Performance-Based Rankings and School Quality*

Claudia Herresthal†

13th June 2019

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 qualities, but this signal also depends on the abilities 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 Pawel Gola and Ricardo Alonso, Leonie Baumann, Inga Deimen, Peter Esö, 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

Every year school rankings are released to help prospective students compare schools in terms of the quality of their teaching. However, schools are ranked based on their performance, which depends not only on their intrinsic qualities but also on the ability of their past student intakes.[1] This paper studies how students learn about schools' qualities from rankings if they cannot observe which schools had the most able intakes in the past.[2] In addition, I analyse how features of the admission and application process affect student intakes and, therefore, influence how informative rankings are about school quality.

I use a dynamic setting to study the interaction between student intakes and rankings. In each cohort, students first observe a ranking of schools and then apply to one of two schools of unknown quality. A school admits all applicants unless it is oversubscribed, in which case it selects among applicants based on an exogenously given admission rule. The probability that a school ranks high depends both on its quality and on the ability of its student intake in the previous cohort. In particular, if intakes across schools were equally able, then the better school is more likely to rank high. In addition, if a school had a more able intake, its chances of ranking high increase, irrespective of its quality. Specifically, if the worse school's intake was sufficiently more able, the ranking is a misleading indicator of quality: The worse school is *more* likely to rank high than the better school. Students do not observe past intakes or past applications.

I study how schools' admission rule affects students' learning about school quality over time. Recent changes to the School Admissions Code in England have restricted the share of students admitted based on ability,[3] and it has been argued that assigning more equal intakes across schools would make it more likely that differences in school quality are reflected in rankings.[4] This seems intuitive at first. Suppose a school's admission rule gives priority to higher-ability applicants when a school is oversubscribed. If students believe that the high-ranked school is more likely to be better, then the high-ranked school will be oversubscribed and admit a more able intake than the other school. Therefore, we could end up in a situation in which the worse school persistently ranks high because its lower quality is covered up by a steady intake of more able students.

Despite this, I show that over time students are strictly more likely to correctly infer which school is of better quality if the admission rule gives priority to higher-ability applicants than if it assigns equally able intakes to each school. To gain some intuition for why this holds, consider the following example. Each cohort consists of two students, one of high and one of low ability. Each

[1] 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.

[2] 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]).

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

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

school has capacity for one student. If both students apply, the high-ability student is accepted and the low-ability student attends the other school. A priori students believe that each school is equally likely to be better. In period 0, the high-ability student is equally likely to attend either school.[5] Students in period 1 observe the ranking in period 0. Since the better school was equally likely to teach the high- or the low-ability student in period 0, they infer that the high-ranked school is more likely to be better. Hence, both students apply to the high-ranked school and the high-ability student is accepted. Then students in period 2 observe the ranking in period 1. Although they do not observe where the high-ability student in period 1 applied or enrolled, they conjecture correctly that he was more likely to attend the better school than the high-ability student in period 0. Therefore, they infer that the better school is even more likely to rank high in period 1 than in period 0. Consequently, students in period 2 are more likely to correctly identify the better school than students in period 1, even though both cohorts observe only the most recent ranking. If an oversubscribed school instead accepted the high- or the low-ability student with equal probability in each cohort, students in period 2 would be no better informed than students in period 1.[6]

Indeed, even conditional on the event that the worse school ranks high initially, students are more likely to correctly identify which school is better over time if the admission rule prioritises high-ability applicants. It is true that, once the worse school ranks high, it is more likely to rank high again if it admits a more able intake than the other school. However, the worse school will not hold on to the high rank forever. In addition, once the better school ranks high, it also will be more likely to rank high again if it admits a more able intake than the other school. In this event, not only the more able intake but also the superior quality work in favour of the better school. This implies that the better school is more likely to maintain a high rank than the worse school. Hence, over time, the better school is more likely to rank high, and the ranking is more likely to reflect schools' relative quality accurately if the admission rule prioritises high-ability applicants.

In addition, I study how students' learning about school quality is affected by the extent to which students perceive schools to be horizontally differentiated, whether due to their location, specialisation or other factors. I introduce horizontal differentiation by assuming that students are distributed along a Hotelling line between the two schools and that each student incurs a transport cost proportional to the distance to the school he attends. Therefore, a student may face a trade-off between attending a school that is closer and a school that is of higher expected quality. I 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), i.e. if students perceive schools to

[5] This allocation arises if applications strategies in period 0 are symmetric with respect to schools' identities.

[6] In the paper, I focus on a steady state in which the application strategy profile of students is constant across periods. To construct students' posterior beliefs in such a steady state, it is not necessary to reason through the choices of students in each previous period as in the example above. Instead, it is possible to directly derive the stationary probability that the better school ranks high and then construct students' posterior beliefs about schools' relative quality.

be less horizontally differentiated. Recent school choice reforms can be represented by a decrease in transport costs, since they allow students to attend a non-local school instead of requiring them to move into the school's catchment area. Therefore, my findings show that school choice reforms interact with performance-based rankings to make it easier for students to identify better schools.

