UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Paper Economics

Cambridge Working Paper Economics: 1754

Performance-Based Rankings and School Quality

Claudia
Herresthal

9 June 2017

Updated 16 January 2018

I study students' inferences about school quality from performance-based rankings in a dynamic setting. Schools differ in location and unobserved quality, students differ in location and ability. Short-lived students observe a school ranking as a signal about schools' relative quality, but this signal also depends on the ability of schools' past intakes. Students apply to schools, trading off expected quality against proximity. Oversubscribed schools select applicants based on an admission rule. In steady-state equilibrium, I find that rankings are more informative if oversubscribed schools select more able applicants or if students care less about distance to school.

Performance-Based Rankings and School Quality*

Claudia Herresthal†

16th January 2018

Abstract

I study students' inferences about school quality from performance-based rankings in a dynamic setting. Schools differ in location and unobserved quality, students differ in location and ability. Short-lived students observe a school ranking as a signal about schools' relative quality, but this signal also depends on the ability of schools' past intakes. Students apply to schools, trading off expected quality against proximity. Oversubscribed schools select applicants based on an admission rule. In steady-state equilibrium, I find that rankings are more informative if more able applicants are given priority in admissions or if students care less about distance to school.

Keywords: performance-based rankings, observational learning, endogenous signal, selection effects, consumer choice

JEL codes: D83, I21, H75

*I would like to thank Meg Meyer, as well as Ricardo Alonso, Inga Deimen, Peter Esö, Pawel Gola, Johannes Hörner, Kohei Kawamura, Gilat Levy, Marco Ottaviani, Ines Moreno de Barreda, Anna Sanktjohanser, Alex Teytelboym, Collin Raymond, the conference and seminar participants at Econometric Society EWM, Econometric Society World Congress, Royal Economic Society Conference, York Symposium for Game Theory, Bonn, Cambridge, Cologne, Edinburgh and Oxford for valuable comments and feedback. I thank the Royal Economic Society for financial support.

†Queens' College, University of Cambridge, Silver Street, Cambridge CB3 9ET, UK, cl.herresthal@gmail.com.

1 Introduction

Performance-based school rankings are released in order to allow families to compare schools in terms of their quality. However, performance is partially determined by the ability of schools' past student intakes and families cannot perfectly observe which schools had the most able intakes in the past. Can families disentangle quality from the effects of student selection in the past? Similarly, treatment outcomes for medical procedures are published so that those seeking treatment can compare the quality of health care providers. Yet, treatment outcomes also depend on the prior health conditions of past patients, and it is often unclear which provider treated the sickest patients in the past.[1] What impact do policy choices have on such inference problems?

This paper studies inference problems in a dynamic infinite-horizon setting, using the example of school rankings. Students face a choice between two capacity-constrained schools of unknown quality. Each cohort of students observes a school ranking, but does not observe past student intakes. The probability that a school ranks high depends both on its quality and on the ability of its student intake in the previous cohort. If intakes across schools were equally able, then the better school would be more likely to rank high. In addition, if a school had a more able intake, its chances of ranking high would increase, irrespective of its quality. This leaves open the possibility that the worse school is *more* likely to rank high than the better school, provided the worse school's intake is sufficiently more able. This in turn implies that the ranking is a misleading indicator of relative quality for some student intakes. Suppose oversubscribed schools give priority to higher-ability applicants in their admissions. If students believe that the high-ranked school is more likely to be better, then the high-ranked school will have a larger applicant pool and admit a more able intake than the other school. Could we end up in a situation in which the worse school always ranks higher because its lower quality is covered up by a steady intake of able students? In the long run, would the chances that students identify the better school improve if oversubscribed schools were not allowed to give priority to high-ability applicants?

I show that, in the long run, students are more likely to infer which school is of better quality if the high-ranked school admits a more able intake rather than if each school admits an equally able intake. This may seem counterintuitive at first. After all, if intakes were equal the better school is more likely to rank high. By contrast, if intakes were not equal there is a risk that the worse school is the one more likely to rank high because it may have admitted a sufficiently more able intake than the better school. In addition, once the worse school is more likely to rank high than the better school, the worse school would also be more likely to attract a very able intake again and rank high again, implying that one misleading ranking is likely to be followed by another. However,

[1] Estimates of value-added have been published, but they do not eliminate the inference problem (Kane et al. [2002], Wilson and Piebalga [2008], Dranove et al. [2003]). In addition, estimates of schools' value-added are rarely used by students (Coldron et al. [2008], Imberman and Lovenheim [2013]).

this reasoning is flawed. What has been ignored is that if the better school happened to admit a more able intake, then its chances of ranking high would increase and, hence, the chances that the ranking accurately reflects schools' relative quality would improve. In this case, better quality and a more able intake both work in favour of the better school. Therefore, the better school is more likely than the worse school to maintain a high rank. Hence, in equilibrium, the better school is more likely to rank high than the worse school, and the ranking is more likely to reflect schools' relative quality accurately if in each cohort the high-ranked school admits a more able intake.

In the baseline model, in each of infinitely many periods, a continuum of short-lived students are matched with two infinitely-lived schools. Students differ in ability and location, schools differ in their location and quality.[2] School quality is unobserved. Students infer school quality from a ranking of schools. Their inference is complicated by the fact that the distribution from which the ranking is drawn depends on schools' most recent intakes, which are endogenously determined and also unobserved. Each student applies to one school. Schools are capacity-constrained. All applicants are accepted if there is sufficient capacity. Otherwise, applicants are selected based on an exogenously given admission rule which is characterised by what share of applicants is admitted based on ability rather than proximity. Each student derives a benefit from attending the better school and incurs a transport cost proportional to the distance to the school they attend. Applications are costless.

I perform comparative statics and show that students are better informed about the relative quality of schools if the distribution of transport costs shifts down (in the sense of first-order stochastic dominance). Recent school choice reforms can be represented by a decrease in transport costs, since they allowed students to attend non-local schools instead of requiring them to move into the school's catchment area. My findings show that school choice reforms interact with performance-based rankings to make it easier for students to identify better schools. In addition, my findings imply that further reducing barriers to choice, e.g. better transport links, facilitates learning about school quality over time. More generally, my result shows that learning about schools' relative quality, a form of vertical differentiation, improves as students perceive schools to be less horizontally differentiated, whether due to location, specialisation or other factors.

Furthermore, I show that students are better informed about the relative quality of schools if admission rules give priority to higher-ability applicants. Recent changes to the admission code in England have made it increasingly difficult for schools to select on ability,[3] yet the fact that this trend may hinder learning about school quality has not been pointed out. On the contrary, it has been argued that assigning more equal intakes across schools would make it more likely that

[2]Students are assumed to care about the intrinsic quality of the school, which reflects the quality of its teaching and leadership. In an extension, I incorporate that students care about their peers and show that this can further reinforce the effects identified in the baseline model.

[3]E.g. see Noden et al. [2014].

differences in school quality are reflected in rankings.[4] My paper is the first to develop a theoretical foundation for how priorities in admissions affect the informativeness of rankings over time and it is able to illustrate dynamic effects ignored by existing arguments. Using this framework, I show that interesting trade-offs may arise in the design of admission codes. Policymakers may want to increase the share of low-ability students at good schools, but if they make it easier for low-ability students to access high-performing schools then performance will become a less reliable indicator of quality over time, making it harder for policymakers to achieve the desired allocation in the future.

The framework developed in this paper is extended to study what role students' inference about school quality plays in the context of quasi-market reforms implemented in both the US and the UK. The idea behind these reforms is to mimic a market mechanism by linking a school's funding to demand for its places, thereby putting pressure on unpopular schools to improve or shut down. Clearly, these reforms are more effective at improving school quality if worse schools are less popular, but this crucially depends on how well students can identify which schools are worse. I incorporate this supply side response into my framework by assuming that the less popular school's quality is replaced by a new draw from an exogenously given quality distribution. This represents the fact that the undersubscribed school is under pressure to make changes, e.g. to experiment with new methodologies, or to replace their leadership, or in the worst case to close and reopen.[5] Since the popularity of schools is determined in equilibrium, so is overall school quality. I find that introducing such a policy improves the average quality of schools, and also improves access to good schools for students of any ability level.

However, in the presence of such policies, if i) an oversubscribed school selects a larger share of their intake based on ability or if ii) transport costs are lower, the average quality of schools improves further, but these additional benefits accrue only to high-ability students. This shows that the regulation of admission rules can affect the overall quality of schools, whereas most policy discussions assume admission rules have purely redistributive effects. In addition, if policymakers aim to improve access to good schools for low-ability students they face an efficiency cost. If they give low-ability students priority in the admissions at oversubscribed schools, then in the short run they improve the quality experienced by low-ability students and reduce the quality experienced by high-ability students, but in the long run they also reduce the overall quality of schools.

My paper relates to the literature on observational learning,[6] because it studies a setting in which the inference problem faced by agents in the current period is influenced by the choices of

[4] E.g. see Burgess and Allen [2010], p.10.

[5] Note that I assume that school quality does not necessarily improve as a consequence of interventions, e.g. a new headteacher is not guaranteed to be more effective than the incumbent headteacher, but more likely so the worse the incumbent. Therefore, the long-term effects of this policy are better the more likely it is that the relatively worse school is the one that is replaced.

[6] E.g. see Bikhchandani et al. [1992].

agents in past periods. However, in observational learning models, past agents' choices are directly observable and convey information about private signals received by these agents. My paper is the first to derive comparative statics when agents observe a limited window of realisations of a public signal, whose distribution depends on past agents' choices. Most work in the social learning literature assumes that agents observe the entire history of predecessors' choices. Lobel et al. [2007] model agents with a limited window of observation, but their focus is on conditions for convergence. Callander and Hörner [2009] propose a steady-state analysis but focus on agents inferring information from the relative frequency with which actions were taken by predecessors.

Many observational learning papers are interested in whether or not eventually all future agents will make the same choice (herd) and whether or not this choice is the one that yields the highest payoff. In the baseline model, I assume that the worse school is never certain to rank high even if it has admitted more able students than the other school. Therefore, a worse school cannot maintain a high rank forever, and it cannot happen that the majority of students will always apply to the worse school. In Section 9.2, I assume that if a school has admitted a sufficiently able intake then it is certain to rank high, irrespective of its quality. Then it is possible that the worse school maintains a high rank forever if an oversubscribed school admits a more able intake, but not if intake ability was equal across schools. However, from an ex-ante point of view, it is more likely that the better school will be the one which maintains a high rank forever.

In addition, my paper relates to models in which a decision-maker runs sequential tests before taking an action under uncertainty. Meyer [1991] studies a decision-maker (DM) who aims to learn which of two (non-strategic) workers is of higher ability. In each of a fixed number of periods, the DM can sequentially design biased contests. In the last period, the DM optimally assigns the bias in favour of the worker he believes to be of higher ability. The reason is that if this worker loses despite the contest being biased in his favour, then this is strong evidence that he is of lower ability. In my paper, learning is facilitated if the school of higher expected quality admits a more able intake and hence enjoys a bias in its favour. However, the reason why is different because students cannot condition their application choices on whether the better-performing school had better students or not. In addition, intakes are not assigned by a forward-looking DM, but by short-sighted students who do not take into account how their application choices will affect future rankings.

A key contribution of my paper is to derive a tractable model for the endogenous link between a school's rank and its pool of applicants, which has not been studied in the existing theoretical literature. Gavazza and Lizzeri [2007] study the impact of making information about school quality public, assuming that otherwise higher-ability students are more likely to be informed about which school is better than low-ability students. They find that the effect on student allocation differs depending on whether or not the school selects students based on ability. By contrast, in my framework, policymakers can never perfectly inform students about school quality, but only give them access to a performance-based ranking. Given that students' make inference about quality from past performance, I show that policymakers cannot choose the admission policy without also affecting how well informed students are about the quality of schools. De Fraja and Landeras [2006] study effects on attainment when students choose between schools based on rankings, but their focus lies on the incentives for schools to exert effort while the average intake ability at each school is assumed to vary exogenously with schools' relative performance. My paper also relates to the literature on matching algorithms, which usually designs the optimal algorithm assuming students have complete information about schools.[7] Yet if students are incompletely informed about the quality of schools then the student allocation today may influence future students' beliefs about schools' qualities and, hence, the preferences submitted to the algorithm. In addition, this insight has not been taken into account by empirical estimation strategies identifying how sensitive consumers' demand is to quality.[8]

My paper's predictions are consistent with recent empirical evidence. In the context of health care, Chandra et al. [2016] study patients' allocation to hospitals in the US over a period in which it became easier for patients to access information about hospital performance. They find that there is a correlation between a hospital's market share, its performance and its quality at a given point in time and that this correlation is growing over time, and that it is stronger when excluding emergency admissions, i.e. patients who had a lower (transport) cost of choosing between hospitals.[9] In the context of higher education, Hoxby [2009] shows that as the cost of long-distance communication and transportation have decreased over the past 60 years, students' choice of college has become less sensitive to the distance of a college from their home and meanwhile top US colleges have become more selective.

This paper outlines the model in Section 2. Section 3 introduces the inference problem in an illustrative example. Sections 4-6 solve for equilibrium and conduct comparative statics and welfare analysis. Section 7 extends the model to incorporate the supply side effects triggered by quasi-market reforms and Section 8 incorporate peer effects. Robustness checks in Section 9 analyse situations in which students observe a longer window of observations, performance can depend solely on intake ability or schools cannot select among applicants based on ability. Section 10 concludes.

[7] E.g. Abdulkadiroglu and Sönmez [2003]

[8] E.g. for education see Black [1999], Bayer and McMillan [2005], Burgess et al. [2015], Imberman and Lovenheim [2013], Hastings and Weinstein [2007], for health care see Gaynor et al. [2012].

[9] Performance is measured by clinical outcomes (survival/readmission) and quality is measured by the adherence to clinical guidelines.

2 Model

In each of infinitely many periods, a continuum of short-lived students are matched with two infinitely-lived schools. Students differ in ability and location, schools differ in their location and unobserved quality. Students observe a (ranking) signal about school quality drawn from an endogenously determined distribution, which depends both on schools' relative quality as well as on their most recent intake. Both school quality and past intakes are unobserved. Each student applies to one school. Schools are capacity-constrained. All applicants are accepted if there is sufficient capacity. Otherwise, applicants are selected based on an exogenously given admission rule. Each student's payoff depends on both the quality and proximity of the school to which they are admitted.

Setting: Time is discrete and the horizon is infinite $t = 0, 1, \ldots$. There is a continuum of students (the players) and two (non-strategic) schools.

Schools: Schools are infinitely-lived. One school is located at each end of a Hotelling line on $[0,1]$. I denote the school located at 0 by X and the school located at 1 by Y. Each school has a capacity of unit mass. An action for a school is to select among applicants based on an exogenously given admission rule as described below. The state of the world $\omega \in \left\{\omega^X, \omega^Y\right\}$ determines if school X is of better quality (ω^X) or if school Y is of better quality (ω^Y) where $Pr\left(\omega^X\right) = Pr\left(\omega^Y\right) = \frac{1}{2}$.

Students: Each student lives for one period only. A student's type is given by (λ, α), where $\lambda \in [0,1]$ denotes the student's location on the Hotelling line and $\alpha \in \{H, L\}$ denotes their ability, which is either high ($\alpha = H$) or low ($\alpha = L$). In each period, there is a unit mass of high-ability students and a unit mass of low-ability students. The distribution of location parameter λ is restricted only in that it is continuous, symmetric about $\lambda = \frac{1}{2}$ and independent of ability.[10] An action $a \in \{a_X, a_Y\}$ for a student is to choose whether to apply to school X ($a = a_X$) or to school Y ($a = a_Y$).

Rankings: Each period $t \geq 0$, a binary signal about school quality is realised, denoted by $W_t \in \left\{W_t^X, W_t^Y\right\}$. I will refer to signal W_t as a *ranking* of schools and refer to W_t^i as the event that school $i = X, Y$ is *the most recent winning school*. The probability that school X wins, denoted by $g\left(W_t^X | h_t^X, \omega\right) \in [0,1]$, is a function of $\left(h_t^X, \omega\right)$, where $h_t^X \in [0,1]$ denotes on the share of high-ability students at school X. Therefore, the signal realisation depends not only on the state of the world, i.e. on schools' qualities, but also on the endogenously determined allocation of students. The signal technology is independent of the school's location, i.e. it holds that for any $\kappa \in [0,1]$:

$$g\left(W_t^X | h_t^X = \kappa, \omega^X\right) = g\left(W_t^Y | h_t^Y = \kappa, \omega^Y\right). \tag{1}$$

In addition, $g\left(W_t^X | h_t^X, \omega^X\right)$ is independent of t, continuous and differentiable in h_t^X and satisfies

[10]Equilibrium existence and comparative statics results would be qualitatively unchanged if the distribution were not continuous. I assume continuity to simplify the exposition.

the following two properties. First, holding fixed the share of high-ability students enrolled at school X, then school X is more likely to win if it is the good rather than the bad school, i.e. for any $\kappa \in [0,1]$:[11]

$$g\left(W_t^X | h_t^X = \kappa, \omega^X\right) - g\left(W_t^X | h_t^X = \kappa, \omega^Y\right) > 0, \tag{2}$$

Second, a more able intake introduces an upward bias in performance. If school X is the better school ($\omega = \omega^X$), its chances of winning increase in the share of high-ability students enrolled at school X, i.e. for any $\kappa \in [0,1]$:

$$\frac{\partial}{\partial \kappa} g\left(W_t^X | h_t^X = \kappa, \omega^X\right) \geq 0, \tag{3}$$

and analogously if $\omega = \omega^Y$.

