ_2 [p-val]		[0.058]	

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The upper panel reports estimates as obtained from modified probit (column 1), truncated regressions (column 2), and OLS on the sub-sample of *concorsi* ended with an appointment (column 3). The lower panel reports the implied structural parameters. Standard errors of columns 1 and 2 are obtained by parametric bootstrapping the two stage procedure.

Figure 1: Single and double qualifications overtime

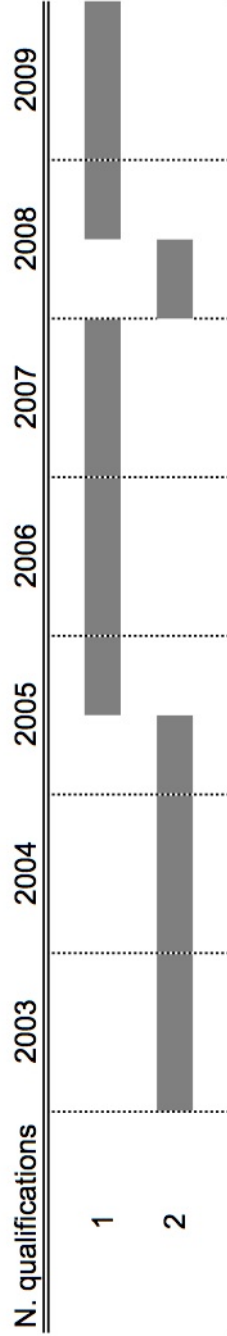


Figure 2: Distribution of $(H_i - H_e)$ net of observable characteristics of the *concorso*

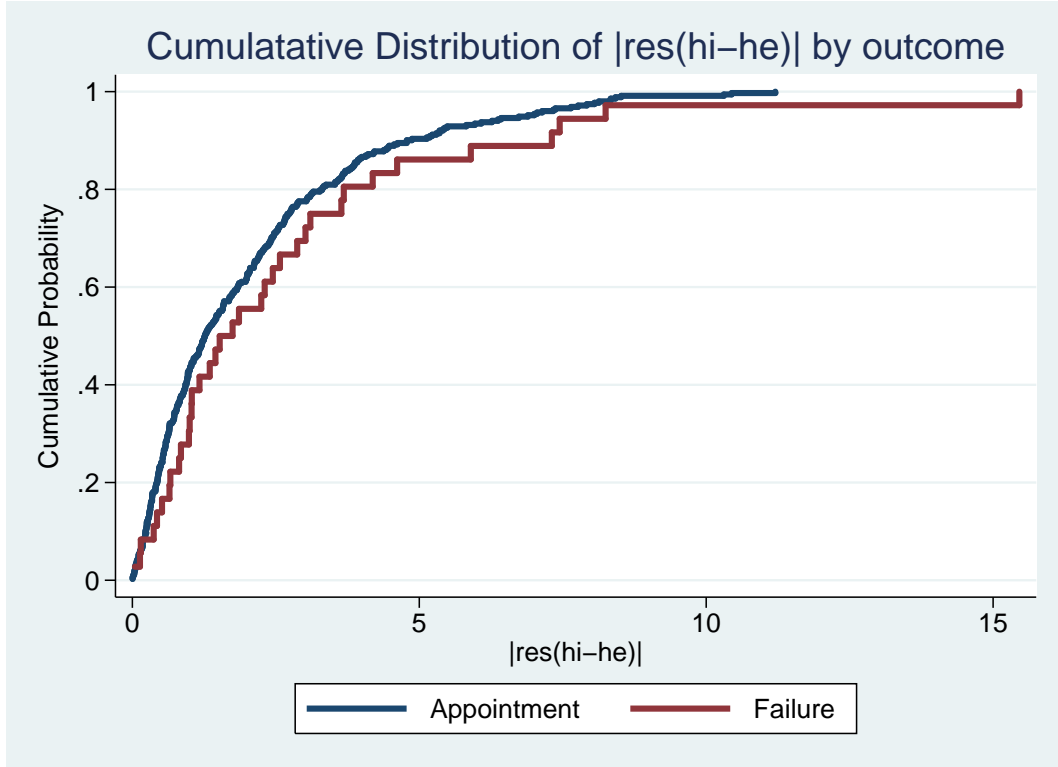


Figure 3: Distribution of $H_1 - H_2$