Finally, I 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 the demand for its places, thereby putting pressure on unpopular schools to make changes 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 students' inference about school quality from rankings, which affects demand for places, and the supply side response, which affects the qualities of schools. I find that introducing such a policy improves the average quality of schools, and improves access to good schools for students of any ability level. However, in the presence of such policies, i) if the admission rule assigns a larger share of spaces at an oversubscribed school based on ability or ii) if transport costs are lower, the average quality of schools improves further, but the additional benefit may 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.

This paper is structured as follows: Section 2 outlines the model. Sections 3-5 solve for equilibrium and conduct comparative statics and welfare analysis. Section 6 studies quasi-market reforms. Section 7 contains robustness checks and Section 8 concludes. Omitted proofs are in the appendix.

1.1 Related Literature

My paper relates to the observational learning literature,[7] because it studies a setting in which the inference by agents in the present is influenced by the choices of agents in the past. Past agents' choices are usually assumed to be observable and they convey information about past agents' private signals. 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. A limited window of observations is also studied by Lobel et al. [2007] but their focus is on conditions for convergence. Callander and Hörner [2009] propose a steady-state analysis, but agents infer information from the relative frequency with which actions were taken by predecessors.

In addition, my paper relates to Meyer [1991]'s work on biased contests, where a decision-maker (DM) sequentially designs contests to learn which of two (non-strategic) workers is of higher ability. The DM optimally assigns the bias in the last contest in favour of the worker he

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

believes to be of higher ability. The reason is that, if this worker loses despite the bias being in his favour, this is strong evidence that he is of lower ability. In my paper, learning about school quality improves if the school of higher expected quality admits a more able intake and, hence, enjoys a bias in its favour. However, the reason is different because students cannot condition their application choices on whether the better-performing school had a more able intake in the past. In addition, intakes are not assigned by a forward-looking DM, but determined by the application choices of short-sighted students who do not care about their effect on 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. 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 depends on whether schools select students based on ability. I show that when students infer school quality from rankings, policymakers cannot choose the admission rule without also affecting how 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 derives optimal algorithms assuming students have complete information about schools.[8] Yet if students are incompletely informed about schools' qualities, 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.[9]

My paper's predictions are consistent with 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.[10] 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.

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

[9] 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].

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

2 Model

Time is discrete and the horizon is infinite $t = 0, 1, \ldots$. A population of students is located on the interval $[0,1]$. The population is comprised of a unit mass of high-ability and a unit mass of low-ability students. A student's type is given by (λ, α), where $\lambda \in [0,1]$ is the student's location and $\alpha \in \{H, L\}$ is his ability, where H is high and L is low. The distribution of location parameter λ is continuous, symmetric about $\lambda = \frac{1}{2}$ and independent of ability. Each student lives for one period.

There are two schools, school 0 and school 1. School i is located at position $i \in \{0,1\}$ on the unit interval. There are two equiprobable states of the world, ω^0 and ω^1, where state ω^i means that school i is of better quality. The state is drawn at the start of period 0 and is unobserved.[11]

An action $a \in \{a^0, a^1\}$ for a student is to choose whether to apply to school 0 or 1. A *local* student for school $i \in \{0,1\}$ is a student whose nearest school is school i. Each school has a capacity of unit mass. Schools are non-strategic and admit students based on the following admission rule. Each school admits all 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 share $s \in [0,1]$ out of the pool of non-local high-ability applicants.[12] 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, it will prioritise high-ability over low-ability applicants. Rejected applicants enrol at the other school.[13]

At the end of each period t, after all students have enrolled at a school, a signal about schools' relative quality is realised. This signal is denoted by $W_t \in \{W_t^0, W_t^1\}$ and can be interpreted as a ranking of schools, where W_t^i means that school i ranks first, i.e. school i is the *winning school*. The signal distribution depends on the endogenous enrolment of students in period t and on school qualities, i.e. the state of the world.[14] The probability that a given school wins increases with the share of high-ability students enrolled at this school and with its quality.[15] Formally, the probability that the better school wins in period t is given by $p(\eta_t)$ where $p : [0,1] \to [0,1]$ and η_t is the share of high-ability students enrolled at the better school in period t. $p(\eta_t)$ is differentiable in η_t and satisfies the following two properties. First, a more able intake introduces an upward bias in

[11] If the state changes with some exogenously given probability $g \in \left[0, \frac{1}{2}\right)$ each period and such changes are unobserved, the results of this paper remain qualitatively unchanged (see Section 7.2).

[12] Hence, if $s = 0$, intake ability will be equal across schools, irrespective of students' application choices.

[13] The fact that schools select on ability is not critical. The crucial feature is that an oversubscribed school admits at least as large a share of high-ability applicants as the other school, as discussed in Section 7.3.

[14] School quality refers to features of the school that affect its performance for a given student intake, e.g. the quality of teaching and leadership. For evidence that these features have a large impact on performance, see e.g. Bloom et al. [2014].