In the Online Appendix in Section B.4, I show the results of the baseline model hold as long as the distribution of signals satisfies the conditions outlined above. For the ease of exposition, I will conduct the analysis for the following specific functional form:

$$g\left(W_t^X | h_t^X, \omega^X\right) = \frac{s}{2} + \left(h_t^X - \frac{1}{2}\right) d, \tag{4}$$

and analogously if $\omega = \omega^Y$, where parameter $s \in (1,2]$ reflects the better school's performance advantage and parameter $d \in [0, 2-s)$ reflects the bias due to an imbalanced allocation of high-ability students. This functional form simplifies the exposition because the better school's performance advantage does not vary with the student allocation.[12]

Admission rule: A *local* student for school $i = X, Y$ is a student whose nearest school is school i. Each school admits all its applicants, unless a school has received more than a unit mass of applicants and is *oversubscribed.* If oversubscribed, a school uses the following rule: first, it admits all local high-ability applicants and a proportion $p \in [0,1]$ out of the pool of non-local high-ability applicants. Then it fills its remaining capacity with local low-ability applicants. If it has admitted all local low-ability applicants and still has spare capacity then it will prioritise high-ability over low-ability applicants.[13] Rejected applicants enrol at the other school.

Timing: Initially, Nature draws a state of the world.[14] In each period $t = 0, 1, ..$, students first

[11] Note that this assumption excludes the possibility that the worse school wins with probability 1, i.e. $Pr\left(W_{t-1}^Y | h_{t-1}^Y, \omega^X\right) = 1$. Section 9.2 studies the implications of allowing for this possibility.

[12] The specific functional form given by (4) satisfies property (2) since $s > 1$ and also satisfies property (3) since $d \geq 0$.

[13] In Section 9.3, I relax the assumption that an oversubscribed school can select applicants based on ability and show that the key insights still apply if high-ability students have a higher valuation for attending a better school than low-ability students.

[14] Equilibrium existence and comparative statics results would be qualitatively unchanged if the schools' relative

observe the ranking realisation in period $t-1$ (W_{t-1}) if $t>0$ and then each student applies to one school. Then schools admit applicants based on the admission rule. Next, students' payoffs are realised. Lastly, a ranking (signal) W_t is drawn.[15]

Information: The location of each school and students' types are public.[16] The state of the world is unobserved, but its distribution is known. Students in period $t>0$ do not observe any past actions or signals except for the most recent signal W_{t-1}.[17] Hence, for students in period $t>0$ an information set corresponds to signal realisation W_{t-1}. Denote the information set of students in period $t=0$ by $W_{-1}=\emptyset$. The structure of the game is common knowledge.

Strategies: A strategy for a student specifies if they apply to school X or school Y at any information set given their type and the current period. Denote the strategy of a student of type (λ,α) in period t by $\sigma(t,(\lambda,\alpha),W_{t-1})\in\{a_X,a_Y\}$ or $\sigma_{t,\lambda,\alpha}(W_{t-1})\in\{a_X,a_Y\}$.

Payoffs: Applications are costless and a student's payoff only depends on the school at which they enrol. Let the payoff of a student of type (λ,α) in period t in state of the world ω be denoted by $\pi(t,(\lambda,\alpha),\varepsilon,\omega)\in\mathbb{R}$ or $\pi_{t,\lambda,\alpha}(\varepsilon,\omega)\in\mathbb{R}$, where $\varepsilon\in\left\{\varepsilon^X,\varepsilon^Y\right\}$ and ε^i denotes the event that they enrol at school i. The student derives a constant benefit of $V>0$ if and only if they enrol at the better school.[18] In addition, they incur a cost equal to their distance from the school. Hence,

$$\pi_{t,\lambda,\alpha}(\varepsilon,\omega)=\begin{cases}V\cdot 1_{\omega=\omega^X}(\omega)-\lambda & \text{if } \varepsilon=\varepsilon^X\\ V\cdot 1_{\omega=\omega^Y}(\omega)-(1-\lambda) & \text{if } \varepsilon=\varepsilon^Y\end{cases}\tag{5}$$

where $1_{\omega=\omega^i}(\omega)=1$ if $\omega=\omega^i$ and 0 otherwise. Note that a student's realised payoff at any given enrolment and state of the world is independent of period t and independent of their ability α.

Let $E\left(\pi_{t,\lambda,\alpha}\left(\varepsilon^i\right)|W_{t-1}\right)$ denote the expected payoff of a student of type (λ,α) in period t if enrolled at school i conditional on observing signal W_{t-1}. Let $Pr\left(\varepsilon^i|\sigma_t,(\alpha,\lambda),W_{t-1},p\right)\in[0,1]$ denote the probability that a student of type (λ,α) in period t enrols at school i given the strategy profile is σ_t and schools' admission rule is characterised by p. Students choose their application strategy such as to maximise their expected payoff taking as given the strategy profile of other students in the same period and schools' exogenous admission rule, i.e.

$$\max_{\sigma_{t,\lambda,\alpha}} E\left(\pi_{t,\lambda,\alpha}|W_{t-1}\right)=\sum_{i=X,Y} Pr\left(\varepsilon^i|\sigma_t,(\alpha,\lambda),W_{t-1},p\right)E\left(\pi_{t,\lambda,\alpha}\left(\varepsilon^i\right)|W_{t-1}\right).\tag{6}$$

quality (state of the world) would change with some exogenously determined probability each period.

[15] In Section 7, I model a policy application in which students' application choices affect the distribution of school quality.

[16] Results are not affected by i) what information students have about their child's ability or by ii) what students observe about other students' types as long as the aggregate distribution of types is known.

[17] My results do not hinge on the assumption that only the most recent ranking is observed, as shown in Section 9.1.

[18] I assume students care about the intrinsic quality of the school, e.g. the quality of teaching and leadership. In Section 8, I incorporate that students care about the ability of their peers and show that this can further reinforce existing effects.

Since students will base their decision on the difference in their expected payoffs between applying to one school rather than the other, it will be convenient to define *transport cost* $c(\lambda) \in [0,1]$ as the additional distance a student in location λ needs to travel to reach their non-local compared to their local school, i.e. $c(\lambda) = |(1-\lambda) - \lambda| \in [0,1]$, with distribution $F(c)$.

Equilibrium Concept: The equilibrium concept is a Perfect Bayesian Equilibrium (PBE). A PBE is a strategy profile $\{\overline{\sigma}_t\}_{t\geq 0}$ and a system of beliefs $\{\mu_t(\omega|W_{t-1})\}_{t\geq 0}$, where beliefs map signal realisations into probability distributions over states of the world, such that i) in every period t for any type of student (λ,α), strategy $\sigma_{t,\lambda,\alpha} \in \overline{\sigma}_t$ is optimal given the profile of strategies of all other types of students in period t, $\widetilde{\sigma}_t = \overline{\sigma}_t \setminus \{\sigma_{t,\lambda,\alpha}\}$, and given their beliefs $\mu_t(\omega|W_{t-1})$ and ii) the system of beliefs $\{\mu_t(\omega|W_{t-1})\}_{t\geq 0}$ is derived from strategy profile $\{\overline{\sigma}_t\}_{t\geq 0}$ according to Bayes' rule. In particular, $\mu_t(\omega|W_{t-1})$ is the posterior belief based on the signal distribution of W_{t-1} constructed from the strategy profile of students in past periods $\{\overline{\sigma}_\tau\}_{0\leq\tau\leq t-1}$.

I am interested in the equilibrium as $t \to \infty$. In order to study this limit, I use the concept of steady-state equilibrium. A *steady state* consists of a time-invariant strategy profile $\{\overline{\sigma}\}_{t\geq 0}$ and a stationary distribution $Pr(W|\omega)$. A time-invariant strategy profile $\{\overline{\sigma}\}_{t\geq 0}$ implies that some constant share $h_t = \overline{h}$ of high-ability students will enrol at the most recent winning school in every period. Given $\{\overline{\sigma}\}_{t\geq 0}$, the stationary distribution solves

$$Pr\left(W^X|\omega^X\right) = \sum_{j=X,Y} g\left(W^X|h^j = \overline{h}, \omega^X\right) Pr\left(W^j|\omega^X\right). \tag{7}$$

A *steady-state equilibrium* is a steady state with a time-invariant profile $\{\overline{\sigma}\}_{t\geq 0}$ such that i) for any type of student (λ,α), strategy $\sigma_{\lambda,\alpha} \in \overline{\sigma}$ is optimal given the profile of strategies of all other types of students $\widetilde{\sigma} = \overline{\sigma} \setminus \{\sigma_{\lambda,\alpha}\}$ and given their beliefs $\mu(\omega|W)$ and ii) beliefs $\mu(\omega|W)$ are the posterior beliefs about the state of the world derived according to Bayes' rule given the stationary distribution $Pr(W|\omega)$. This implies that if the signal distribution happened to be equal to the stationary distribution in steady-state equilibrium in some period T, then there is no reason for students in periods $T+1, T+2, ...$ to deviate from the steady-state equilibrium strategy profile. I focus on steady-state equilibria in which the strategy profile is *symmetric with respect to schools' identities*, i.e. $\sigma_{t,\lambda,\alpha}\left(W^i_{t-1}\right) = \sigma_{t,1-\lambda,\alpha}\left(W^j_{t-1}\right)$ for $i = X,Y$ and $j \neq i$ for any t. This implies that in equilibrium students' application strategies only depend on their distance to the most recent winning school, irrespective of whether school X or school Y is the most recent winning school.

2.1 Remarks on Modelling Choices

Admission Policy - Schools' admission rule is characterised by parameter p and subject to regulation. The higher p, the more spaces at the oversubscribed school will be allocated based on ability rather than proximity.[19] If $p = 0$ then intake ability would be equal across schools, independent of students' application choices.[20] If $p = 1$ then an oversubscribed school would admit all high-ability applicants and, consequently, intakes are less balanced the more high-ability students have applied to the oversubscribed school.

Transport Costs - Transport costs capture that students perceive schools as horizontally differentiated along dimensions unrelated to quality. If all students were located at $\lambda = \frac{1}{2}$ then $c = 0$, and students would evaluate schools based on their relative quality only. If all students were evenly split between the end points, i.e. between $\lambda = 0$ and $\lambda = 1$, then $c(\lambda) = 1$ for all students. A negative shift in the sense of FOSD of $F(c)$ reduces the extent to which students perceive schools as horizontally differentiated. While proximity to home is an important school characteristic, students may also perceive schools as horizontally differentiated because schools specialise in different areas, e.g. humanities versus sciences, and then transport costs could also be thought of as representing students' preferences over the area of specialisation.

School Quality - School's intrinsic quality represents the quality of teaching and leadership, which have been shown to have a large impact on school performance (e.g. see Bloom et al. [2014]). In the baseline model, I assume students only care about this quality, but I relax this assumption in Section 8, where I incorporate that students care about the quality of their peers and show that this can strengthen the findings of the baseline model. In addition, I assume that school qualities are initially randomly allocated, consistent with the symmetry of my framework. The insights of the baseline model still apply if school qualities were to change exogenously over time. However, this is not necessarily the case if changes in school quality depend on the school's rank, as shown for the case of quasi-market reforms in Section 7.[21]

Rankings - The school ranking is drawn from a distribution which depends on schools' relative quality and on the allocation of high-ability students. For evidence that there is residual noise in rankings see Kane et al. [2002]. The property that the better school enjoys a strict performance advantage (see (2)) ensures that the worse school never wins with probability 1, even if it admitted all high-ability students. This assumption is crucial to prevent a "self-fulfilling" cycle of rankings. In Section 9.2, I relax this assumption and assume that if the share of high-ability students exceeds a certain threshold then a school will win with probability 1, irrespective of its quality. Then a form of herding can arise. The assumption that students observe rank-order information rather than continuous performance measures is motivated by the fact that rank-order information is easy

[19] For the results on learning about school quality, the important feature of the admission process is that an oversubscribed school takes at least as large a share of high-ability applicants as the other school (see also Section 9.3).

[20] Admission lotteries ($p = 0$) are commonly used by charter schools and by some magnet schools in the US and are becoming more common practice for oversubscribed UK state schools.

[21] School qualities may be endogenously determined for reasons other than quasi-market reforms. One may think that the likelihood of a quality change is exogenously determined but that the new quality depends on the school's rank, e.g. if higher-ranked schools can attract more capable headmasters or teachers. This would, unsurprisingly, strengthen the results about informativeness because a school with a better intake would also have higher expected quality.

to process, whereas making sense of the size of the difference in performance measures is relatively difficult. The fact that only the ranking based on the most recent performance can be observed is motivated by the fact that this ranking is the most easily accessible, e.g. in England this is the one circulated by the media, but my results do not hinge on this assumption as shown in Section 9.1.

3 Illustrative Example

The following example will illustrate how students in successive periods draw inferences about schools' relative quality from the performance-based ranking they observe, starting in period 0 when the first ranking is constructed. In particular, I will show that students in period 2 are better informed than students in period 1, despite the fact that students in each period only observe the most recent ranking. This example is useful to build intuition for how students draw inferences about school qualities in steady-state equilibrium, as analysed in detail in Section 4.2. In addition, it will highlight why analysing a steady-state equilibrium makes comparative statics more tractable. To this end, the example will use a simplified set-up.

There are two students in each period, one is of high ability and one is of low ability. Suppose there are no transport costs, i.e. $c = 0$. Each school has capacity for one student. If both students apply to the same school then the high-ability student is admitted with probability $\frac{1+p}{2}$, and the rejected student attends the other school. Denote by h_t^i the event that the high-ability student enrols at school i. If the high-ability student enrols at the better school, the better school wins with probability

$$g\left(W_t^X|h_t^X,\omega^X\right) = g\left(W_t^Y|h_t^Y,\omega^Y\right) = \frac{1}{2}(s+d). \tag{8}$$

In period 0 students have no access to rankings. They believe each school is equally likely to be better. Suppose each student picks a school at random to which they apply.

In period 1 students observe which school won based on the allocation of students in period 0, i.e. they observe W_0. To form a belief about which distribution this signal is drawn from conditional on the state of the world, they need to form a conjecture about the student allocation in period 0. Consistent with the strategies of students in period 0, they conjecture that the high-ability student in period 0 was equally likely to enrol at the better or the worse school for any given state of the world, i.e.

$$Pr\left(h_0^X|\omega^X\right) = Pr\left(h_0^Y|\omega^Y\right) = \frac{1}{2}. \tag{9}$$

Based on this conjecture, students in period 1 infer that the most recent winning school, W_0, was drawn from the distribution which assigns the following probability to the better school winning:

$$
\begin{aligned}
Pr\left(W_0^X|\omega^X\right) =& g\left(W_0^X|h_0^X,\omega^X\right)Pr\left(h_0^X|\omega^X\right) \\
&+ g\left(W_0^X|h_0^Y,\omega^X\right)Pr\left(h_0^Y|\omega^X\right) \\
=& \frac{1}{2}\left[\frac{1}{2}(s+d)\right]+\frac{1}{2}\left[\frac{1}{2}(s-d)\right]=\frac{s}{2}\geq\frac{1}{2}
\end{aligned}
\tag{10}
$$

and $Pr\left(W_0^Y|\omega^Y\right)=Pr\left(W_0^X|\omega^X\right)$. And hence, they update their beliefs that the most recent winning school, W_0, is the better school as follows:

$$
\begin{aligned}
Pr\left(\omega^X|W_0^X\right) =& \frac{Pr\left(W_0^X|\omega^X\right)Pr\left(\omega^X\right)}{Pr\left(W_0^X|\omega^X\right)Pr\left(\omega^X\right)+\left(1-Pr\left(W_0^X|\omega^X\right)\right)Pr\left(\omega^X\right)} \\
=& Pr\left(W_0^X|\omega^X\right)=\frac{s}{2}\geq\frac{1}{2}
\end{aligned}
\tag{11}
$$

and analogously for $Pr\left(\omega^Y|W_0^Y\right)$.

As students in period 1 infer that the most recent winning school, W_0, is more likely to be the better school, both students will apply to W_0.

In period 2 students observes which school that won in period 1, W_1, but not which school won in period 0, W_0. Although they can conjecture where students in period 1 applied conditional on W_0, they cannot infer whether school X or school Y enrolled the high-ability student in period 1 due to the fact that they do not observe whether school X or school Y was W_0 . However, they can infer how likely it is that the high-ability student in period 1 enrolled at the better school, consistent with the strategy profile of students in the previous two periods:

$$
\begin{aligned}
Pr\left(h_1^X|\omega^X\right) &= Pr\left(W_0^X|\omega^X\right)\left(\frac{1+p}{2}\right)+Pr\left(W_0^Y|\omega^X\right)\left(\frac{1-p}{2}\right) \\
&= \frac{1+p(s-1)}{2}\geq\frac{1}{2}
\end{aligned}
\tag{12}
$$

and $Pr\left(h_1^Y|\omega^Y\right)=Pr\left(h_1^X|\omega^X\right)$. Based on this conjecture, they conclude that that the most recent winner, W_1, was drawn from the distribution which assigns the following probability to the better school winning:

$$
Pr\left(W_1^X|\omega^X\right)=Pr\left(W_1^Y|\omega^Y\right)=\frac{s+dp(s-1)}{2}\geq\frac{s}{2}
\tag{13}
$$

And hence, they update their beliefs that the most recent winner, W_1, is the better school as follows:

$$
Pr\left(\omega^X|W_1^X\right)=Pr\left(W_1^X|\omega^X\right)=\frac{s+dp(s-1)}{2}\geq\frac{s}{2}
\tag{14}
$$

and analogously for $Pr\left(\omega^Y|W_1^Y\right) = Pr\left(\omega^X|W_1^X\right)$.

I can now make the key comparison between posterior beliefs held by students in period 2 and period 1, i.e. between (11) and (14). Students in period 2 are better informed than students in period 1, despite the fact that in each period students only have access to the most recent ranking. The reason is that the ranking observed in period 2, relative to the ranking observed in period 1, is more likely to have been generated in a situation in which the high-ability student attended the better school. This is the case because students in period 1 had better information about school quality than students in period 0. The same logic applies to students in future periods, and therefore the inference about quality will improve over time.

In the remainder of the paper, I will focus on a steady state in which the application strategy profile of students is constant across periods. In equilibrium, students' posterior beliefs that the higher-ranked school is the better of the two schools is also constant. To construct students' posterior beliefs in such a steady state, it is not necessary to reason through choices of students in each previous period as in the example above. Instead, it is possible to directly derive the stationary probability that a better school wins and then construct students' posterior beliefs about how likely it is that the school that they observe winning is the better of the two schools.

4 Steady-State Equilibrium

This section will solve for a steady-state equilibrium. I will show that, in equilibrium, students believe that the most recent winning school is more likely to be the better school and students apply to the most recent winning school if and only if i) it is their local school or ii) it is their non-local school and their transport costs fall below a cut-off level.

Definition 1: *Given the most recent winning school W_{t-1},* ***mobility in period t,*** *$m_t \in [0,1]$, is defined as the share of non-local students who apply to this school.*

There is a one-to-one map between the strategy profile of students in period t and the level of mobility m_t (as illustrated in Figure 1). In an abuse of terminology, I will refer to the level of mobility m_t as the strategy profile in period t.

I will solve for steady-state equilibrium mobility m^* in three steps. First, I will take as given posterior beliefs about the state of the world held by students in period t and derive the optimal strategy profile $\overline{m}_t$ in terms of these beliefs. Second, I will take as given a time-invariant strategy profile, $m_t = \widehat{m}$ for all t, and derive the stationary distribution, $Pr\left(W|\omega^i\right)$, in steady state for $i = X,Y$. Given this stationary distribution, I will compute (time-invariant) posterior beliefs about the state of the world in terms of $\widehat{m}$. Finally, I will solve for fixed points, i.e. I will find a time-

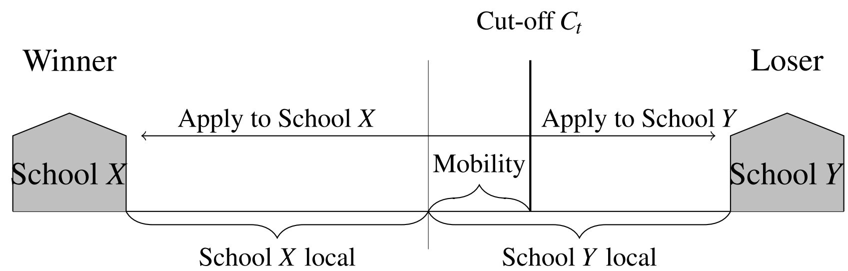

Figure 1: Students are distributed along a line between school X and school Y. Those with high transport costs live closer to their respective local school. The figure depicts the situation in which school X is the most recent winner and therefore receives applications from all local students, and from the share of non-local students whose transport costs lie below the cut-off C_t. Mobility is defined as the share of students who apply to their non-local winning school.

invariant strategy profile m^* such that m^* is optimal given posterior beliefs and posterior beliefs are based on the stationary distribution in steady state, i.e. $m^* = \overline{m}_t = \widehat{m}$.

A fixed point always exists since the optimal mobility level $\overline{m}_t$ increases in the posterior belief that the most recent winner is the better school and this posterior belief increases in steady-state mobility level $\widehat{m}$. However, the fixed point is not necessarily unique. At the end of the section, I will motivate my selection of the smallest fixed point by showing that this corresponds to the limit of the sequence of (non-steady-state) equilibrium strategy profiles $\{m_t\}_{t\geq 0}$ as $t \to \infty$, assuming no ranking is available to students in period $t = 0$, as in the example in Section 3.

4.1 Optimal Mobility

First, I will derive optimal mobility given beliefs. Suppose students in period t believe that the most recent winning school is weakly more likely to be the better of the two schools and their beliefs are independent of whether this is school X or Y. It is helpful to introduce informativeness I_t as a shortcut for how these beliefs enter students' expected payoff in period t.[22]

Definition 2: ***Informativeness in period t***, $I_t \in [0,1]$, *is defined as*

$$I_t \equiv Pr\left(\omega^i|W^i_{t-1}\right) - Pr\left(\omega^j|W^i_{t-1}\right) \tag{15}$$

for $i,j = X,Y$ *and* $i \neq j$, *where* $Pr\left(\omega^i|W_{t-1}\right)$ *denotes the posterior beliefs held by students in period t that school i is the better school conditional on the realisation of the most recent winning school.*

[22] Off the equilibrium path, $I_t \in [-1,1]$ and if $I_t < 0$ then $\overline{m}_t = F\left(|V \cdot I_t|\right)$ would refer to the share of students who apply to their non-local losing school.

Lemma 1 [Optimal mobility given Informativeness]

The optimal strategy profile of students in period t given informativeness I_t is given by

$$\overline{m}_t = F\left(V \cdot I_t\right). \tag{16}$$

Proof of Lemma 1: There is no downside for a student to apply to the school at which their expected payoff conditional on enrolment is higher.[23] In the worst case their application gets rejected and they enrol at the other school, which results in the same expected payoff as if they had applied there.[24] Hence, a student in period t of type (λ, α) applies to school X if and only if the expected payoff conditional on enrolment is higher:

$$E\left(\pi_{t,\lambda,\alpha}\left(\varepsilon_t^X\right)|W_{t-1}\right) - E\left(\pi_{t,\lambda,\alpha}\left(\varepsilon_t^Y\right)|W_{t-1}\right) \geq 0, \tag{17}$$

where ε_t^i denotes the event that the student enrols at school i.

Suppose $W_{t-1} = W_{t-1}^X$. Then by enrolling at school X rather than at school Y a student derives a benefit V if school X is better, but loses V if school X is worse. In addition, they incur the additional transport costs $c(\lambda)$ if school X is their non-local school, but save these costs if school X is local:

$$\begin{aligned} &E\left(\pi_{t,\lambda,A}\left(\varepsilon_t^X\right)|W_{t-1}^X\right) - E\left(\pi_{t,\lambda,A}\left(\varepsilon_t^Y\right)|W_{t-1}^X\right) \\ &= \begin{cases} V \cdot I_t + c(\lambda) & \text{if } \lambda \leq \frac{1}{2} \\ V \cdot I_t - c(\lambda) & \text{if } \lambda > \frac{1}{2} \end{cases} \end{aligned} \tag{18}$$

and analogously for $W_{t-1} = W_{t-1}^Y$.

Therefore, any student for whom school W_{t-1} is local will apply to W_{t-1}. However, a student for whom school W_{t-1} is non-local will apply if and only if their transport cost $c(\lambda)$ satisfy $C_t \geq c(\lambda)$ where the cut-off C_t is defined as follows:

$$C_t \equiv V \cdot I_t. \tag{19}$$

Hence, the share of non-local students who apply to the most recent winning school is

$$\overline{m}_t = F\left(C_t\right) = F\left(V \cdot I_t\right). \square \tag{20}$$

[23]This implies that a student's optimal strategy is independent of the strategy of other students in the same period.

[24]Note that a non-local low-ability applicant is indifferent between applying to either school, but for all other types it is a strictly dominant strategy to apply to the school with a higher expected payoff conditional on enrolment.

Mobility is a measure of how responsive application choices are to informativeness in equilibrium. If students believed that each school was equally likely to be better, $I_t = 0$, then each student would apply to their local school, $m_t = 0$. If students believed that the period-t-1 winning school is likely to be better, $I_t > 0$, then some students face a trade-off between attending the closer school and attending the school of higher expected quality. Mobility measures the share of students who choose to apply to the school of higher expected quality.[25]

4.2 Steady-state Informativeness

Next, I will derive informativeness in steady state given a time-invariant strategy profile $m_t = \widehat{m}$ for all t.

Lemma 2 [Informativeness given time-invariant Mobility]
The unique steady-state level of informativeness given mobility $m_t = \widehat{m}$ is

$$I(\widehat{m}) = \frac{s-1}{1-p \cdot d \cdot \widehat{m}} \geq 0. \tag{21}$$

Proof of Lemma 2: The sequence of signal realisations $\{W_t\}_{t>0}$ follows a time-homogeneous Markov process. It is a Markov process because the distribution $g\left(W_t | h_t^i, \omega\right)$ from which W_t is drawn depends on the share h_t^i of high-ability students at school i, which in turn depends on only the most recent signal realisation W_{t-1} for any given mobility level m_t. Given that mobility is time-invariant, the process is time-homogeneous and the share of high-ability students at the most recent winning school is $h_t = \frac{1+p\widehat{m}}{2}$ for every t. If $s < 2$, the process is irreducible and hence has a unique stationary distribution which for any t solves:

$$\begin{aligned}
Pr\left(W^X | \omega^X\right) =& g\left(W^X | h^X = \frac{1+p\widehat{m}}{2}, \omega^X\right) Pr\left(W^X | \omega^X\right) + \\
& g\left(W^X | h^X = \frac{1-p\widehat{m}}{2}, \omega^X\right) \left[1 - Pr\left(W^X | \omega^X\right)\right] \\
=& \left[\frac{s}{2} + \left(\frac{1+p\widehat{m}}{2} - \frac{1}{2}\right) d\right] Pr\left(W^X | \omega^X\right) + \\
& \left[\frac{s}{2} + \left(\frac{1-p\widehat{m}}{2} - \frac{1}{2}\right) d\right] \left[1 - Pr\left(W^X | \omega^X\right)\right].
\end{aligned} \tag{22}$$

[25] That students tend to prefer higher-performing schools has been shown by Burgess et al. [2015].

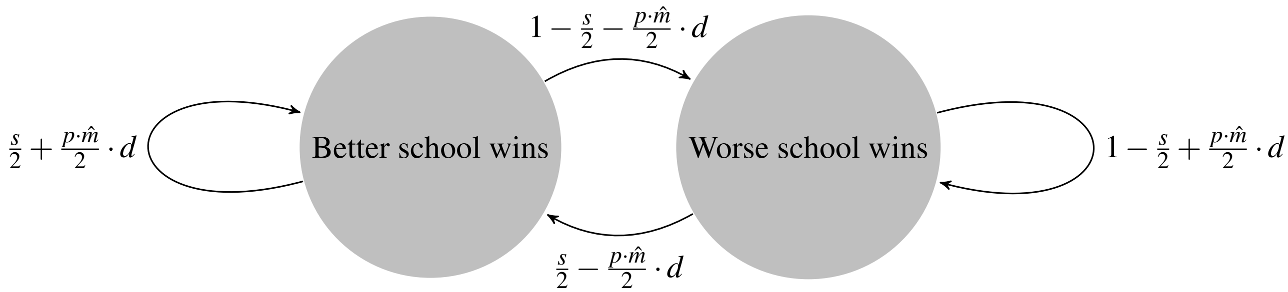

Figure 2: The graph illustrates the dynamics of the Markov process of ranking realisations. The arrows show the possible transitions between ranking realisations and the likelihood with which they occur, given time-invariant mobility level $\widehat{m}$.

and analogously for $Pr\left(W^Y|\omega^Y\right)$. Hence,

$$Pr\left(W^X|\omega^X\right)=Pr\left(W^Y|\omega^Y\right)=\frac{s-p\cdot d\cdot\widehat{m}}{2\left(1-p\cdot d\cdot\widehat{m}\right)}. \tag{23}$$

By Bayes' rule and by the symmetry of the setting:

$$Pr\left(\omega^X|W^X\right)=\frac{Pr\left(W^X|\omega^X\right)Pr\left(\omega^X\right)}{Pr\left(W^X|\omega^X\right)Pr\left(\omega^X\right)+\left[1-Pr\left(W^X|\omega^X\right)\right]Pr\left(\omega^X\right)}=Pr\left(W^X|\omega^X\right) \tag{24}$$

and analogously for $Pr\left(\omega^Y|W^Y\right)$. The level of informativeness follows. If $s=2$ then $d=0$ and $I=1$. $I\geq 0$ since $s>1$ and $p,d,\widehat{m}\in[0,1]$.$\square$

Informativeness strictly increases in the level of steady-state mobility $\widehat{m}$ if an oversubscribed school selects a strictly positive proportion of non-local high-ability applicants ($p>0$) and if the chances of winning strictly increase in the share of high-ability students ($d>0$). It is true that if steady-state mobility is positive ($\widehat{m}>0$), then it is possible that the student allocation is such that the worse school is more likely to rank high than the better school, whereas this is never the case if steady-state mobility is zero ($\widehat{m}=0$). Nevertheless, as I will argue below, the likelihood that the better school wins increases in steady-state mobility $\widehat{m}$, and, therefore, informativeness increases in steady-state mobility $\widehat{m}$.[26]

Consider the transition probabilities of the Markov process between the signal realisation that the the better school wins and the signal realisation that the worse school wins given time-invariant mobility level $\widehat{m}$ (illustrated in Figure 2). At $\widehat{m}=0$, the probability that the better school wins in period t is independent of the signal realisation in period $t-1$, and the better school is strictly more likely to win than the worse school since $s>1$. By contrast, at $\widehat{m}>0$, the signal realisation in period t depends on the signal realisation in period $t-1$. As mobility $\widehat{m}$ rises, it is more likely

[26]This corresponds to the dynamics illustrated in the example in Section 3.

that the signal realisation in period t is the same as the signal realisation in period $t-1$, provided $p > 0$ and $d > 0$. Importantly, the chances that the signal realisation is the same across two periods is independent of whether the better or the worse school won in period $t-1$. This implies that not only a truthful ranking (better school wins) but also a misleading ranking (worse school wins) is more likely to be repeated. Yet, an increase in mobility raises the stationary probability that the better school wins, since the better school is strictly more likely to win than the worse school at any given level of mobility $\widehat{m}$.[27]

4.3 Fixed Point

I characterise a steady-state equilibrium by the fixed point of mobility m^* such that mobility m^* is optimal given posterior beliefs and posterior beliefs are based on the signal distribution in steady state at mobility m^* (as illustrated in Figure 3):

$$m^* = F\left(V \cdot I\left(m^*\right)\right). \tag{25}$$

Proposition 1 [Equilibrium]: *A steady-state equilibrium level of mobility m^* is characterised by*

$$m^* = F\left(V \cdot \frac{s-1}{1-p \cdot d \cdot m^*}\right) \tag{26}$$

and the corresponding steady-state equilibrium level of informativeness $I\left(m^*\right)$ is given by

$$I\left(m^*\right) = \frac{s-1}{1-p \cdot d \cdot m^*}. \tag{27}$$

Such a steady-state equilibrium level of mobility (and informativeness) always exists.

Proof of Proposition 1: $I\left(\widehat{m}\right)$ is increasing in $\widehat{m}$ since $s > 1$, $p, d > 0$. Then $F\left(V \cdot I\left(\widehat{m}\right)\right)$ is monotone increasing in $\widehat{m} \in [0,1]$ since $V > 0$ and $F\left(\cdot\right)$ is increasing. By Tarski's fixed point theorem, there exists an m^* such that $F\left(V \cdot I\left(m^*\right)\right) = m^*$.□

4.4 Convergence to Steady-State Equilibrium

In the remainder of the paper, I will focus on the smallest steady-state equilibrium level of mobility and informativeness. This is a natural choice because the sequence of (non-steady-state)

[27] Note that I assume that the worse school never wins with probability 1, which implies that the worse school winning is never an absorbing state. For a discussion about how results are affected when this assumption is relaxed see Section 9.2.

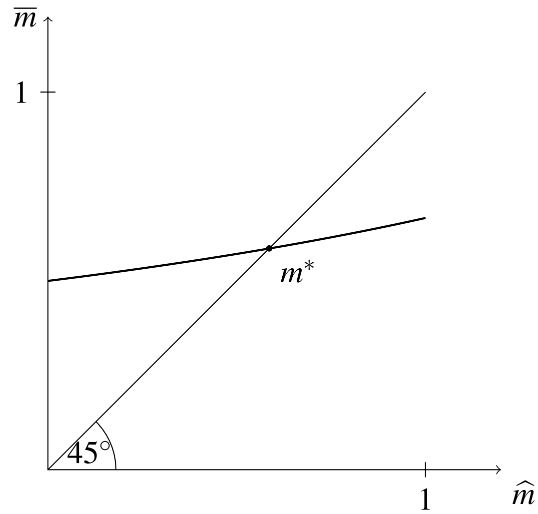

Figure 3: The graph shows the optimal mobility level $\overline{m}$, which characterises the optimal strategy profile of students in steady state, as a function of steady-state mobility level $\widehat{m}$. The intersection with the 45-degree line shows the steady-state equilibrium level of mobility m^*. This graph is drawn assuming F is a Uniform distribution on $[0,1]$, $V=1$, $\theta=\frac{1}{2}$, $s=\frac{3}{2}$, $p=d=\frac{1}{2}$.

equilibrium mobility levels $\{m_t\}_{t\geq 0}$ converges to the the smallest steady-state equilibrium mobility level as $t\to\infty$, given that no ranking is available to students in period 0. The dynamics in a non-steady-state equilibrium are similar to those illustrated in the example in Section 3.