[15] For evidence that there is residual noise in rankings see Kane et al. [2002].

performance for either type of school, that is for any $\eta_t \in [0,1]$,

$$\frac{\partial}{\partial \eta_t} p(\eta_t) \geq 0. \tag{1}$$

Second, assume a share κ_t of high-ability students are enrolled at school i, then school i is more likely to win if it is the better school than if it is the worse school, that is for any $\kappa_t \in [0,1]$,

$$p(\kappa_t) > 1 - p(1-\kappa_t).^{16} \tag{2}$$

In each period $t > 0$, students first observe the most recent signal W_{t-1} and then apply to schools.[17] Note that they do not observe any past actions or signals predating W_{t-1}.[18] A strategy for a student of type (λ, α) in period $t > 0$ is: $\sigma_{\lambda,\alpha,t} : \{W_{t-1}^0, W_{t-1}^1\} \to \Delta\{a^0, a^1\}$. It selects the probabilities with which the student applies to school 0 and school 1 conditional on which is the most recent winning school. Denote the profile of strategies in period t by $\overline{\sigma}_t$.

Applications are costless and a student's payoff only depends on the school at which he enrols. Let $u_{\lambda,\alpha,t}(\varepsilon^i, \omega) \in \mathbb{R}$ be the payoff of a student of type (λ, α) in period t enrolled in school i (ε^i) in state ω, where $i \in \{0,1\}$. The student derives benefit $V > 0$ if and only if he enrols at the better school. In addition, he incurs a cost equal to his distance from school i. Hence,

$$u_{\lambda,\alpha,t}(\varepsilon, \omega) = \begin{cases} V \cdot 1_{\omega=\omega^0} - \lambda & \text{if } \varepsilon = \varepsilon^0 \\ V \cdot 1_{\omega=\omega^1} - (1-\lambda) & \text{if } \varepsilon = \varepsilon^1 \end{cases} \tag{3}$$

where $1_{\omega=\omega^i} = 1$ if $\omega = \omega^i$ and 0 otherwise. Given W_{t-1}, his expected payoff from being enrolled at school i is given by

$$E\left(u_{\lambda,\alpha,t}(\varepsilon^i) | W_{t-1}\right) = \sum_{k=0,1} u_{\lambda,\alpha,t}\left(\varepsilon^i, \omega^k\right) Pr\left(\omega^k | W_{t-1}\right). \tag{4}$$

The expected payoff from a given application choice depends on the expected payoff from being enrolled at each school and the probability of being admitted at each school. Given W_{t-1}, the probability of being admitted at a given school depends on the student's type (λ, α), the profile of strategies of all other types in period t, $\overline{\sigma} \setminus \{\sigma_{\lambda,\alpha,t}\}$, and on the admission rule characterised by s. Each student chooses his strategy such as to maximise his expected payoff.

[16] Note that I do not assume that the better school's probability of winning increases more than the worse school's for a given increase in the share of high-ability students admitted.

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

[18] The insights still apply if students observed a longer window of rankings as long as they do not have access to all rankings available to students in previous periods (see Section 7.1). The most recent ranking is clearly the most easily accessible, e.g. in England this is the one circulated by the media.

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

I am interested in the Perfect Bayesian Equilibrium (PBE) as $t \to \infty$. For this reason, I use the concept of a steady-state equilibrium. In *steady state,* the strategy profile $\{\overline{\sigma}_t\}_{t\geq 0}$ is time-invariant, i.e. $\overline{\sigma}_t = \overline{\sigma}$ for any $t > 0$, and the distribution of rankings is stationary in any given state of the world. Denote beliefs of students in period t about the state of the world by $\mu_t : \left\{W^0_{t-1}, W^1_{t-1}\right\} \to \Delta\left\{\omega^0, \omega^1\right\}$. A *steady-state equilibrium* is a steady state with a strategy profile $\{\overline{\sigma}_t\}_{t\geq 0}$, where $\overline{\sigma}_t = \overline{\sigma}$ for any $t > 0$, and a system of beliefs $\{\mu_t\}_{t\geq 0}$, where $\mu_t = \mu$ for any $t \geq 0$, such that i) for any period $t > 0$ and type of student (λ, α), strategy $\sigma_{\lambda,\alpha}$ is optimal given the profile of strategies of all other students and given beliefs μ and ii) the system of beliefs $\{\mu_t\}_{t\geq 0}$ is derived from the stationary distribution using Bayes' rule. 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 $T+1, T+2, \ldots$ to deviate from the steady-state equilibrium strategy profile. I focus on strategy profiles which are symmetric with respect to schools' identities, i.e. $\sigma_{\lambda,\alpha,t}\left(W^i_{t-1}\right) = \sigma_{1-\lambda,\alpha,t}\left(W^j_{t-1}\right)$ for any $i, j \in \{0,1\}$ and $j \neq i$.

3 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 (Mobility) *Given school i is the winner in period t-1, mobility in period t, $m_t \in [0,1]$, is defined as the share of non-local students who apply to school i where $i \in \{0,1\}$.*

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 of rankings in steady state. Given the stationary probability that the better school ranks high, I will derive (time-invariant) posterior beliefs about the state of the world in terms of $\widehat{m}$ using Bayes' rule. Finally, I will solve for fixed points, i.e. I will find a time-invariant strategy profile m^* such that m^* is optimal given posterior

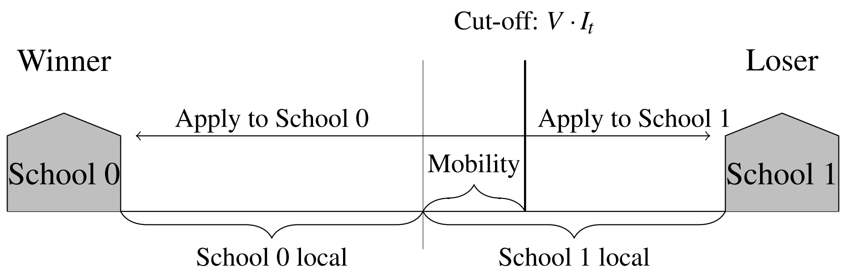

Figure 1: Students are distributed along a line between school 0 and school 1. Those with high transport costs live closer to their respective local school. The figure depicts the situation in which school 0 is the most recent winner and therefore receives applications from all local students, and from the share of non-local students whose transport costs lie below the cut-off $V \cdot I_t$. Mobility is defined as the share of non-local students who apply to the most recent winner.

beliefs and posterior beliefs are derived from the stationary probability that the better school ranks high 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 $t = 0$.

3.1 Optimal Mobility

First, I will derive students' optimal mobility given beliefs. Suppose students in period t believe that the most recent winner is more likely to be the better school and their beliefs are independent of whether the recent winner is school 0 or 1. It is helpful to introduce informativeness I_t as a shortcut for how these posterior beliefs enter students' expected payoff in period t.[19]

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

$$I_t \equiv Pr\left(\omega^i | W^i_{t-1}\right) - Pr\left(\omega^j | W^i_{t-1}\right) \tag{5}$$

for $i, j \in \{0,1\}$ and $i \neq j$, where $Pr\left(\omega^k | W^i_{t-1}\right)$ denotes the posterior belief of students in period t that school $k \in \{0,1\}$ is the better school conditional on observing that school i won in period $t-1$.

[19] 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 non-local students who apply to the most recent loser.

Lemma 1 (Optimal Mobility given Informativeness) *The optimal strategy profile of students in period t given informativeness I_t is given by*

$$\overline{m}_t = F\left(V \cdot I_t\right). \tag{6}$$

Proof: There is no downside for a student to apply to the school at which his expected payoff conditional on enrolment is higher.[20] By enrolling at a better school, a student gains benefit V. Hence, his expected benefit of enrolling at the winning school instead of the losing school is $V \cdot I_t \geq 0$. If the winning school is local, a student incurs no transport cost when attending the winning school. Therefore, any local student applies to this school. If the winning school is non-local, a student incurs transport costs $c(\lambda)$ when attending the winning school. Hence, any non-local student applies to the winning school if and only if the expected benefit of enrolling at the winning school rather than the losing school outweighs these transport costs $c(\lambda)$, which is the case for a share $F\left(V \cdot I_t\right)$ of non-local students.[21]$\square$

3.2 Steady-State Informativeness

Next, I will derive informativeness in steady state with a time-invariant strategy profile $\widehat{m}$. Given mobility $\widehat{m}$ and an admission rule characterised by s, denote the share of high-ability students enrolling at the most recent winner by $h(\widehat{m}, s)$, where $h : [0,1] \times [0,1] \to [0,1]$ and

$$h(\widehat{m}, s) = \frac{1 + s \cdot \widehat{m}}{2}. \tag{7}$$

Lemma 2 (Steady-State Informativeness given Mobility) *The unique steady-state level of informativeness given $m_t = \widehat{m}$ for all $t > 0$ is*

$$I(\widehat{m}) = \frac{p(h(\widehat{m}, s)) - (1 - p(1 - h(\widehat{m}, s)))}{p(1 - h(\widehat{m}, s)) + 1 - p(h(\widehat{m}, s))} > 0, \tag{8}$$

and satisfies

$$\frac{\partial I(\widehat{m})}{\partial \widehat{m}} \geq 0, \tag{9}$$

where the inequality is strict if and only if $s > 0$ and equation (1) holds with strict inequality.

In steady state, informativeness is derived from the stationary distribution of rankings using Bayes' rule. The distribution of ranking is stationary because, in a given state of the world, the

[20] In the worst case, his application gets rejected and he enrols at the other school, which results in the same expected payoff as if he had applied there.

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

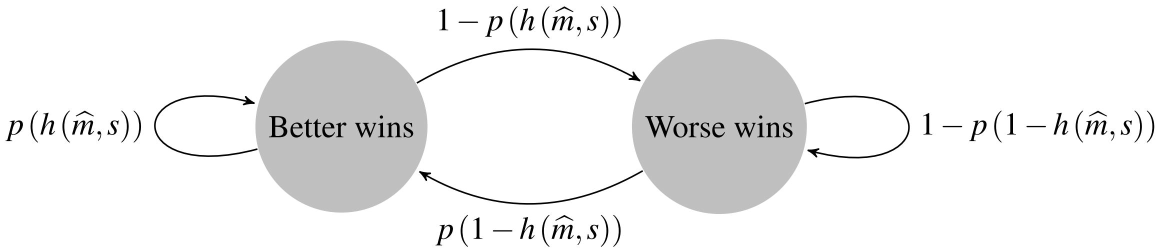

Figure 2: The graph illustrates the dynamics of the Markov process of ranking realisations. The arrows show the possible transitions between ranking realisations and the likelihood with which they occur, given time-invariant mobility level $\widehat{m}$.

sequence of ranking realisations $\{W_t\}_{t>0}$ follows a time-homogeneous Markov process. This is because the distribution from which W_t is drawn depends on the share of high-ability students at each school in period t, which depends only on the most recent realisation W_{t-1} and on mobility level $\widehat{m}$.