Proposition 2 [Convergence] *Consider the sequence $\{I_t\}_{t\geq 0}$ defined by informativeness in period* 0*,*

$$I_0 = 0, \tag{28}$$

and by the following recurrence equation for informativeness in all periods $t>0$,

$$I_t = s - 1 + F\left(V\cdot I_{t-1}\right) p\cdot d\cdot I_{t-1}. \tag{29}$$

The sequence $\{I_t\}_{t\geq 0}$ corresponds to the sequence of levels of informativeness in non-steady-state equilibrium. As $t\to\infty$, $\{I_t\}_{t\geq 0}$ converges to the smallest steady-state equilibrium level of informativeness, denoted by I^{1*}. As $\{I_t\}_{t\geq 0}$ converges so does the sequence of mobility levels $\{m_t\}_{t\geq 0}$. The sequence $\{m_t\}_{t\geq 0}$ converges to the smallest steady-state equilibrium level of mobility, denoted by m^{1*}, where

$$m^{1*} = F\left(V\cdot I^{1*}\right). \tag{30}$$

See Appendix for proof. In period 0, students do not observe a ranking of schools and, hence, believe that each school is equally likely to be better. In equilibrium, students in any period $t>0$

know the distribution from which the ranking in period $t-1$ is drawn (see example in Section 3). The ranking distribution in period t depends only on ranking distribution in period $t-1$ because students' optimal strategy profile depends only on the ranking realisation in period $t-1$ (Markov property). The sequence of levels of informativeness that arises in equilibrium is increasing and will converge as $t \to \infty$. Starting from a level of informativeness equal to zero, such a sequence will converge to the smallest steady-state equilibrium level of informativeness. As steady-state equilibrium levels of informativeness and mobility are jointly ordered, this also implies that the sequence of mobility levels will converge to the smallest steady-state equilibrium level of mobility.[28]

5 Comparative statics

This section will analyse how both the informativeness of rankings and the share of high-ability students at the better school vary across different environments. The following analysis will focus on the steady-state equilibrium associated with the smallest level of mobility (m^{1*}), henceforth equilibrium level of mobility (see Section 4.4). For the remainder of the paper, all proofs can be found in the Appendix.

Theorem 1 [Comparative Statics]

1. *An increase in the proportion p of non-local high-ability applicants admitted by an oversubscribed school, or*
2. *a negative shift in the sense of FOSD of the distribution $F(c)$ of transport costs, or*
3. *an increase in the performance advantage d due to a more able intake, or*
4. *an increase in the performance advantage s due to superior school quality*

- *increases the equilibrium level of mobility and the informativeness of rankings, and*
- *increases the share of high-ability students attending the better school.*

The first three exogenous changes increase the chances that the most recent winning school wins again. Importantly, these changes increase the chances that the most recent winning school wins again by an equal amount whether the most recent winner is the good school or the bad school. Therefore, it may at first seem as if these changes would not make the ranking more informative. However, by the same argument used to show that steady-state informativeness increases in the

[28] For any time-invariant mobility level, the sequence of levels of informativeness converges. However, it may not converge to a steady-state equilibrium level.

time-invariant mobility level in Section 4.2, these changes increase the likelihood that the better school wins in steady state and, consequently, informativeness increases. In addition, an increase in the performance advantage s due to superior quality raises informativeness independent of student allocations. As informativeness increases at any given level of steady-state mobility, optimal mobility rises and hence equilibrium mobility rises (as shown in Figure 4).

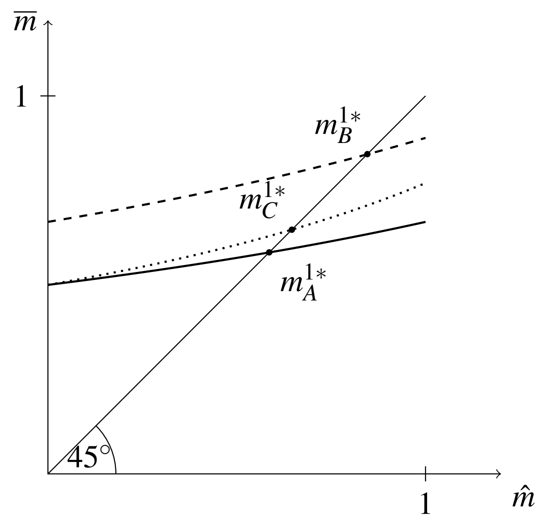

Figure 4: Consider the equilibrium level of mobility m_A^{1*}. A negative shift in the sense of FOSD of the distribution $F(\cdot)$ or an increase in s shifts up optimal mobility at any level of steady-state mobility, as illustrated by the dashed line. The equilibrium level of mobility increases to m_B^{1*}. An increase in p or an increase in d shifts up optimal mobility at any positive level of steady-state mobility, as illustrated by the dotted line. The equilibrium level of mobility increases to m_C^{1*}.
The solid line is drawn for F being a Uniform distribution on $[0,\overline{c}]$, $\overline{c}=1$, $V=2$, $\theta=\frac{1}{2}$, $s=\frac{3}{2}$, $p=d=\frac{1}{2}$. The dashed line is drawn for $\frac{V(s-1)}{\overline{c}}=\frac{8}{3}$, and the dotted line for $p\cdot d=\frac{7}{20}$, all else equal.

The first result shows that an admission code which balances intakes across schools ($p=0$), e.g. an admissions lottery, may hinder learning about school quality compared to an admission code which leads oversubscribed schools to prioritise more able applicants ($p>0$). The reason is that imbalanced intakes do not simply add noise to performance-based rankings, but help the current ranking to convey information about past performance. This paper is the first to formally analyse the impact of admission codes on the informativeness of rankings and it is able to contribute novel insights to the recent policy discussion about admission codes. For example, Burgess and Allen [2010] argue that "where schools are very similar in their intakes, the excellence of teaching and learning will be critical to where the school is placed in a local league table of academic performance", whereas if intakes are very imbalanced then differences in school qualities are less likely to affect the ranking, "because differences in pupil intakes will produce very large differences in raw outcomes" (p.10). The authors conclude that equal intakes have the advantage that differences in school qualities are more likely to be reflected in rankings. Their statements are consistent with my assumptions about how the distributions of rankings is determined for given intakes. However, in a dynamic setting in which intakes are endogenously determined, I come to the opposite conclusion, i.e. differences in school qualities are less likely to be reflected in rankings when intakes are equal.

In addition, the second result shows that learning about school quality from rankings is facilitated if it is less costly for students to choose a school based on its perceived quality. Recent school choice reforms can be modelled as a decrease in transport costs, because under these reforms students no longer had to move to the school's attendance area to be admitted, but could apply from outside this area and then commute. Therefore, my findings imply that school choice reforms should improve the informativeness of rankings and thereby help to target accountability pressures towards lower quality schools. A further prediction is that, in areas where students have a larger set of schools within a reasonable distance, these schools' relative performance should be a stronger indicator about their quality than in areas where students have less choice. The result fits with the observation that the rising selectivity at top US colleges coincides with a decrease in the cost of attending a university further from home (Hoxby [2009]). In addition, it fits with empirical evidence that for patients who require non-emergency care (lower transport cost) the correlation between hospital quality and market share is higher than for patients who require emergency treatment (Chandra et al. [2016]). Finally, transport costs in the model can represent other forms of perceived horizontal differentiation between schools, e.g. different schools may differ in their area of specialisation (arts versus sciences) and students differ in their preferences over specialisations. In areas where schools are more homogeneous, rankings should be more informative about schools' relative quality. This suggests that if students are offered a more diverse set of schools to choose from then they are more likely to find a school that matches their individual needs, but they are less likely to learn which school is of better quality.

Recent policy efforts have focused on providing more students with information about schools' performance. In my framework, the cost c could be interpreted as the cost to look up the most recent ranking, but unlike in the baseline model these costs would then be incurred independent of where the student ends up applying. Students would trade off researching schools and potentially applying to a better-performing school against remaining uninformed and applying to their local school. Therefore, a reduction in costs would again trigger more students to apply to the better-performing school, and my results suggest that such policies would contribute to improving the informativeness of rankings.[29]

Finally, results about increasing the performance advantage due to a more able intake or superior quality are interesting in the light of a recent trend in the US and UK towards using value-added measures of school performance. In the context of this model, I would expect a value-added

[29]That access to performance information can increase applications to well-performing schools has been shown by Hastings and Weinstein [2007].

measure to simultaneously raise the performance advantage s of the better school and lower the performance advantage d due to a more able intake. The rise in s improves informativeness, the fall in d reduces informativeness. This shows that reducing the influence of student ability on performance is not equivalent to reducing other forms of noise in rankings because past allocations convey information about earlier performance. Therefore, if value-added measures are less sensitive to intake ability, then to increase informativeness these measures need to raise sensitivity to underlying school quality sufficiently.

6 Welfare

This section analyses how a social planner would optimally assign students to schools if he were subject to the same informational constraints as students, i.e. he can condition this assignment only on each student's type (their transport costs and their ability) and on the identity of the most recent winning school W_{t-1}. This exercise will highlight how students' equilibrium strategy profile deviates from the socially optimal allocation. In addition, it highlights trade-offs that a forward-looking social planner may incur when choosing the optimal assignment.

I will start by defining *expected social welfare* in period t, given school X is the most recent winning school:

$$\begin{aligned} E\left(\Pi_t | W_{t-1}^X\right) =& VPr\left(\omega^X | W_{t-1}^X\right)\left[\eta h_t + 1 - h_t\right] \\ &+ V\left(1 - Pr\left(\omega^X | W_{t-1}^X\right)\right)\left[h_t + \eta\left(1 - h_t\right)\right] \\ &- 2\int_0^{F^{-1}\left(\left|h_t - \frac{1}{2}\right|\right)} c dF(c), \end{aligned} \tag{31}$$

where h_t denotes the share of high-ability students being allocated to the most recent winner, F^{-1} denotes the inverse of F and $\eta \in (0, \infty)$ measures the extent to which society values an increase in the share of high-ability students at the better school.[30] An analogous definition applies if school Y is the winning school. The social welfare function allows for the possibility that society derives a larger or lower benefit from assigning a high-ability student rather than a low-ability student to the better school, i.e. school quality and intake ability could be *complements* ($\eta > 1$) or *substitutes* ($\eta < 1$) in the social welfare function.[31]

For any given share $h_t \neq \frac{1}{2}$ of high-ability students assigned to W_{t-1}, the social planner has to assign some students to their non-local schools. The social welfare function only counts transport

[30] Note that if a share h_t of high-ability students are allocated to the most recent winner then this implies that the remaining capacity is equal to $1 - h_t$ and filled with low-ability students.

[31] If $\eta = 1$ then the social benefit is constant and independent of the ability level of students at each school.

costs incurred in excess of the (unavoidable) cost of assigning each student to their local school and already incorporates that the social planner minimises the total expenditure on transport costs by choosing the students who incur the lowest additional transport costs within each ability group.

Next, I will define the social planner's problem. I will assume the social planner commits to allocating share $h_t = h$ of high-ability students to the most recent winning school in each period t such as to maximise expected social welfare in steady state. In steady state, the signal distribution is stationary given $h_t = h$ and, hence, expected welfare is constant across periods, i.e. $\Pi_t = \Pi$. This implies that the social planner solves

$$\begin{aligned}\max_h E\left(\Pi|W^X\right) = {} & V Pr\left(\omega^X|W^X\right)[\eta h + 1 - h] \\ & + V\left(1 - Pr\left(\omega^X|W^X\right)\right)[h + \eta(1-h)] \\ & - 2\int_0^{F^{-1}\left(\left|h-\frac{1}{2}\right|\right)} c dF(c), \end{aligned} \tag{32}$$

such that

$$Pr\left(W^X|\omega^X\right) = \sum_{j=X,Y} g\left(W^X|h^j = h, \omega^X\right) Pr\left(W^j|\omega^X\right) \tag{33}$$

and similarly for $Pr\left(W^Y|\omega^Y\right)$. The social planner's assignment choice will influence social welfare at any given signal distribution, and importantly, it will also influence the stationary signal distribution. This captures that the current assignment of students to schools not only affects current welfare, but also affects future welfare indirectly because it affects how informative future rankings will be about school quality. As a benchmark, suppose the social planner is *myopic*, i.e. he chooses his optimal allocation taking the signal distribution as given.[32] The assignment in steady state is denoted by h_m. Contrast this with the scenario in which the social planner is *forward-looking*, i.e. he chooses his optimal allocation h_f also taking into account how this will affect the stationary signal distribution.

Proposition 3 [Welfare]

A myopic (forward-looking) social planner commits to allocating a share h_m (h_f) of high-ability students to the most recent winning school in each period such as to maximise expected welfare in steady state. If $\eta = 1$, i.e. if there are neither complements nor substitutes between school quality and intake ability, then $h_f = h_m = \frac{1}{2}$. If $\eta > 1$, i.e. if there are complements, then $h_f \geq h_m \geq \frac{1}{2}$. If $\eta < 1$, i.e. if there are substitutes, then $h_m \leq \frac{1}{2}$ and $h_f \geq h_m$.

[32] A myopic social planner disregards how the allocation today will affect the informativeness of future signals, but welfare in steady state incorporates these effects on informativeness in the expected per-period payoff.

First, consider $\eta = 1$, which corresponds to a utilitarian social planner. The optimal allocation is such that each student is assigned to their local school, irrespective of whether the social planner is myopic or forward-looking, i.e. $h_m = h_f = \frac{1}{2}$. This is because the benefit V is independent of the match between school quality and intake ability and transport costs are minimised if each student attends their local school.

Second, consider $\eta > 1$, i.e. the case of complements. A myopic social planner will allocate the majority of high-ability students to the school of higher expected quality, i.e. $h_m \geq \frac{1}{2}$. This is because the net benefit from increasing the share of high-ability students and reducing the share of low-ability students at the school of higher expected quality outweighs the additional spending on transport costs. In addition, if the social planner is forward-looking, he would like to allocate an even larger share of high-ability students to the school of higher expected quality than when he is myopic, i.e. $h_f \geq h_m \geq \frac{1}{2}$. This is because he also takes into account that increasing the share of high-ability students at the school of higher expected quality improves how likely it is that the most recent winner is the better school and therefore allows him to exploit complementarities even more.

Third, consider $\eta < 1$, i.e. the case of substitutes. If the social planner is myopic, he will allocate the majority of low-ability students to the school of higher expected quality, i.e. $h_m \leq \frac{1}{2}$. Again, a forward-looking social planner would allocate more high-ability students to the better performing school than a myopic one in order to improve how informative the ranking is about relative school quality. However, it is unclear whether or not a forward-looking planner will allocate the majority of low-ability students to the school of higher expected quality, i.e. whether $h_f \geq \frac{1}{2}$ or $h_f \leq \frac{1}{2}$. This is because when school quality and intake ability are substitutes then the social planner faces a tension between utilizing the acquired information about school quality to improve the assignment and choosing the assignment to improve the information about quality.

To highlight several externalities between students, compare students' equilibrium profile of strategies with the social planner's optimal allocation. First, consider the socially optimal benchmark if $\eta = 1$. In equilibrium, some students apply to their non-local school because each student compares only their individual gain from attending the school of higher expected quality with the additional transport costs incurred, but they do not take into account that they may take up a place that would have otherwise been available for a student living closer to the school. Second, consider the socially optimal benchmark if $\eta > 1$ ($\eta < 1$). In equilibrium, students may apply to the school of higher expected quality despite the fact that society would derive a greater benefit if a student of higher (lower) ability took their place instead. Conversely, a student may not apply to the school of higher expected quality even if it would be socially optimal, because they do not take into account that benefit derived by society may exceed the additional transport costs. Finally, students do not take into account that their allocation will influence the ranking available to students in future periods.

7 Extension: Quasi-Market Reforms

This extension will analyse how overall school quality and students' access to good schools is affected by quasi-market reforms. These reforms intend to mimic market forces by linking a school's funding to demand for its places, thereby putting pressures on unpopular schools to improve or shut down. Clearly, these reforms are more effective at improving school quality if worse schools are less popular, but this crucially depends on how well students can identify which schools are worse. My framework is well suited to study the dynamic interaction between student's inference about school quality from rankings, which affects demand for places, and the supply side response to this demand, which affects the qualities of schools.

In the baseline model, school qualities were assumed to be exogenously determined, i.e. the supply side was exogenous. Therefore, students' application strategies affected the allocation of students to schools, but not the overall quality of schools. To incorporate the supply side response triggered by quasi-market reforms, I assume that the undersubscribed school's quality is, with some probability, replaced by a new draw from an exogenously given quality distribution. This represents the fact that the undersubscribed school is under pressure to make changes, e.g. to experiment with new methodologies, or to replace their leadership, or in the worst case to close down. Such changes do not necessarily lead to improvements and, therefore, directing pressure at bad schools is valuable in order to observe long-term improvements in school quality.[33] This extended framework allows me to study a steady-state equilibrium in which the overall quality of schools is endogenous because it depends on students' application strategies, and students' application strategies are optimal given the distribution of school qualities.

The set-up of the baseline model is amended in the following way. A school's quality in period t, denoted by $Q_t \in \{G, B\}$, is either good ($Q_t = G$) or bad ($Q_t = B$). At the beginning of period $t = 0$, each school's quality is drawn independently at random such that $Pr(Q_0 = G) = \frac{1}{2}$. In each period $t \geq 0$, after students' payoffs have been realised, but before the ranking of schools is drawn, a school which was undersubscribed in period t has its quality replaced with probability $\gamma \in [0, 1]$ by a new quality draw, where $Pr(Q_t = G) = \frac{1}{2}$.[34] The other school's quality is not replaced and remains the same as in the previous period. Denote the underlying pair of school qualities at the end of period t by $Q^X Q^Y_t \in \{GG_t, GB_t, BG_t, BB_t\}$. As in the baseline model, if school X is the better school ($Q^X Q^Y_t = GB_t$) then the probability that school X wins with a share h^X_t of high-ability applicants is given by

$$g\left(W^X_t | h^X_t, GB_t\right) = \frac{s}{2} + \left(h^X_t - \frac{1}{2}\right) d \tag{34}$$

[33] It is unrealistic to assume that every intervention weakly improves school quality and, in addition, if this were the case school quality would be certain to improve over time irrespective of which schools come under pressure.