Informativeness weakly increases in the level of steady-state mobility $\widehat{m}$. To understand why, consider the transition probabilities of the time-homogeneous Markov process for a given state of the world (illustrated in Figure 2). At $\widehat{m}=0$, $h\left(0,s\right)=\frac{1}{2}$ and, hence, the probability that the better school wins in period t is independent of the ranking in period $t-1$. By equation (2), the better school wins more frequently than the worse school in steady state. By contrast, if $\widehat{m}$ rises above 0, then $h\left(\widehat{m},s\right)$ raises above $\frac{1}{2}$ and the school that currently ranks high is more likely to rank high again, whether it is of better or worse quality. Given the better school's performance advantage, this must raise the frequency with which the better school wins relative to the worse school in steady state. Consequently, the better school ranks high even more frequently than the worse school at a higher level of mobility, which raises informativeness.[22]

3.3 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 derived using Bayes' rule from the ranking distribution in steady state at mobility m^* (as illustrated in Figure 3):

$$m^* = F\left(V \cdot I\left(m^*\right)\right). \tag{10}$$

[22]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 7.1.

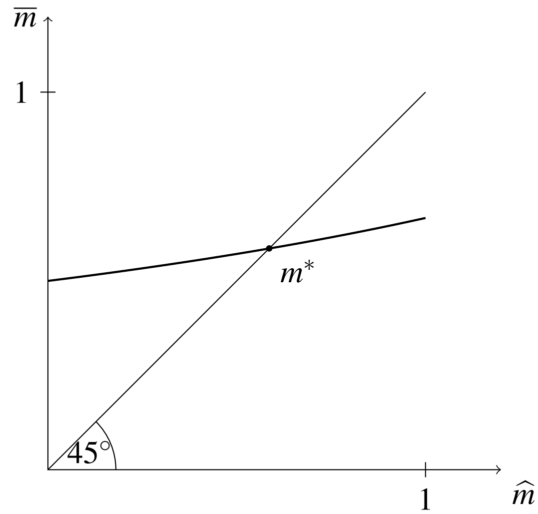

Figure 3: The graph shows the optimal mobility level $\overline{m}$, which characterises the optimal strategy profile of students in steady state, as a function of steady-state mobility level $\widehat{m}$. The intersection with the 45-degree line shows the steady-state equilibrium level of mobility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

Proposition 1 (Equilibrium) *A steady-state equilibrium level of mobility m^* is characterised by*

$$m^* = F\left(V \cdot \frac{p(h(m^*,s)) - (1 - p(1 - h(m^*,s)))}{p(1 - h(m^*,s)) + 1 - p(h(m^*,s))}\right) \tag{11}$$

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

$$I(m^*) = \frac{p(h(m^*,s)) - (1 - p(1 - h(m^*,s)))}{p(1 - h(m^*,s)) + 1 - p(h(m^*,s))}. \tag{12}$$

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

Proof: $I(\widehat{m})$ is increasing in $\widehat{m}$ given Lemma 2. Then $F(V \cdot I(\widehat{m}))$ is monotone increasing in $\widehat{m} \in [0,1]$ since $V > 0$ and $F(\cdot)$ is increasing. By Tarski's fixed point theorem, there exists an m^* such that $F(V \cdot I(m^*)) = m^*$.$\square$

3.4 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) 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.

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

$$I_0 = 0, \tag{13}$$

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

$$\begin{aligned} I_t = & I_{t-1}\left(p\left(h\left(m_{t-1},s\right)\right) - p\left(1-h\left(m_{t-1},s\right)\right)\right) \\ & + p\left(h\left(m_{t-1},s\right)\right) - \left(1-p\left(1-h\left(m_{t-1},s\right)\right)\right), \end{aligned} \tag{14}$$

where $m_{t-1} = F\left(V \cdot I_{t-1}\right)$.

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

$$m^{1*} = F\left(V \cdot I^{1*}\right). \tag{15}$$

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. 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.[23]

4 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 3.4).

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

Theorem 1 (Comparative Statics) *The equilibrium levels of mobility and informativeness as well as the share of high-ability students attending the better school rises with*

1. *an increase in the share s 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.*

These exogenous changes raise the share of high-ability students who enrol at the most recent winning school for any given posterior belief about school qualities. This raises optimal mobility at any given level of steady-state mobility and, hence, equilibrium mobility rises (as shown in Figure 4).