[34] The distribution of the new quality could be dependent on the quality level it replaces. The more likely it is that bad quality improves and the more likely it is that good quality deteriorates with an intervention, the more overall school quality improves when bad schools rather than good schools are selected for replacement.

and $g\left(W_t^Y|h_t^Y, BG_t\right) = g\left(W_t^X|h_t^X, GB_t\right)$. In addition, if schools are of equal quality, the probability that school X wins is given by

$$g\left(W_t^X|h_t^X, GG_t\right) = g\left(W_t^X|h_t^X, BB_t\right) = \frac{1}{2} + \left(h_t^X - \frac{1}{2}\right)d, \tag{35}$$

since neither school has an advantage due to superior quality, i.e. $s = 1$.[35] I will continue to assume that students only observe the most recent ranking of schools. School quality replacements are not observed.

A steady state is a time-invariant mobility, i.e. $m_t = \widehat{m}$ for all t, and a joint distribution of the pair of school qualities ($Q^X Q_t^Y$) and the signal of schools' relative quality (W_{t-1}) such that the joint distribution is stationary given $m_t = \widehat{m}$. A *steady-state equilibrium* is a steady state such that the time-invariant mobility $m_t = m^*$ is optimal given the stationary joint distribution. A student's optimal application strategy will depend on the difference between the posterior belief that the most recent winner is strictly better and the likelihood that the most recent winner is strictly worse. Therefore, informativeness in period t will be defined as

$$\begin{aligned} I_t \equiv & Pr\left(GB_{t-1}|W_{t-1}^X\right) - Pr\left(BG_{t-1}|W_{t-1}^X\right) \\ = & Pr\left(BG_{t-1}|W_{t-1}^Y\right) - Pr\left(GB_{t-1}|W_{t-1}^Y\right). \end{aligned} \tag{36}$$

To study the effect of quasi-market reforms, I will compare the case in which the supply side is independent of the demand side, i.e. $\gamma = 0$, with the case in which the supply side is affected by the demand side, i.e. $\gamma > 0$.[36]

Proposition 5 [Quasi-Market Reforms]
If $\gamma > 0$ rather than $\gamma = 0$, then

- *the level of mobility and informativeness remains unchanged, and*
- *the average fraction of good schools increases, and*
- *both the share of low-ability students and the share of high-ability students who attend a good school increases.*

[35] In steady state, it is irrelevant how the relative performance of schools of the same quality is determined, due to the symmetry of the set-up.

[36] Note that $\gamma = 0$ does not recover the baseline model, because there is a likelihood that both schools are good or both are bad. However, the qualitative insights of the baseline model apply if $\gamma = 0$ because optimal mobility still only varies with the difference between the posterior belief that the winning school is better and the posterior belief that the winning school is worse.

See Online Appendix for proofs. Mobility and informativeness remain unaffected as supply side responses are incorporated for two reasons. First, informativeness in steady-state equilibrium is independent of γ for $\gamma > 0$, because the relative frequency with which a good rather than a bad school is replaced is unaffected by γ, and hence the relative frequency with which a good school ranks higher than a bad school is also unaffected.[37] Second, the chances that schools are of different quality at any given point in time is unaffected by whether $\gamma = 0$ or $\gamma > 0$ because a new school is equally likely to be good or bad.

Replacing an undersubscribed school by a new school increases the overall quality of schools in equilibrium. To remain in steady state, the proportion of good schools closing down must equal to the proportion of good schools opening up. However, the average length of time that a good school operates increases and, therefore, the overall quality improves. Given that mobility is unaffected, the allocation of students conditional on the most recent ranking is unaffected. Consequently, both ability groups benefit from linking school replacement to students' application choices.

It is interesting to repeat some of the comparative static exercises of Theorem 1 when supply side responses are incorporated.

Proposition 4 [Quasi-Market Reforms - Comparative Statics]
If $\gamma > 0$, then

1. *an increase in the proportion p of non-local high-ability applicants admitted by an oversubscribed school, or*

2. *a negative shift in the sense of FOSD of the distribution $F(c)$ of transport costs*

- *increases the equilibrium level of mobility and the informativeness of rankings, and*

- *increases the average fraction of good schools, and*

- *increases the share of high-ability students but decreases the share of low-ability students who attend a good school.*

Any change in the environment which raises equilibrium informativeness and mobility at any given distribution of school qualities will have two effects. First, it will raise the share of high-ability students at the relatively better school (sorting effect), as in the case of exogenous school qualities. Second, it will raise the average fraction of good schools (quality effect). Both the quality and the sorting effect cause the share of high-ability students at good schools to increase.

[37] In particular, for any given time-invariant mobility level $\widehat{m}$, γ only affects how quickly the joint distribution of school qualities and rankings converges to its stationary distribution, but not the stationary distribution itself.

However, the share of low-ability students at good schools decreases because the sorting effect is of first order, while the quality effect is of second order.

This has implications for policy. If a policymaker only cares about the share of students accessing good schools, then it is desirable to make schools' admission policies more selective or to reduce the cost of applying to a non-local school. However, a policymaker may also care about who achieves access to a good school. Since these policies only raise the share of high-ability students at good schools, they may not be desirable if the policymaker has concerns for equity or places a higher value on improving low-ability student's productivity.

To highlight the additional dynamic effects arising with supply side effects, it is helpful to repeat the exercise in Section 6 and compare the social planner's optimal allocation when he is myopic compared to when he is forward-looking. Suppose the most recent losing school is replaced with probability γ and the social planner can determine the share h of high-ability students at the most recent winning school.

Proposition 6 [Quasi-Market Reforms - Welfare]
Suppose there are neither complements nor substitutes between school quality and intake ability, i.e. $\eta = 1$. *Under quasi-market reforms, i.e.* $\gamma > 0$, *a forward-looking social planner would allocate a larger share of high-ability students to the most recent winner than a myopic social planner:* $h^f \geq h^m = \frac{1}{2}$.

Given there are neither complements nor substitutes between school quality and intake ability, a myopic social planner will allocate each student to their local school, just as he would if school qualities were exogenous. However, a forward-looking social planner will allocate more high-ability students to the most recent winner because he thereby raises the average fraction of good schools in steady state.

In the baseline model with exogenous school qualities, it is socially inefficient that some students apply to their non-local school when it is the most recent winner. However, in the presence of quasi-market reforms, there is a further inefficiency which may alleviate the first: Students do not take into account that applying to the most recent winner entails a benefit for students in future periods because it increases the average fraction of good schools in future periods.

8 Extension: Peer Effects

This extension will incorporate that students may not only care about the quality of the school but also about the ability of their peers. Students' inference problem, which is the focus of this paper, would still persist even in the presence of such peer effects. The ranking of schools is influenced by students who attended these schools in the past, not by their contemporary peers. If I incorporate concerns for peers, the findings of the baseline model change in two important ways. First, the smallest equilibrium levels of mobility and informativeness increase relative to the baseline model which demonstrates that peer effects can reinforce the increase in informativeness and sorting described in the baseline model. Second, unlike in the baseline model, a negative equilibrium level of mobility may exist.[38]

The set-up of the baseline model is amended in the following way. Students derive a benefit $V > 0$ in the event that the school they attend wins, rather than in the event that this school is better.[39] Hence, the realised payoff of a student in period t of type (λ, α) conditional on enrolling at school $i = X, Y$ and conditional on the ranking realisation in period t is given by:

$$\pi_{t,\lambda,\alpha}(\varepsilon, W_t) = \begin{cases} V \cdot 1_{W_t = W_t^X}(W_t) - \lambda & \text{if } \varepsilon = \varepsilon^X \\ V \cdot 1_{W_t = W_t^Y}(W_t) - (1-\lambda) & \text{if } \varepsilon = \varepsilon^Y \end{cases} \tag{37}$$

where $1_{W_t = W_t^i}(W_t) = 1$ if $W_t = W_t^i$ and 0 otherwise for $i = X, Y$. This implies that their payoff increases in both the school's quality and the ability of the current intake. Parameter d measures how strongly the signal of school quality is biased in favour of the school with more high-ability students. In the context of this extension, parameter d can be interpreted as the strength of peer effects. The stronger peer effects, the stronger students' preference for the school with a higher number of high-ability students, all else equal.

Proposition 7 [Peer Effects] *1.) The smallest positive equilibrium levels of mobility and informativeness increase as i) peer effects are stronger (d larger), or as ii) an oversubscribed school selects a larger proportion p of non-local high-ability applicants. 2.) Some share of students may apply to the non-local losing school, i.e. a negative equilibrium level of mobility may exist, and it is more likely to exist as i) peer effects are stronger (d larger), or as ii) an oversubscribed school selects a larger proportion p of non-local high-ability applicants.*

For proof see Online Appendix. Given the presence of peer effects, if students expect a larger share of high-ability students to apply to the most recent winner, then this raises their benefit of attending this school for any given level of informativeness. In addition, it raises the benefit by more the more an oversubscribed school selects on ability or the stronger peer effects are. As a

[38] Rothstein [2006] examines the relative valuation of peers versus observed school effectiveness in Tiebout choice, and points out that a multiplicity of equilibria arises if students value peers, in some of which students do not sort according to school effectiveness. The negative level of equilibrium mobility which can arise in my model has a similar flavour, but my model also shows that peer effects can strengthen the informativeness of school performance rankings.

[39] This could also represent that students care about whether or not their school is perceived to be good, e.g. because it improves their employment prospects.

larger share of students applies to the winning school, informativeness rises by more, reinforcing the initial effect. Therefore, stricter selection on ability or stronger peer effects unambiguously raise the smallest equilibrium levels of mobility and informativeness.

The introduction of peer effects can give rise to steady-state equilibria associated with a negative level of mobility, i.e. students incur costs to travel to the *most recent loser*. Students may still believe that the most recent winning school is more likely to be the better school. Yet if they expect the share of high-ability students who will enrol at the most recent loser to be sufficiently high, the expected benefit derived from better peers at a worse school may outweigh the expected benefit derived from a better school with worse peers. These equilibria are more likely to arise if peer effects are stronger (d higher) or if oversubscribed schools admit an even larger share of high-ability students (p higher).

9 Robustness Checks

This section shows to what extent my findings are robust to changes in some of the assumptions.

9.1 Longer Window of Rankings

I have assumed that students only observe the most recent ranking. The insights of this paper still apply if students observed a longer window of rankings. If this were the case, students would have access to some of the information based on which students in previous periods applied to schools. However, crucially, they still would not have access to *all* the information available to students in all previous periods. Therefore, it is still true that rankings would convey some additional information about the relative quality of schools if students in each period applied to the school which they expect to be of higher quality. This is the feature on which the proofs for equilibrium existence and comparative statics results are based.

In the Online Appendix, I analytically solve for steady-state informativeness, when, in each period t, students observe the two most recent rankings. Students' strategy profile is characterised by a pair of mobility levels, one conditional on the same school winning in both of the two most recent periods and one conditional on a different school winning in each of these periods. I show that the stationary distribution of rankings is such that it is uninformative to observe a different school win in each ranking, independent of the pair of mobility levels. In addition, I show that observing the same school win is informative, and weakly more so the higher the level of mobility conditional on observing the same school win and independent of the level of mobility conditional on observing different schools win.

9.2 Self-fulfilling Rankings and Herding

The property that the better school has a strict performance advantage relative to the worse school (see (2)) implies that the worse school never wins for certain, independent of the share of high-ability students admitted. Suppose I were to allow for the possibility that there exists some critical level $\overline{h}$ for the share of high-ability students such that if $h^i > \overline{h}$ then school $i = X,Y$ is guaranteed to win.[40] Then for some sufficiently high level of mobility, the stationary distribution of rankings in steady state would no longer be unique. It would either assign probability 1 to the worse school winning or probability 1 to the better school winning. This is because for a sufficiently high level of mobility, a self-perpetuating cycle could arise in which the same school continues to win, because once it has attracted sufficiently many high-ability students to win it will then be able to attract these high-ability students again and again. In the terminology of the social learning literature, a herd will arise. In this situation, it is no longer true that the realised level of informativeness necessarily increases in the proportion p of non-local high-ability applicants admitted by an oversubscribed school. However, the higher the proportion p of non-local high-ability applicants selected, the more likely it is that the better school rather than the worse school will be the first to reach the critical level $\overline{h}$. Therefore, from an ex-ante perspective, informativeness is still at least as high if oversubscribed schools select on ability as if intakes are balanced across schools.

9.3 Admission not based on Ability

Even in situations in which schools cannot select among applicants based on their ability, the insights developed in this paper can be useful. In the baseline model, an increase in mobility strictly increases steady-state informativeness only if the most recent winner (and hence oversubscribed school) admits some non-local high-ability applicants, i.e. $p > 0$. However, an alternative set of assumptions could also ensure that the most recent winner admits more high-ability students and that this share of high-ability students increases in informativeness.

As an example, I will show that this is the case if I assume students with high-ability students derive a larger benefit from attending a better school, i.e. $V^H > V^L$, and if the distribution for transport costs is Uniform. The assumption that $V^H > V^L$ is supported by empirical evidence that students of higher socio-economic status are more likely to seek out better-performing schools, e.g. see Hastings et al. [2005], Allen et al. [2014]. Students' application strategies would not only depend on their location, but also on their ability. High-ability students would be willing to incur higher transport costs than low-ability students to attend the most recent winning school at any

[40] The baseline model has abstracted away from this to keep the analysis tractable. The baseline model can be thought of as capturing situations in which there is sufficiently high differentiation between schools such that intakes will not become too unbalanced across schools, or situations in which students' abilities are relatively homogeneous so that the performance advantage from taking on more able students is not too large.

given level of informativeness. Hence, the applicant pool of the most recent winning school will contain more high-ability than low-ability students. Therefore, even if this school selects among applicants at random, it will admit more high-ability students than the other school. In addition, for any increase in informativeness, the most recent winner will receive relatively more applicants from high-ability students.[41] More detail can be found in the Online Appendix.

In addition, even if a school had to admit those students who live nearest to the school, some of the effects captured in this model may still be present. Students would consider the quality of the school when choosing where to live (Tiebout choice). If students with higher-ability are more likely to move into the proximity of the better-performing school, either because they value school quality more or because they are more likely to be able to afford housing there, then again a better-performing school may admit more high-ability students.

10 Conclusion

This paper studies students' inferences about school quality from rankings, when students in each period observe which school performed better in the previous period, but they do not observe past allocations or applications. I develop a dynamic framework in which the pool of applicants at a school and its relative performance are endogenously linked and analyse comparative statics in a steady-state equilibrium. I find that a performance-based ranking is more informative about school quality if the admission code leads oversubscribed school to prioritise more able applicants. I also find that such a ranking is more informative if it is less costly for students to attend a non-local school. My paper is the first in the observational learning literature to derive comparative statics when agents observe a limited window of realisations of a public signal, whose distribution depends on past agents' choices. Furthermore, my findings contribute to recent discussions on school choice and on the design of school admission codes. I view the framework developed as a building block for future research on analysing the link between information and match outcomes. In addition, the framework is also suitable to explore what could cause persistent differences between schools to arise, and a starting point to explore how schools build and maintain reputations over time when students have limited access to past performance information.

[41] I assume $V^H < 1$ such that not all high-ability students apply to the most recent winner.

A Appendix

A.1 Proposition 2 [Convergence]

The sequence of levels of informativeness in non-steady-state equilibrium satisfies (28) because students in period 0 do not observe a signal and, hence, their posterior beliefs equal their (uninformative) prior beliefs. In addition, the equilibrium sequence satisfies (29), since by (24) it follows that

$$\begin{aligned} I_t \equiv& Pr\left(\omega^X|W^X_{t-1}\right) - Pr\left(\omega^Y|W^X_{t-1}\right) = Pr\left(W^X_{t-1}|\omega^X\right) - Pr\left(W^X_{t-1}|\omega^Y\right) \\ =& Pr\left(W^X_{t-1}|\omega^X\right) - Pr\left(W^Y_{t-1}|\omega^X\right) = 2Pr\left(W^X_{t-1}|\omega^X\right) - 1, \end{aligned} \tag{38}$$

and since the stationary distribution must satisfy

$$\begin{aligned} Pr\left(W^X_{t-1}|\omega^X\right) =& g\left(W^X_{t-1}|h^X_{t-1} = \frac{1+m_{t-1}}{2}, \omega^X\right) Pr\left(W^X_{t-2}|\omega^X\right) + \\ & + g\left(W^X_{t-1}|h^X_{t-1} = \frac{1-m_{t-1}}{2}, \omega^X\right) Pr\left(W^Y_{t-2}|\omega^X\right) \\ =& \left[\frac{s}{2} + \left(\frac{p \cdot m_{t-1}}{2}\right) d\right] Pr\left(W^X_{t-2}|\omega^X\right) + \\ & \left[\frac{s}{2} - \left(\frac{p \cdot m_{t-1}}{2}\right) d\right] Pr\left(W^Y_{t-1}|\omega^X\right) \\ =& \frac{s}{2} + \frac{p \cdot d \cdot m_{t-1}}{2} I_{t-1} \end{aligned} \tag{39}$$

and since (16) holds.