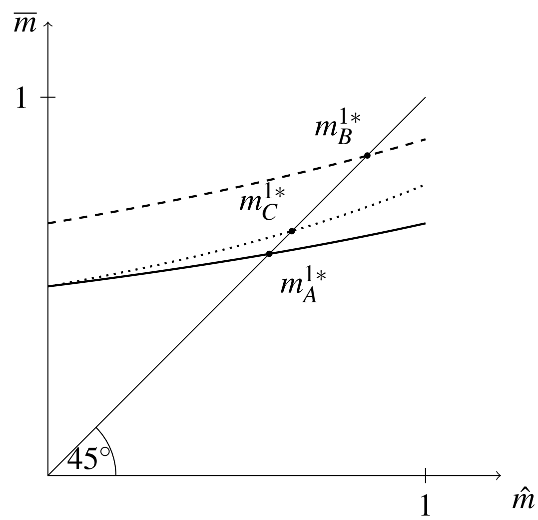

Figure 4: Consider the equilibrium mobility m_A^{1*}. A negative shift in the sense of FOSD of the distribution $F(\cdot)$ shifts up optimal mobility at any level of steady-state mobility, as illustrated by the dashed line. Equilibrium mobility increases to m_B^{1*}. An increase in s shifts up optimal mobility at any positive level of steady-state mobility, as illustrated by the dotted line. Equilibrium mobility increases to m_C^{1*}. The solid line is drawn for F being a Uniform distribution on $[0, \overline{c}]$, $\overline{c} = 1$, $V = 2$, $\theta = \frac{1}{2}$, $p(h) = \frac{1+h}{2}$. The dashed line is drawn for $\frac{V}{\overline{c}} = \frac{16}{3}$, and the dotted line for $s = \frac{7}{10}$, all else equal.

My results show that an admission rule which assigns priority to high-ability students ($s > 0$) facilitates learning about school quality relative to an admission rule which assigns equally able intakes to each school ($s = 0$). This is contrary to what has been suggested in recent discussions on how intakes should be assigned to improve the informativeness of rankings. 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 on how the ranking realisation depends on the ability of exogenously 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, my results show that learning about school quality from rankings is facilitated if schools are perceived to be less horizontally differentiated. Recent school choice reforms reduced the cost of choosing a school that is further from home, 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, 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, in areas where schools are more homogeneous, e.g. in terms of the curriculum they offer, rankings should be more informative about schools' relative quality. This suggests that there exists a trade-off as schools diverge in their specialisations: students are more likely to find a school that caters to their desired specialisation, 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.[24]

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

4.1 Discussion

Both exogenous changes, a decrease in transport costs and a rise in the share of non-local high-ability applicants admitted by an oversubscribed school, have the following feature in common: at any given level of informativeness, these changes increase the share of high-ability students admitted by the high-ranked school and, therefore, increase the chances that the same school will rank high again. In this sense, both contribute to making the ranking more persistent.

One could also build persistence into rankings by basing the ranking on some longer window of schools' recent performance, e.g. let $p(\cdot)$ be the probability that the better school has higher performance and construct the ranking as follows: in period 0, the school with higher performance ranks high, and in each period $t > 0$, the low-ranked school in period $t-1$ becomes the high-ranked school in period t if and only if it had higher performance in periods $t-1$ and $t-2$. Such built-in inertia would increase the chances that the high-ranked school will rank high again, and would raise the equilibrium levels of mobility and informativeness.

To understand why this is the case, it is important to realise that this inertia increases the steady-state level of informativeness at any given level of mobility. In particular, as a result of this inertia, a school becomes less likely to change its rank from one period to the next, irrespective of its quality. Without any inertia, the better school is more likely to maintain a high rank over time than the worse school, due to its superior quality. Therefore, introducing inertia must make the better school even more likely to maintain a high rank relative to the worse school, which increases the steady-state level of informativeness at any given level of mobility.[25] As a consequence, the equilibrium levels of mobility and informativeness increase.[26]

5 Welfare

This section analyses how a utilitarian social planner would optimally assign students to schools if he were subject to the same informational constraints as students, that is, if he could condition this assignment of students only on the identity of the most recent winning school. The planner may face a trade-off between maximising welfare in a given period and maximising future welfare.

If all students derive the same benefit V from attending the better school, the planner's objective amounts to minimising expenditure on transport costs. Hence, he optimally assigns students to their respective local schools, independent of the ranking. Suppose instead that a high-ability

[25] This reason is similar to the one for why steady-state informativeness increases in mobility. See Subsection 3.2

[26] To maximise equilibrium informativeness, such inertias should be made as strong as possible. However, if school qualities change exogenously over time, a trade-off arises: with unchanged qualities, a longer window of past performance provides a more informative ranking, but with changing qualities, recent past performance is more likely than earlier past performance to have been generated by the current school qualities. By contrast, if the ranking is based on last period's performance, a decrease in transport costs and or a rise in the share of non-local high-ability applicants admitted by an oversubscribed school always raises equilibrium informativeness (see Subsection 7.2).

student values attending the better school at $V^H = V$ and a low-ability student at $V^L = \delta \cdot V$, where $\delta \in (0,\infty)$.[27] From the planner's point of view, school quality and intake ability are complements if $\delta \leq 1$ and substitutes if $\delta \geq 1$. Let $h \in [0,1]$ be the share of high-ability students assigned to the most recent winner and let $Pr\left(\omega^i|W^i\right)$ be the posterior belief that the most recent winner is the better school. Then expected utilitarian welfare is given by

$$\begin{aligned} U\left(h, Pr\left(\omega^i|W^i\right), \delta\right) =& V\left\{Pr\left(\omega^i|W^i\right)\left[h+(1-h)\delta\right] + \left[1-Pr\left(\omega^i|W^i\right)\right]\left[1-h+h\delta\right]\right\} \\ & -2\int_0^{F^{-1}\left(\left|h-\frac{1}{2}\right|\right)} c dF(c),^{28} \end{aligned} \tag{16}$$

where F^{-1} denotes the inverse of F.