Next, I will show that the sequence $\{I_t\}_{t\geq 0}$ converges to its smallest fixed point I^{1*}. Any increasing sequences converges to its least upper bound. The sequence $\{I_t\}_{t\geq 0}$ is increasing by the following induction argument. Define

$$Z(I_{t-1}) = s - 1 + p \cdot d \cdot F(V \cdot I_{t-1}) \cdot I_{t-1} \in [0,1] \tag{40}$$

$Z(I_t)$ increases in I_t given $F(\cdot)$ is positive and increasing, $V > 0$, $p \geq 0$, $d \geq 0$ and $I_t \geq 0$. Since $s > 1$,

$$I_1 = Z(I_0) \geq I_0 = 0,$$

and due to $Z(\cdot)$ being increasing for all I_t,

$$I_t = Z(I_{t-1}) \geq Z(I_{t-2}) = I_{t-1}.$$

Since $\{I_t\}_{t\geq 0}$ is increasing, it converges to its least upper bound. Next, I will show that this least upper bound is given by the smallest fixed point I^{1*}. Suppose I^{1*} was not the least upper bound. Then there would be some $\widetilde{I} \in [0,1]$ such that $\widetilde{I} < I^{1*}$ and $Z\left(\widetilde{I}\right) > I^{1*}$. Since I^{1*} is a fixed point, this implies $Z\left(\widetilde{I}\right) > Z\left(I^{1*}\right)$. But this contradicts the fact that $Z(\cdot)$ is increasing.

Given (20), $V > 0$ and $F(\cdot)$ is increasing, m_t increases in I_t. Hence, as $\{I_t\}_{t\geq 0}$ converges so does $\{m_t\}_{t\geq 0}$. Further, the smallest equilibrium level of informativeness I^{1*} corresponds to the smallest equilibrium level of mobility m^{1*}.

A.2 Theorem 1 [Comparative Statics]

Define

$$\Gamma(I,F,p,d,s,V) = \frac{s-1}{1-p\cdot d\cdot F(V\cdot I)} \in [0,1].$$

Denote the parameter subject to exogenous change by $z \in Z$, where Z is a subset of $\mathbb{R}$. To simplify notation, I will suppress all arguments other than z and I and write $\Gamma(z,I) : Z \times [0,1] \to [0,1]$. For any $z \in Z$, $\Gamma(z,I)$ is continuous and increasing in I since $s > 1$, $d, p \geq 0$, $F(\cdot)$ is positive and increasing. I will show that $\Gamma(z,I)$ is increasing in z for all $I \in [0,1]$. After Corollary 1, (p. 446), in Milgrom and Roberts [1994] (henceforth MR), this implies that the smallest fixed point of $\Gamma(z,I)$, denoted by $I^{1*}(z)$, is increasing in z. I will prove statements 1.-4. in Theorem 1 by choosing z to be the variable of interest:

1. $\Gamma(p,I)$ is increasing in p at every I since $s > 1$, $d \geq 0$ and $F(\cdot) \geq 0$.

2. For any I, and any $F(\cdot)$ and $\widetilde{F}(\cdot)$, such that $F(\cdot)$ first-order stochastically dominates $\widetilde{F}(\cdot)$, it holds that

$$F(V\cdot I) \leq \widetilde{F}(V\cdot I). \tag{41}$$

Hence,

$$\frac{s-1}{1-p\cdot d\cdot F(V\cdot I)} \geq \frac{s-1}{1-p\cdot d\cdot \widetilde{F}(V\cdot I)}.$$

3. $\Gamma(d,I)$ is non-decreasing in d at every I since $s > 1$, $p \geq 0$ and $F(\cdot) \geq 0$.

4. Since $\Gamma(d,I)$ is non-decreasing in s at every I since $p, d, F(\cdot) < 1$.

For 1.-4., by MR, I^{1*} is non-decreasing in z. For 1., 3. and 4., since $m^{1*} = F\left(V\cdot I^{1*}\right)$, $F(\cdot)$ is increasing and $V > 0$, m^{1*} is increasing in z. For 2., since $m^{1*} = F\left(V\cdot I^{1*}\right)$, $F(\cdot)$ is increasing, $V > 0$ and (41), m^{1*} is increasing with a negative shift in the sense of FOSD of $F(\cdot)$.

A.3 Proposition 3 [Welfare]

A social planner solves (32) subject to (33).

If $\eta = 1$, then this simplifies to

$$\max_{h} E\left(\Pi|W\right) = V - 2\int_{0}^{F^{-1}\left(\left|h-\frac{1}{2}\right|\right)} cdF\left(c\right),$$

which is independent of posterior beliefs $Pr\left(\omega|W\right)$ and therefore independent of the stationary ranking distribution $Pr\left(W|\omega\right)$. Whether the social planner is myopic or forward-looking, welfare is maximised at $h_m = h_f = \frac{1}{2}$. For any $h \in [0,1]$, there exists a unique distribution $Pr\left(W|\omega\right)$ such that (33) is satisfied because the sequence of ranking realisations follows a irreducible time-homogeneous Markov process.

Suppose $\eta \neq 1$. The social welfare function is not differentiable at $h = \frac{1}{2}$. Suppose the social planner solves

$$\max_{h \in \mathscr{H}} E\left(\Pi|W\right) \text{ subject to (33)},$$

first for $\mathscr{H} = \left[0, \frac{1}{2}\right]$ and then for $\mathscr{H} = \left[\frac{1}{2}, 1\right]$ and finally chooses the global maximum.

If the *social planner is myopic*, i.e. if he treats the stationary distribution as given, then the first derivative of expected social welfare if $\mathscr{H} = \left[0, \frac{1}{2}\right]$ or if $\mathscr{H} = \left[\frac{1}{2}, 1\right]$ is given by

$$V\left[2Pr\left(\omega^X|W^X\right) - 1\right](\eta - 1) - \frac{\partial}{\partial h}\left[2\int_{0}^{F^{-1}\left(\left|h-\frac{1}{2}\right|\right)} cdF\left(c\right)\right]. \tag{42}$$

By the proof of Lemma 2 for $h = \frac{1+p\cdot\widehat{m}}{2}$, it holds that

$$2Pr\left(\omega^X|W^X\right) - 1 = 2Pr\left(W^X|\omega^X\right) - 1 = \frac{s-1}{1-d\cdot(2h-1)} \geq 0. \tag{43}$$

Suppose $\eta > 1$. Then if $\mathscr{H} = \left[0, \frac{1}{2}\right]$ the optimum is $h_m = \frac{1}{2}$ since (42) is positive for any $h_m \in \left[0, \frac{1}{2}\right]$.[42] Hence, the optimal $h_m \in [0,1]$ must satisfy $h_m \geq \frac{1}{2}$.

Suppose $\eta < 1$. Then if $\mathscr{H} = \left[\frac{1}{2}, 1\right]$ the optimum is $h_m = \frac{1}{2}$ since (42) is negative for any $h_m \in \left[\frac{1}{2}, 1\right]$.[43] Hence, the optimal $h_m \in [0,1]$ must satisfy $h_m \leq \frac{1}{2}$.

If the *social planner is forward-looking*, i.e. if he takes into account that his choice will influence the stationary distribution, then the first derivative of expected social welfare if $\mathscr{H} = \left[0, \frac{1}{2}\right]$

[42] As h rises for $\mathscr{H} = \left[0, \frac{1}{2}\right]$ more students are assigned to their local school, so expenditure on transport costs decreases.

[43] As h rises for $\mathscr{H} = \left[\frac{1}{2}, 1\right]$ fewer students are assigned to their local school, so expenditure on transport costs increases.

or if $\mathscr{H}=\left[\frac{1}{2},1\right]$ is given by

$$V\left[2Pr\left(\omega^X|W^X\right)-1\right](\eta-1)+\frac{\partial}{\partial h}Pr\left(\omega^X|W^X\right)\left[V\left((\eta-1)(2h-1)\right)\right] \\ -\frac{\partial}{\partial h}\left[2\int_0^{F^{-1}\left(\left|h-\frac{1}{2}\right|\right)}cdF(c)\right]. \tag{44}$$

Since (43) is increasing in h, it holds that

$$\frac{\partial}{\partial h}Pr\left(\omega^X|W^X\right)=\frac{\partial}{\partial h}\frac{s-d\cdot(2h-1)}{2\left(1-d\cdot(2h-1)\right)}\geq 0. \tag{45}$$

Suppose $\eta>1$. Then if $\mathscr{H}=\left[0,\frac{1}{2}\right]$, then $h_f=\frac{1}{2}$ since (44) is positive at $h_m=\frac{1}{2}$ given (43) and (45). If $\mathscr{H}=\left[\frac{1}{2},1\right]$, then $h_f\geq h_m\geq\frac{1}{2}$ since either i) $h_m=\frac{1}{2}$ and (44) is positive at $h_m=\frac{1}{2}$, or ii) $h_m>\frac{1}{2}$ and (44) is positive at h_m since (42) must be zero at h_m and (45) holds. Hence, $h_f\geq h_m$ for $\mathscr{H}\in[0,1]$.

Suppose $\eta<1$. Then if $\mathscr{H}=\left[\frac{1}{2},1\right]$, then $h_f\geq\frac{1}{2}$ since (44) is positive at $h_m=\frac{1}{2}$ given (43) and (45). If $\mathscr{H}=\left[0,\frac{1}{2}\right]$, then $h_f\geq h_m$ since either i) $h_m=\frac{1}{2}$ and (44) is positive at $h_m=\frac{1}{2}$ or ii) $h_m<\frac{1}{2}$ and (44) is positive at h_m since (42) must be zero at h_m and (45) holds. Hence, $h_f\geq h_m$ for $\mathscr{H}\in[0,1]$.

B Online Appendix

B.1 Extensions: Quasi-Market Reform

Lemma 3: *For $\gamma\geq 0$, a steady-state equilibrium level of mobility m^* is characterised by*

$$m^*=F\left(V\cdot\frac{s-1}{2\left(1-p\cdot d\cdot m^*\right)}\right) \tag{46}$$

and the corresponding steady-state equilibrium level of informativeness $I(m^*)$ is given by

$$I(m^*)=\frac{s-1}{2\left(1-p\cdot d\cdot m^*\right)}. \tag{47}$$

Such a steady-state equilibrium level of mobility (and informativeness) always exists.

I will first derive the stationary joint distribution of the pair of school qualities and of the signal W_{t-1} given $m=\widehat{m}$. Since I focus on symmetric equilibria, it is sufficient to define a realisation by the quality of the winning school and the quality of the losing school, but not their identities. I will denote a realisation by an unordered pair of school qualities and whenever schools differ in quality

I indicate the quality of the winning school by an upper bar, i.e. $\overline{G}B_t$ denotes the realisation that one school is good and one is bad and the good school is the most recent winner, and similarly $\overline{B}G_t$ denotes the realisation that one school is good and one is bad and the bad school is the most recent winner. Therefore, I denote the stationary joint distribution by the vector ρ where

$$\rho \equiv \left(Pr\left(GG\right), Pr\left(\overline{G}B\right), Pr\left(\overline{B}G\right), Pr\left(BB\right)\right).$$

The Markov process is defined by the following matrix T of transition probabilities between possible realisations in two consecutive periods:

$$T = \begin{pmatrix} Pr\left(GG_t|GG_{t-1}\right) & Pr\left(\overline{G}B_t|GG_{t-1}\right) & Pr\left(\overline{B}G_t|GG_{t-1}\right) & Pr\left(BB_t|GG_{t-1}\right) \\ Pr\left(GG_t|\overline{G}B_{t-1}\right) & Pr\left(\overline{G}B_t|\overline{G}B_{t-1}\right) & Pr\left(\overline{B}G_t|\overline{G}B_{t-1}\right) & Pr\left(BB_t|\overline{G}B_{t-1}\right) \\ Pr\left(GG_t|\overline{B}G_{t-1}\right) & Pr\left(\overline{G}B_t|\overline{B}G_{t-1}\right) & Pr\left(\overline{B}G_t|\overline{B}G_{t-1}\right) & Pr\left(BB_t|\overline{B}G_{t-1}\right) \\ Pr\left(GG_t|BB_{t-1}\right) & Pr\left(\overline{G}B_t|BB_{t-1}\right) & Pr\left(\overline{B}G_t|BB_{t-1}\right) & Pr\left(BB_t|BB_{t-1}\right) \end{pmatrix}.$$

Each transition probability is comprised of two parts: i) the undersubscribed losing school in period $t-1$ is replaced with probability γ by a new school which is of a different quality with probability $\frac{1}{2}$, and ii) the oversubscribed winning school in $t-1$ admits a share $\widehat{m}p$ of high-ability non-local students. The pair of school qualities and the allocation of high-ability students then determine the winner in period t. The transition matrix T has the following entries:

$$\begin{aligned}
Pr\left(\overline{G}B_t|GG_{t-1}\right) &= \frac{\gamma}{2}\left[\frac{1}{2}\left(s+\widehat{m}pd\right)\right] \\
Pr\left(\overline{B}G_t|GG_{t-1}\right) &= \frac{\gamma}{2}\left[1-\frac{1}{2}\left(s+\widehat{m}pd\right)\right] \\
Pr\left(BB_t|\overline{G}B_{t-1}\right) &= \frac{\gamma}{2}\left[\frac{1}{2}\left(s-\widehat{m}pd\right)\right] \\
Pr\left(BB_t|\overline{B}G_{t-1}\right) &= \frac{\gamma}{2}\left[1-\frac{1}{2}\left(s-\widehat{m}pd\right)\right] \\
Pr\left(\overline{G}B_t|\overline{G}B_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]\frac{1}{2}\left(s+\widehat{m}pd\right) \\
Pr\left(\overline{B}G_t|\overline{G}B_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]\left[1-\frac{1}{2}\left(s+\widehat{m}pd\right)\right] \\
Pr\left(\overline{G}B_t|\overline{B}G_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]\frac{1}{2}\left(s-\widehat{m}pd\right) \\
Pr\left(\overline{B}G_t|\overline{B}G_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]\left[1-\frac{1}{2}\left(s-\widehat{m}pd\right)\right] \\
Pr\left(GG_t|GG_{t-1}\right) = Pr\left(BB_t|BB_{t-1}\right) &= 1-\frac{\gamma}{2} \\
Pr\left(GG_t|\overline{G}B_{t-1}\right) = Pr\left(BB_t|\overline{B}G_{t-1}\right) &= \frac{\gamma}{2}
\end{aligned}$$

and the remaining entries are equal to 0.

If the Markov process characterised by T is both irreducible and aperiodic it has a unique stationary distribution which is defined by the row vector ρ that satisfies both

$$\rho \quad = \quad \rho T \tag{48}$$

and

$$Pr(GG) + Pr\left(\overline{G}B\right) + Pr\left(\overline{B}G\right) + Pr(BB) = 1. \tag{49}$$

If $\gamma > 0$ then the stationary distribution is defined by

$$Pr(GG) = Pr\left(\overline{G}B\right) = \frac{1}{4}\frac{(s - p \cdot d \cdot \widehat{m})}{1 - p \cdot d \cdot \widehat{m}} \tag{50}$$

$$Pr\left(\overline{B}G\right) = Pr(BB) = \frac{1}{4}\frac{(2 - s - p \cdot d \cdot \widehat{m})}{1 - p \cdot d \cdot \widehat{m}}. \tag{51}$$

If $\gamma = 0$, then the stationary distribution is defined by

$$Pr(GG) = Pr(BB) = \frac{1}{4} \tag{52}$$

$$Pr\left(\overline{G}B\right) = \frac{1}{4}\frac{(s - p \cdot d \cdot \widehat{m})}{1 - p \cdot d \cdot \widehat{m}} \tag{53}$$

$$Pr\left(\overline{B}G\right) = \frac{1}{4}\frac{(2 - s - p \cdot d \cdot \widehat{m})}{1 - p \cdot d \cdot \widehat{m}}. \tag{54}$$

Students' optimal application strategy is again described by (17) and (18), where informativeness is defined as

$$\begin{aligned} I_t(\widehat{m}) \equiv & Pr\left(GB_{t-1}|W_{t-1}^X\right) - Pr\left(BG_{t-1}|W_{t-1}^X\right) \\ = & Pr\left(BG_{t-1}|W_{t-1}^Y\right) - Pr\left(GB_{t-1}|W_{t-1}^Y\right), \end{aligned}$$

since by enrolling at school X rather than at school Y a student derives a benefit V if school X is better, loses V if school X is worse and neither gains nor loses if schools are of the same quality.

For $\gamma \geq 0$, informativeness in steady state is given by (47). Since optimal mobility in terms of informativeness is given by (16), the steady-state equilibrium mobility and informativeness solve (25). Existence follows by Tarski's fixed point theorem for the same reasons given in the proof of Proposition 1.

B.1.1 Proposition 4

Part 1: The equilibrium levels of mobility and informativeness of rankings in Lemma 3 are identical to the equilibrium levels in the baseline model up to a scaling factor. Hence, by the proof of Theorem 1, mobility and informativeness weakly increase (1.) in p and (2.) with a negative shift in the sense of FOSD of F.

Part 2: The average fraction of good schools in equilibrium, denoted by Θ, is given by:

$$\Theta = \begin{cases} Pr(GG) + \frac{1}{2}Pr\left(\overline{G}B\right) + \frac{1}{2}Pr\left(\overline{B}G\right) = \frac{s-1+2\left(1-p\cdot d\cdot m^{1*}\right)}{4\left(1-p\cdot d\cdot m^{1*}\right)} & \text{if } \gamma > 0 \\ \frac{1}{2} & \text{if } \gamma = 0 \end{cases} \tag{55}$$

1. $\frac{d\Theta}{dp} \geq 0$ since

$$\frac{d\Theta}{dp} = \frac{\partial\Theta}{\partial p} + \frac{\partial\Theta}{\partial m^{1*}}\frac{\partial m^{1*}}{\partial p}, \tag{56}$$

where $\frac{\partial m^{1*}}{\partial p} \geq 0$ given Part 1,

$$\frac{\partial\Theta}{\partial m^{1*}} = \frac{p\cdot d\cdot(s-1)}{2\left(1-p\cdot d\cdot m^{1*}\right)^2} \geq 0 \tag{57}$$

and

$$\frac{\partial\Theta}{\partial p} = \frac{m^{1*}\cdot d\cdot(s-1)}{2\left(1-p\cdot d\cdot m^{1*}\right)^2} \geq 0.$$

2. Θ increases with a negative shift in the sense of FOSD of $F(\cdot)$ since $F(\cdot)$ affects Θ only through its effect on m^{1*}, Θ weakly increases in $m^{1*}$by (57) and m^{1*} weakly increases with a negative shift in the sense of FOSD of $F(\cdot)$.