A forward-looking planner maximises expected welfare in steady state, that is

$$\max_{h\in[0,1]} U\left(h, \pi(h), \delta\right) \tag{17}$$

where

$$\pi(h) = \frac{p(1-h)}{1-p(h)+p(1-h)}.^{29} \tag{18}$$

In particular, he not only takes into account how his assignment influences expected welfare at any given posterior belief about school quality, but also how his assignment influences the stationary distribution of rankings and, hence, posterior beliefs. This reflects 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. To illustrate how the concern for future welfare affects the optimal choice, consider the benchmark of *myopic* planner who maximises expected welfare taking as given the posterior beliefs at the forward-looking planner's optimal choice: for $h_F \in \text{argmax}_h U\left(h, \pi(h), \delta\right)$, the myopic planner solves

$$\max_{h\in[0,1]} U\left(h, \pi(h_F), \delta\right). \tag{19}$$

Proposition 3 (Welfare) *Suppose the probability of winning strictly increases with the share of high-ability students, i.e.* $\frac{\partial}{\partial h}p(h) > 0$. *Consider any* $h_F \in \text{argmax}_h U\left(h, \pi(h), \delta\right)$.
1) For $\delta \in (0,1]$, *i.e. the case of complements,* $\text{argmax}_h U\left(h, \pi(h_F), \delta\right) \subseteq \left[\frac{1}{2}, 1\right]$ *and for* $\delta \in [1,\infty)$, *i.e. the case of substitutes,* $\text{argmax}_h U\left(h, \pi(h_F), \delta\right) \subseteq \left[0, \frac{1}{2}\right]$.

[27] Note that the preceding analysis is unaffected by these assumptions, since only the high-ability students' application strategies affect the equilibrium allocation of students.

[28] The last term of equation (16) already incorporates that the expenditure on transport costs is minimised for a given h by assigning those students to non-local schools who incur the lowest transport costs within each ability group.

[29] Posterior beliefs in steady state satisfy (18) by the proof of Lemma 2.

2) For $\delta \in (0,\infty)$,

$$h_F \geq \sup\left\{\underset{h}{\operatorname{argmax}}\, U\left(h, \pi\left(h_F\right), \delta\right)\right\}, \tag{20}$$

with strict inequality if either i) $\delta \in (0,1)$ and $h_F \in \left(\frac{1}{2},1\right)$ or ii) $\delta \in (1,\infty)$ and $h_F \notin \left\{0,\frac{1}{2}\right\}$.

In the case of substitutes, the myopic planner allocates the majority of high-ability students to the most recent loser. The forward-looking planner allocates fewer high-ability students to the most recent loser than the myopic planner. This is because he is willing to forego some of the expected benefit derived from the optimal allocation at any given beliefs about school qualities in exchange for obtaining more accurate beliefs about school qualities in steady state.

In the case of complements, the myopic social planner allocates the majority of high-ability students to the most recent winner. By contrast, the forward-looking planner allocates an even larger share of high-ability students to the most recent winner. This is because he takes into account that he will thereby raise the likelihood of identifying the better school, which in turn allows him to exploit complementarities between student ability and school quality even more.

6 Extension: Quasi-Market Reforms

This extension will analyse how overall school quality and students' access to good schools are affected by quasi-market reforms. These reforms intend to mimic market forces by linking a school's funding to the 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, which affects the qualities of schools.

In the baseline model, school qualities were assumed to be exogenously determined. 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.[30] This extended framework allows me to study a steady-state

[30] 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.

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 baseline model is amended as follows. Each school's quality is either good or bad and drawn independently at the start of period 0. In addition, a given school's quality is redrawn with probability $\gamma \in [0,1]$ in period t before the ranking of schools W_t is realised if and only if this school was undersubscribed in period t. Any draw of school quality is equally likely to be good or bad.[31] Neither the event of a quality draw nor its realisation is observed. A *steady-state equilibrium* is a steady state such that i) the time-invariant mobility $m_t = m^*$ is optimal given beliefs and ii) beliefs are derived from the stationary joint distribution of the pair of school qualities and the school ranking using Bayes' rule.[32] Optimal mobility depends on informativeness, which, as before, is given by the difference between the posterior belief that the most recent winner is a strictly better school and the posterior belief that the most recent winner is a strictly worse school. All proofs can be found in the Online Appendix.

To study the effect of introducing quasi-market reforms, I 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. $\gamma > 0$.[33]

Proposition 4 (Quasi-Market Reforms) *If $\gamma > 0$ rather than $\gamma = 0$, then*

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

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

[31] 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.

[32] Since I focus on a symmetric steady-state equilibrium, it is irrelevant how the relative performance of schools of the same quality is determined.

[33] 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.

also unaffected.[34] Second, the chances that schools are of different quality at any time is unaffected by whether $\gamma = 0$ or $\gamma > 0$ because a new quality draw is equally likely to be good or bad.