Part 3: Denote the share of high-ability students at a good school in equilibrium by h_G, and the share of low-ability students at a good school in equilibrium by l_G. Since the share of high-ability students at the most recent winner is $\frac{1}{2}\left(1+m^{1*}p\right)$,

$$h_G = Pr(GG) + \frac{1}{2}\left(1+m^{1*}p\right)Pr\left(\overline{G}B\right) + \frac{1}{2}\left(1-m^{1*}p\right)Pr\left(\overline{B}G\right) \tag{58}$$

$$l_G = Pr(GG) + \frac{1}{2}\left(1-m^{1*}p\right)Pr\left(\overline{G}B\right) + \frac{1}{2}\left(1+m^{1*}p\right)Pr\left(\overline{B}G\right). \tag{59}$$

Using the stationary probabilities derived in the proof of Lemma 3 for $\gamma > 0$:

$$
\begin{aligned}
h_G &= \frac{s+1-m^{1*}p\left(2d-(s-1)\right)}{4\left(1-dm^{1*}p\right)} \\
l_G &= \frac{s+1-m^{1*}p\left(2d+(s-1)\right)}{4\left(1-dm^{1*}p\right)}.
\end{aligned}
$$

1. $\frac{dh_G}{dp} \geq 0$ since

$$\frac{dh_G}{dp} = \frac{\partial h_G}{\partial p} + \frac{\partial h_G}{\partial m^{1*}} \frac{\partial m^{1*}}{\partial p}, \tag{60}$$

where $\frac{\partial m^{1*}}{\partial p} \geq 0$ given Part 1,

$$\frac{\partial h_G}{\partial p} = \frac{m^{1*}\left(s-1\right)\left(1+d\right)}{4\left[1-dpm^{1*}\right]^2} \geq 0$$

and

$$\frac{\partial h_G}{\partial m^{1*}} = \frac{p\left(s-1\right)\left(1+d\right)}{4\left[1-dpm^{1*}\right]^2} \geq 0. \tag{61}$$

In addition, $\frac{dl_G}{dp} \geq 0$ since

$$\frac{dl_G}{dp} = \frac{\partial l_G}{\partial p} + \frac{\partial l_G}{\partial m^{1*}} \frac{\partial m^{1*}}{\partial p}, \tag{62}$$

where $\frac{\partial m^{1*}}{\partial p} \geq 0$ given Part 1,

$$\frac{\partial l_G}{\partial p} = -\frac{m^{1*}\left(s-1\right)\left(1-d\right)}{4\left[1-dpm^{1*}\right]^2} \leq 0 \tag{63}$$

and

$$\frac{\partial l_G}{\partial m^{1*}} = -\frac{p\left(s-1\right)\left(1-d\right)}{4\left[1-dpm^{1*}\right]^2} \leq 0. \tag{64}$$

2. h_G increases and l_G decreases with a negative shift in the sense of FOSD of $F\left(\cdot\right)$ because such a shift affects h_G and l_G only through its effect on m^{1*}, m^{1*} weakly increases with such a shift by Part 1, h_G weakly increases in m^{1*} by (61) while l_G weakly decreases in m^{1*} by (64).

B.1.2 Proposition 5

Part 1: Informativeness and mobility are unaffected by γ as shown in Lemma 3.

Part 2: The difference in the average fraction of good schools when $\gamma > 0$ versus $\gamma = 0$ is given by:

$$\begin{aligned}\Theta(\gamma>0)-\Theta(\gamma=0) &= \frac{s-1+2\left(1-p\cdot d\cdot m^{1*}\right)}{4\left(1-p\cdot d\cdot m^{1*}\right)}-\frac{1}{2} \\ &= \frac{s-1}{4\left(1-p\cdot d\cdot m^{1*}\right)}\geq 0\end{aligned}$$

Part 3: Since mobility is unchanged by Part 1, the share of high-ability students at the most recent winner is unchanged. In steady state, this implies that the share of high-ability students at the relatively better school is unchanged. If $\gamma>0$, the average fraction of good schools increases by Part 2 and, hence, the fraction of students of each ability type at a good school increases. Using (58) and (59):

$$h_G(\gamma>0)-h_G(\gamma=0)=l_G(\gamma>0)-l_G(\gamma=0)=\frac{s-1}{4\left(1-p\cdot d\cdot m^{1*}\right)}\geq 0.$$

B.1.3 Proposition 6

Given that there are no complements or substitutes, the social benefit is equal to the benefit V weighted by the share of students enrolled at a good school in steady state:

$$\begin{aligned}E(\Pi) &= V\left[2Pr(GG)+Pr\left(\overline{G}B\right)+Pr\left(\overline{B}G\right)\right]-2\int_0^{F^{-1}\left(\left|h-\frac{1}{2}\right|\right)} c\,dF(c) \\ &= 2V\Theta-2\int_0^{F^{-1}\left(\left|h-\frac{1}{2}\right|\right)} c\,dF(c)\end{aligned}$$

where $Pr(GG), Pr\left(\overline{G}B\right)$ and $Pr\left(\overline{B}G\right)$ are given by the stationary joint distribution ρ of school qualities and rankings derived in Lemma 3 and the fraction Θ of good schools is given by (55).

A myopic social planner takes ρ as given, and therefore Θ as given, when he chooses h such as to maximise $E(\Pi)$. He chooses $h_m=\frac{1}{2}$ because he treats the benefit as independent of h and cost is minimised at $h=\frac{1}{2}$. A forward-looking social planner takes into account that his choice of h affects ρ and hence Θ. Since the social benefit increases in the fraction Θ of good schools, and since the fraction Θ of good schools weakly increases in the proportion of non-local high-ability students admitted to the most recent winning school by (57), the social benefit weakly increases in h. Therefore, a forward-looking social planner will never choose a lower h than a myopic social planner: $h_f\geq h_m$.

B.2 Extensions: Peer Effects

Lemma 4 *A steady-state equilibrium level of mobility is characterised by*

$$m^* = F\left[V \cdot \left(\frac{(s-1)^2}{1-p\cdot d\cdot m^*} + p\cdot d\cdot m^*\right)\right] \tag{65}$$

and the corresponding steady-state equilibrium level of informativeness is given by

$$I(m^*) = \frac{s-1}{1-p\cdot d\cdot m^*}. \tag{66}$$

Such equilibrium steady-state levels of mobility (and informativeness) always exist.

A student still optimally applies to the school at which their expected payoff conditional on enrolment is higher. Suppose $W_{t-1} = W_{t-1}^X$. Then by enrolling at school X rather than at school Y a student derives a benefit V if school X wins, but loses V if school X loses in period t. As in the baseline model, they incur the additional transport costs $c(\lambda)$ if school X is their non-local school, but save these costs if school X is local:

$$\begin{aligned} & E\left(\pi_{t,\lambda,\alpha}|e_t^X, W_{t-1}^X\right) - E\left(\pi_{t,\lambda,\alpha}|e_t^Y, W_{t-1}^X\right) \\ = & \begin{cases} V\cdot\left[Pr\left(W_t^X|W_{t-1}^X\right) - Pr\left(W_t^Y|W_{t-1}^X\right)\right] + c(\lambda) & \text{if } \lambda \leq \frac{1}{2} \\ V\cdot\left[Pr\left(W_t^X|W_{t-1}^X\right) - Pr\left(W_t^Y|W_{t-1}^X\right)\right] - c(\lambda) & \text{if } \lambda > \frac{1}{2}, \end{cases} \end{aligned}$$

where

$$
\begin{aligned}
&Pr\left(W_t^X|W_{t-1}^X\right)-Pr\left(W_t^Y|W_{t-1}^X\right)\\
=&\sum_{i=X,Y}Pr\left(W_t^X|W_{t-1}^X,\omega^i\right)Pr\left(\omega^i|W_{t-1}^X\right)\\
&-\sum_{i=X,Y}Pr\left(W_t^Y|W_{t-1}^X,\omega^i\right)Pr\left(\omega^i|W_{t-1}^X\right)\\
=&\left[g\left(W_t^X|h_{t-1}^X=\frac{1+p\cdot\widehat{m}_t}{2},\omega^X\right)+g\left(W_t^Y|h_{t-1}^X=\frac{1+p\cdot\widehat{m}_t}{2},\omega^X\right)\right]Pr\left(\omega^X|W_{t-1}^X\right)\\
&-\left[g\left(W_t^X|h_{t-1}^X=\frac{1-p\cdot\widehat{m}_t}{2},\omega^Y\right)+g\left(W_t^Y|h_{t-1}^X=\frac{1-p\cdot\widehat{m}_t}{2},\omega^Y\right)\right]Pr\left(\omega^Y|W_{t-1}^X\right)\\
=&\left[s-1+p\cdot d\cdot\widehat{m}_t\right]Pr\left(\omega^X|W_{t-1}^X\right)\\
&+\left[1-s+p\cdot d\cdot\widehat{m}_t\right]Pr\left(\omega^Y|W_{t-1}^X\right)\\
=&\left[s-1+p\cdot d\cdot\widehat{m}_t\right]Pr\left(W_{t-1}^X|\omega^X\right)\\
&+\left[1-s+p\cdot d\cdot\widehat{m}_t\right]Pr\left(W_{t-1}^Y|\omega^X\right)\\
=&(s-1)I_t+p\cdot d\cdot\widehat{m}_t,
\end{aligned}
$$

where the second to last equality follows by Bayes' rule and symmetry. Therefore, students apply to their non-local winning school if and only if their transport costs satisfy $c(\lambda)\leq C_t$, where

$$C_t\left(\widehat{m}_t,I_t\right)=V\cdot\left[(s-1)I_t+p\cdot d\cdot\widehat{m}_t\right].$$

Hence, the optimal strategy profile given informativeness is

$$m_t=F\left(V\cdot\left[(s-1)I_t+p\cdot d\cdot m_t\right]\right).$$

Steady-state informativeness $I(m)$ is still given by (57). I will allow for $m_t\in[-1,1]$ where a negative level of mobility represents the situation in which a share $|m_t|$ of students in period t apply to their non-local losing school. It is still true that $I(m)\in[0,1]$ and I will allow F to evaluate negative values such that $F(-c)=-F(c)$. Equilibrium steady-state is characterised by

$$
\begin{aligned}
m^*&=F\left(V\cdot\left[(s-1)I(m^*)+p\cdot d\cdot m^*\right]\right)\\
&=F\left(V\cdot\left[(s-1)\frac{s-1}{1-p\cdot d\cdot m^*}+p\cdot d\cdot m^*\right]\right)
\end{aligned}
\tag{67}
$$

Since $F\left(V\cdot\left[(s-1)\frac{s-1}{1-p\cdot d\cdot m}+p\cdot d\cdot m\right]\right)$ is weakly increasing in m and since m is bounded, Tarski's fixed point theorem implies that there exists some m^* such that (67) is satisfied.

B.2.1 Proposition 7

Define

$$\Gamma \equiv F\left(V \cdot \left[(s-1)\frac{s-1}{1-p\cdot d\cdot m} + p\cdot d\cdot m\right]\right).$$

1. For any $m \in [0,1]$, $\frac{\partial}{\partial m}\Gamma \geq 0$, $\frac{\partial}{\partial d}\Gamma \geq 0$ and $\frac{\partial}{\partial p}\Gamma \geq 0$, therefore the smallest equilibrium level of mobility increases in d or in p after Corollary 1 in Milgrom and Roberts [1994]. Given (66), the smallest equilibrium level of informativeness also increases in d or in p.

2. A negative equilibrium level of mobility exists if and only if $m = \Gamma(m)$ for some $m < 0$. Since there exist feasible parameter combinations such that $V \cdot \left[(s-1)\frac{s-1}{1-p\cdot d\cdot m} + p\cdot d\cdot m\right] < 0$ for $m < 0$, there exists some F such that $m = \Gamma(m)$ for some $m < 0$. Such a fixed point is weakly more likely to exist the lower $V \cdot \left[(s-1)\frac{s-1}{1-p\cdot d\cdot m} + p\cdot d\cdot m\right]$ at any $m < 0$. Therefore, it is more likely to exist if d increases or if p increases.

B.3 Robustness Checks

B.3.1 Longer window of Rankings

Denote the realisation of the past two rankings by a pair $(W_{t-1}, W_{t-2}) = \left(W^i_{t-1}, W^j_{t-2}\right)$, where $i, j = X, Y$. Given $(W_{t-1}, W_{t-2}) = \left(W^i_{t-1}, W^j_{t-2}\right)$, denote by m_{ij} the share of non-local students who apply to W^i_{t-1} and denote the level of informativeness by

$$I_{ij} \equiv Pr\left(\omega^i | W^i_{t-1}, W^j_{t-2}\right) - Pr\left(\omega^j | W^i_{t-1}, W^j_{t-2}\right).$$

Proposition 8 *A steady-state equilibrium vector of mobility $m^* = (m^*_{XX}, m^*_{XY}, m^*_{YX}, m^*_{YY})$ is characterised by*

$$m^*_{XX} = F\left[V \cdot \frac{s-1}{1-p\cdot d\cdot m^*_{XX}}\right] \tag{68}$$

and

$$m^*_{XX} = m^*_{YY}$$

and

$$m^*_{XY} = m^*_{YX} = 0,$$

and the corresponding steady-state equilibrium vector of informativeness is given by

$$I_{XX}(m^*) = I_{YY}(m^*) = \frac{s-1}{1-p\cdot d\cdot m^*} \tag{69}$$

and

$$I_{XY}(m^*) = I_{YX}(m^*) = 0.$$

Such equilibrium steady-state vectors of mobility (and informativeness) always exist.

Optimal mobility is given by $m_{ij} = F\left(V \cdot I_{ij}\right)$ for the same reason as in Lemma 1. Next, I will find informativeness in terms of mobility. First, derive steady-state informativeness at time-invariant vector of mobility levels $\widehat{m} = (\widehat{m}_{XX}, \widehat{m}_{XY}, \widehat{m}_{YX}, \widehat{m}_{YY})$. In a symmetric steady-state equilibrium, students' optimal mobility does not depend on the identity of the winning schools, but only on whether the same or a different school won in period $t-1$ than in period $t-2$. Hence, I restrict attention to $\widehat{m}_{XX} = \widehat{m}_{YY}$ and $\widehat{m}_{XY} = \widehat{m}_{YX}$.

Suppose the state is ω^X. The stationary distribution of (W_{t-1}, W_{t-2}) is defined by the vector ρ:

$$\rho \equiv \left(Pr\left(W^X_{t-1}, W^X_{t-2}\right), Pr\left(W^X_{t-1}, W^Y_{t-2}\right), Pr\left(W^Y_{t-1}, W^X_{t-2}\right), Pr\left(W^Y_{t-1}, W^Y_{t-2}\right)\right).$$

The Markov process of (W_{t-1}, W_{t-2}) is defined by the following matrix T of transition probabilities:

$$T = \begin{pmatrix} \frac{1}{2}(s+\widehat{m}_{XX}pd) & 0 & 1-\frac{1}{2}(s+\widehat{m}_{XX}pd) & 0 \\ \frac{1}{2}(s+\widehat{m}_{XY}pd) & 0 & 1-\frac{1}{2}(s+\widehat{m}_{XY}pd) & 0 \\ 0 & \frac{1}{2}(s-\widehat{m}_{XY}pd) & 0 & 1-\frac{1}{2}(s-\widehat{m}_{XY}pd) \\ 0 & \frac{1}{2}(s-\widehat{m}_{XX}pd) & 0 & 1-\frac{1}{2}(s-\widehat{m}_{XX}pd) \end{pmatrix}.$$

The Markov process of (W_{t-1}, W_{t-2}) has a unique stationary distribution defined by the row vector ρ that satisfies both

$$\rho = \rho T \tag{70}$$

and

$$Pr\left(W^X_{t-1}, W^X_{t-2}\right) + Pr\left(W^X_{t-1}, W^Y_{t-2}\right) + Pr\left(W^Y_{t-1}, W^X_{t-2}\right) + Pr\left(W^Y_{t-1}, W^Y_{t-2}\right) = 1. \tag{71}$$

By Bayes' rule:

$$Pr\left(\omega^X | W^i_{t-1}, W^j_{t-2}\right) = \frac{Pr\left(W^i_{t-1}, W^j_{t-2} | \omega^X\right) Pr\left(\omega^X\right)}{Pr\left(W^i_{t-1}, W^j_{t-2} | \omega^X\right) Pr\left(\omega^X\right) + Pr\left(W^i_{t-1}, W^j_{t-2} | \omega^Y\right) Pr\left(\omega^Y\right)}.$$

The same reasoning holds for ω^Y.

Solving for posterior beliefs yields:

$$I_{XY}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)=I_{YX}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)=0$$

and

$$I_{YY}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)=I_{XX}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)=\frac{s-1}{1-\widehat{m}_{XX}\cdot d\cdot p}\geq 0.$$

If $\widehat{m}_{XY}$ increases, then the levels of informativeness $I_{XY}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)$ and $I_{XX}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)$ remain unchanged. If $\widehat{m}_{XX}$ increases, $I_{XY}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)$ remains unchanged, while $I_{XX}\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)$ weakly increases. Hence, the vector of optimal mobilities $\left(\overline{m}_{XX},\overline{m}_{XY}\right)$ increases in the vector of levels of informativeness $\left(I_{XX},I_{XY}\right)$, which increases in the vector of steady-state mobilities $\left(\widehat{m}_{XX},\widehat{m}_{XY}\right)$. In addition, $\left(m_{XX},m_{XY}\right)\in[0,1]^2$.Therefore, Tarski's fixed point theorem applies.

B.3.2 Admission not based on Ability

Suppose a student of ability $A=H,L$ derives benefit V_A of enrolling at the better school, where $1>V_H>V_L>0$. Assume that an oversubscribed school uses a lottery to assign places to applicants, and that a student's chances of being allocated a place conditional on applying are independent of their location and their ability. Let m_A denote the share of students of ability $A=H,L$ who apply to their non-local winning school. Suppose F is Uniform on $[0,1]$, such that $F(c)=c$.

Proposition 9 *A steady-state equilibrium level of informativeness is given by*

$$I^*=\frac{(s-1)\left(2+\left(V_H+V_L\right)\cdot I^*\right)}{2+\left(V_H+V_L\right)\cdot I^*-2d\left(\left(V_H-V_L\right)\cdot I^*\right)}$$

and the corresponding steady-state equilibrium vector of mobility is characterised by

$$\left(m_H^*,m_L^*\right)=\left(V_H\cdot I^*,V_L\cdot I^*\right).$$

Such equilibrium steady-state level of informativeness (and vector of mobility) always exist.

By Lemma 1, optimal mobility of students in period t in terms of period-t informativeness is given by:

$$\left(m_{t,H},m_{t,L}\right)=\left(V_H\cdot I_t,V_L\cdot I_t\right). \tag{72}$$

The stationary distribution of the signal $Pr\left(W^i|\omega^i\right)$ for $i=X,Y$ at time-invariant mobility vector

$\widehat{m} = (\widehat{m}_H, \widehat{m}_L)$ satisfies

$$\begin{aligned} Pr\left(W^i|\omega^i\right) =& g\left(W^i|h^i = \frac{1+\widehat{m}_H}{2+\widehat{m}_H+\widehat{m}_L}, \omega^i\right) Pr\left(W^i|\omega^i\right) \\ &+ g\left(W^i|h^i = \frac{1+\widehat{m}_L}{2+\widehat{m}_H+\widehat{m}_L}, \omega^i\right)\left(1-Pr\left(W^i|\omega^i\right)\right) \end{aligned}$$

where the most recent winner chooses at random from a mass $\frac{1}{2}+\frac{1}{2}\widehat{m}_H$ of high-ability applicants and mass $\frac{1}{2}+\frac{1}{2}\widehat{m}_L$ of low-ability applicants. By Bayes' rule and symmetry, informativeness in steady state is given by:

$$I\left(\widehat{m}_H, \widehat{m}_L\right) \quad = \quad \frac{(s-1)\left(2+\widehat{m}_H+\widehat{m}_L\right)}{2+\widehat{m}_H+\widehat{m}_L-2d\left(\widehat{m}_H-\widehat{m}_L\right)}. \tag{73}$$

Note that if $\widehat{m}_H = \widehat{m}_L$ then $I\left(\widehat{m}_H, \widehat{m}_L\right) = s-1$ which is the level of informativeness if intakes are balanced across schools.

Given (72) and (73), equilibrium steady-state is characterised by

$$I^* = \frac{(s-1)\left(2+\left(V_H+V_L\right)\cdot I^*\right)}{2+\left(V_H+V_L\right)\cdot I^*-2d\left(\left(V_H-V_L\right)\cdot I^*\right)}.$$

Define

$$\Gamma(I) \equiv \frac{(s-1)\left(2+\left(V_H+V_L\right)\cdot I\right)}{2+\left(V_H+V_L\right)\cdot I-2d\left(V_H-V_L\right)\cdot I}. \tag{74}$$

$$\frac{\partial \Gamma(I)}{\partial I} = \frac{(s-1)\,4d\left(V_H-V_L\right)}{\left(2+\left(V_H+V_L\right)\cdot I-2d\left(V_H-V_L\right)\cdot I\right)^2} \geq 0,$$

since $d \geq 0$, $s > 1$ and $V_H > V_L$. Since I is bounded, by Tarski's fixed point theorem, a fixed point I^* satisfying $I^* = \Gamma(I^*)$ exists.

B.4 General Signal Distribution

To simplify notation, define $g : h \to [0,1]$ to be the probability that the better school wins in period t if a share h of high-ability students is enrolled at the better school, i.e.

$$g(h) \equiv Pr\left(W^X|h^X_{t-1} = h, \omega^X\right) = Pr\left(W^Y|h^Y_{t-1} = h, \omega^Y\right)$$

for $i = X, Y$ and $t = 0, 1, \ldots$. For any $\kappa \in [0,1]$, property (2) corresponds to

$$g(h) > 1 - g(1-h) \tag{75}$$

and property (3) corresponds to

$$\frac{\partial}{\partial h} g(h) \geq 0. \tag{76}$$

Define $u(m) = \frac{1+p \cdot m}{2} \in [0,1]$. Then $u(-m) = 1 - u(m)$.

Lemma 2A: *The unique steady-state level of informativeness given mobility $m_t = \widehat{m}$* is

$$I(\widehat{m}) = \frac{g(u(\widehat{m})) - (1 - g(1 - u(\widehat{m})))}{g(1 - u(\widehat{m})) + 1 - g(u(\widehat{m}))} \geq 0. \tag{77}$$

Proof Lemma 2A: A general solution to (23) is given by

$$Pr\left(W^X|\omega^X\right) = Pr\left(W^Y|\omega^Y\right) = \frac{g(1 - u(\widehat{m}))}{1 - g(u(\widehat{m})) + g(1 - u(\widehat{m}))}. \tag{78}$$

Given (24), informativeness follows. $I(\widehat{m}) \geq 0$ since $g(u(\widehat{m})) > (1 - g(u(\widehat{m})))$ by (75) and $g(1 - u(\widehat{m})) - g(u(\widehat{m})) \leq 0$ given $g \in [0,1]$.

Proposition 1A [Equilibrium]: *A steady-state equilibrium level of mobility m^* is characterised by*

$$m^* = F\left(V \cdot \frac{g(u(m^*)) - (1 - g(1 - u(m^*)))}{g(1 - u(m^*)) + 1 - g(u(m^*))}\right) \tag{79}$$

and the corresponding steady-state equilibrium level of informativeness $I(m^*)$ is given by

$$I(m^*) = \frac{g(u(m^*)) - (1 - g(1 - u(m^*)))}{g(1 - u(m^*)) + 1 - g(u(m^*))}. \tag{80}$$

Such a steady-state equilibrium level of mobility (and informativeness) always exists.

Proof of Proposition 1A: Substitute (77) into (16). Follow the proof of Proposition 1 given

$$\frac{\partial}{\partial \widehat{m}} I(\widehat{m}) \geq 0$$

$$\Leftrightarrow \left[\frac{\partial}{\partial \widehat{m}} g(u(\widehat{m})) - \frac{\partial}{\partial \widehat{m}} g(1 - u(\widehat{m}))\right] [g(u(\widehat{m})) - (1 - g(1 - u(\widehat{m})))] \geq 0,$$

since $g(u(\widehat{m})) - (1 - g(1 - u(\widehat{m}))) > 0$ by property (75) and both $\frac{\partial}{\partial \widehat{m}} (g(u(\widehat{m})) - g(1 - u(\widehat{m}))) \geq 0$ by property (76).

Proposition 2A [Convergence] *Take Proposition 2 and substitute (29) for*

$$\begin{aligned} I_t =& I_{t-1}\left(g\left(u\left(F\left(V\cdot I_{t-1}\right)\right)\right)-g\left(1-u\left(F\left(V\cdot I_{t-1}\right)\right)\right)\right)\\ &+g\left(u\left(F\left(V\cdot I_{t-1}\right)\right)\right)-\left(1-g\left(1-u\left(F\left(V\cdot I_{t-1}\right)\right)\right)\right). \end{aligned} \tag{81}$$

Proof Proposition 2A: The general recurrence equation (81) is constructed using (38) and (78):

$$\begin{aligned} I_t = 2Pr\left(W_{t-1}^X|\omega^X\right)-1 =& \left(g\left(u\left(m_{t-1}\right)\right)-g\left(1-u\left(m_{t-1}\right)\right)\right)2Pr\left(W_{t-2}^X|\omega^X\right)\\ &+2g\left(1-u\left(m_{t-1}\right)\right)-1\\ =& I_{t-1}\left(g\left(u\left(m_{t-1}\right)\right)-g\left(1-u\left(m_{t-1}\right)\right)\right)\\ &+g\left(u\left(m_{t-1}\right)\right)-\left(1-g\left(1-u\left(m_{t-1}\right)\right)\right). \end{aligned}$$

Since $m_{t-1}=F\left(V\cdot I_{t-1}\right)$, (81) follows. Let

$$\begin{aligned} Z\left(I_{t-1}\right) =& g\left(u\left(F\left(V\cdot I_{t-1}\right)\right)\right)-\left(1-g\left(1-u\left(F\left(V\cdot I_{t-1}\right)\right)\right)\right)\\ &+\left(g\left(u\left(F\left(V\cdot I_{t-1}\right)\right)\right)-g\left(1-u\left(F\left(V\cdot I_{t-1}\right)\right)\right)\right)\cdot I_{t-1}\in[0,1] \end{aligned} \tag{82}$$

$Z\left(I_{t-1}\right)$ is increasing in I_{t-1}:

$$\begin{aligned} &\frac{\partial}{\partial I_{t-1}}Z\left(I_{t-1}\right)=g\left(u\left(F\left(V\cdot I_{t-1}\right)\right)\right)-g\left(1-u\left(F\left(V\cdot I_{t-1}\right)\right)\right)\\ &+\frac{\partial}{\partial I_{t-1}}\left(g\left(u\left(F\left(V\cdot I_{t-1}\right)\right)\right)-g\left(1-u\left(F\left(V\cdot I_{t-1}\right)\right)\right)\right)\cdot I_{t-1}\geq 0, \end{aligned}$$

since $F(\cdot)$ is positive and increasing, $V>0$ and $I_{t-1}\geq 0$ and given property (3). By property (2), it follows that

$$I_1 = Z\left(I_0\right) = g\left(\frac{1}{2}\right)-\left(1-g\left(\frac{1}{2}\right)\right)\geq I_0=0.$$

The remaining steps are as in the proof of Proposition 2.

Proof of Theorem 1: Define

$$\Gamma\left(I,p,F,V\right)=\frac{g\left(u\left(F\left(V\cdot I\right)\right)\right)-\left(1-g\left(1-u\left(F\left(V\cdot I\right)\right)\right)\right)}{1-g\left(u\left(F\left(V\cdot I\right)\right)\right)+g\left(1-u\left(F\left(V\cdot I\right)\right)\right)}.$$

$\Gamma\left(z,I\right)$ is increasing in I at any $I\in[0,1]$ since $\frac{d\Gamma}{dI}=\frac{\partial\Gamma}{\partial u}\frac{\partial u}{\partial I}$, $\frac{d\Gamma}{du}\geq 0$ by Lemma 2A and $\frac{\partial u}{\partial I}\geq 0$ since $p\geq 0$, $V>0$ and $F(\cdot)$ is increasing.

1. $\Gamma\left(p,I\right)$ is increasing in p since $\frac{d\Gamma}{dp}=\frac{\partial\Gamma}{\partial u}\frac{\partial u}{\partial p}$ and $\frac{d\Gamma}{du}\geq 0$ by Lemma 2A and $\frac{\partial u}{\partial p}\geq 0$ since

$I \geq 0$, $V > 0$ and $F(\cdot)$ is increasing.

2. If $F(\cdot)$ first-order stochastically dominates $\widetilde{F}(\cdot)$, then $u(F(V \cdot I)) \leq u\left(\widetilde{F}(V \cdot I)\right)$. Since $\Gamma(F, I)$ increases in u, this implies that $\Gamma(F, I) \leq \Gamma\left(\widetilde{F}, I\right)$.

The remaining steps are as in the proof of Theorem 1 for the special function form.

Proof of Proposition 3: For any $h \in [0,1]$

$$2Pr\left(\omega^X | W^X\right) - 1 = 2Pr\left(W^X | \omega^X\right) - 1 = \frac{g(h) - (1 - g(1-h))}{1 - g(h) + g(1-h)} \geq 0 \tag{83}$$

since (77) is positive at $\widehat{m} = \frac{2h-1}{p}$ as shown in the proof of Lemma 2A.

For any $h \in [0,1]$

$$\frac{\partial}{\partial h} Pr\left(\omega^X | W^X\right) \geq 0 \tag{84}$$

since (77) is increasing in $\widehat{m}$ and $\widehat{m} = \frac{2h-1}{p}$ as shown in the proof of Proposition 1A.

References

A. Abdulkadiroglu and T. Sönmez. School choice: A Mechanism Design Approach. *The American Economic Review*, 93(3):729–747, 2003.

R. Allen, S. Burgess, and L. McKenna. School Performance and Parental Choice of School: Secondary Data Analysis. 2014.

L. Barseghyan, D. Clark, and S. Coate. Public School Choice: An Economic Analysis. Working Paper 20701, National Bureau of Economic Research, November 2014.

P. Bayer and R. McMillan. Choice and Competition in Local Education Markets. Working Paper 11802, National Bureau of Economic Research, November 2005.

S. Bikhchandani, D. Hirshleifer, and I. Welch. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

S. Black. Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, 114(2):577–599, 1999.

N. Bloom, R. Lemos, R. Sadun, and J. Van Reenen. Does Management Matter in Schools. Working Paper 20667, National Bureau of Economic Research, November 2014.

S. Burgess and R. Allen. The Future of Competition and Accountability in Education. *2020 Public Services Trust*, 2010.

S. Burgess, A. Briggs, B. McConnell, and H. Slater. *School choice in England: Background facts*. Centre for Market and Public Organisation, University of Bristol Bristol, 2006.

S. Burgess, E. Greaves, A. Vignoles, and D. Wilson. What Parents Want: School Preferences and School Choice. *The Economic Journal*, 125(587):1262–1289, 2015.

S. Callander and J. Hörner. The Wisdom of the Minority. *Journal of Economic Theory*, 144(4): 1421–1439, 2009.

C. Calsamiglia and M. Güell. The Illusion of School Choice: Empirical Evidence from Barcelona. 2014.

A. Chandra, A. Finkelstein, A. Sacarny, and C. Syverson. Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector. *American Economic Review*, 106(8): 2110–44, 2016.

J. Coldron, E. Tanner, S. Finch, L. Shipton, C. Wolstenholme, B. Willis, S. Demack, and B. Stiell. Secondary School Admissions. *DCSF research report*, 2008.

G. De Fraja and P. Landeras. Could do better: The Effectiveness of Incentives and Competition in Schools. *Journal of Public Economics*, 90(1):189–213, 2006.

D. Dranove, D. Kessler, M. McClellan, and M. Satterthwaite. Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy*, 111(3): 555–588, 2003.

A. Gavazza and A. Lizzeri. The Perils of Transparency in Bureaucracies. *American Economic Review*, 97(2):300–305, 2007.

M. Gaynor, C. Propper, and S. Seiler. Free to choose? Reform and Demand Response in the English National Health Service. Technical report, National Bureau of Economic Research, 2012.

H. Gravelle and P. Sivey. Imperfect Information in a Quality-competitive Hospital Market. *Journal of Health Economics*, 29(4):524–535, 2010.

J. Hastings and J. Weinstein. Information, School Choice, and Academic Achievement: Evidence from Two Experiments. NBER Working Papers 13623, National Bureau of Economic Research, Inc, November 2007.

J. Hastings, T. Kane, and D. Staiger. Parental Preferences and School Competition: Evidence from a Public School Choice Program. Technical report, National Bureau of Economic Research, 2005.

C. Hoxby. Does Competition among Public Schools Benefit Students and Taxpayers? *American Economic Review*, 90(5):1209–1238, 2000.

C. Hoxby. The Changing Selectivity of American Colleges. *Journal of Economic Perspectives*, 23 (4):95–118, December 2009.

S. Imberman and M. Lovenheim. Does the Market Value Value-Added? Evidence from Housing Prices After a Public Release of School and Teacher Value-Added. Working Paper 19157, National Bureau of Economic Research, June 2013.

T. Kane, D. Staiger, D. Grissmer, and H. Ladd. Volatility in School Test Scores: Implications for Test-Based Accountability Systems. *Brookings Papers on Education Policy*, (5):235–283, 2002.

D. Kessler and M. McClellan. Is Hospital Competition Socially Wasteful? *The Quarterly Journal of Economics*, 115(2):577–615, 2000.

I. Lobel, D. Acemoglu, M. Dahleh, and A. Ozdaglar. Preliminary Results on Social Learning with Partial Observations. In *Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools*, 2007.

M. A. Meyer. Learning from Coarse Information: Biased Contests and Career Profiles. *The Review of Economic Studies*, 58(1):15–41, 1991.

P. Milgrom and J. Roberts. Comparing Equilibria. *The American Economic Review*, 84(3):441–459, 1994.

P. Noden, A. West, and A. Hind. Banding and Ballots. *The Sutton Trust*, 2014.

J. Rothstein. Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions. *American Economic Review*, 96(4):1333–1350, 2006.

D. Wilson and A. Piebalga. Accurate Performance Measure but Meaningless Ranking Exercise? An Analysis of the English School League Tables. Technical report, Department of Economics, University of Bristol, UK, 2008.