If an undersubscribed school's quality is redrawn, the overall quality of schools increases in equilibrium. To remain in steady state, the proportion of good schools closing down must equal 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 5 (Quasi-Market Reforms - Comparative Statics) *If $\gamma > 0$, then*

1. *an increase in the share s 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 levels of mobility and informativeness, and*

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

- *increases the share of high-ability students at a good school.*

Given $\gamma > 0$, (1.) or (2.) decreases the share of low-ability students at a good school if at all $h \in \left[\frac{1}{2}, 1\right]$,

$$(2-2h)\frac{\partial}{\partial h}p(h) \leq 1 - p(h) + p(1-h). \tag{21}$$

These exogenous changes have two effects. First, they raise the share of high-ability students at the relatively better school (sorting effect), as in the case of exogenous school qualities. Second, they 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. However, the effect on the share of low-ability students at good schools is ambiguous. It is negative if the impact of student ability on performance is sufficiently weak, as this causes the sorting effect to dominate.

Therefore, a policymaker may face a trade-off between equity and efficiency. While lowering transport costs or increasing how selective the admission rule is raises school quality overall, these benefits may accrue only to high-ability students.[35]

[34] 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.

[35] There may be other reasons outside of this model for why an admission rule that selects a larger share of applicants based on ability is undesirable, e.g. one may worry that if oversubscribed schools get to select a larger share of their intake based on ability, these schools will invest more effort into screening applicants and will divert effort away from other tasks that would improve their students' education.

7 Robustness Checks

This section shows to what extent my findings are robust to different assumptions about how rankings are generated, how persistent school qualities are and how the admission rule selects between applicants.

7.1 Rankings

My paper shows that equilibrium informativeness increases as transport costs are lowered or as the share of non-local high-ability applicants admitted by an oversubscribed school increases. The argument is based on the assumption that the worse school is never guaranteed to win even if it admitted all high-ability students, as implied by equation (2). Suppose I instead allowed for the possibility that there exists some critical level $\overline{h}$ for the share of high-ability students such that, if any school $i \in \{0,1\}$ enrols h students, where $h > \overline{h}$, then school i is guaranteed to win.[36] 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 a self-perpetuating cycle arises in which the same school continues to rank high: once it has attracted sufficiently many high-ability students to rank high it will then be able to attract these high-ability students again and again, irrespective of its quality.[37] Given these assumptions, it is no longer true that the level of informativeness necessarily increases in the share s of non-local high-ability applicants admitted by an oversubscribed school. However, the higher the share s of non-local high-ability applicants admitted, 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 equal across schools.

Furthermore, I assume students observe only the most recent ranking. The insights of this paper still apply if students instead 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. This is the feature on which the proofs for equilibrium existence and comparative statics results are based. In Section B.2 of the Online Appendix, I analyse the situation when, in each period t, students observe the two most recent rankings. I solve for equilibrium and show that Theorem 1 still applies.

[36] The model of this paper can be thought of as capturing situations in which there is sufficiently high differentiation between schools such that intakes will not become too unequal 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.

[37] In the terminology of the social learning literature, a herd will arise.

7.2 School Qualities

In the baseline model, school qualities are assumed to be fixed but, in reality, they might change over time. The results of my model are qualitatively unaffected when school qualities change exogenously. In particular, suppose that schools' relative quality, i.e. the state of the world, changes with some fixed probability $g \in \left[0, \frac{1}{2}\right)$ each period and that such changes are unobserved by students. Then the steady-state level of informativeness still increases with mobility and, hence, Theorem 1 still holds. The proof can be found in Section B.3 of the Online Appendix.[38]

7.3 Admission Rule

Even in situations in which schools' admission rules do not select among applicants based on their ability, the insights developed in this paper can be useful. Informativeness increases over time if and only if, in any period t, the share of high-ability students enrolling at the school which ranked high in period t-1 increases in their level of informativeness, I_t. In the model studied, this requires that the admission rule selects a strictly positive share of applicants based on their ability, i.e. $s > 0$. However, there are alternative sets of assumptions under which this condition is met even if the admission school selects among applicants at random.

For example, suppose that high-ability students derive a larger benefit from attending a better school, i.e. $V^H > V^L > 0$, and that the distribution for transport costs is Uniform.[39] Then high-ability students are willing to incur higher transport costs than low-ability students to attend the most recent winning school at any given level of informativeness. Hence, the applicant pool of the most recent winning school contains more high-ability than low-ability students and the share of high-ability students is greater the higher the level of informativeness.[40] Therefore, even if places at this school were allocated to applicants at random independent of their ability, the school would admit a larger share of high-ability students than the other school and the share of high-ability students admitted would increase with the level of informativeness. More detail can be found in Section B.4 in the Online Appendix.

In addition, even if an oversubscribed school had to admit those students who live nearest to the school irrespective of their ability, some of the effects captured in the analysis may still be present. Students would consider the quality of schools when choosing where to live (Tiebout choice). If

[38] If a high-ranked school were more likely to maintain or improve its quality than a low-ranked school, e.g. because it can attract more capable teachers, then a high rank would also serve as a predictor for better relative quality in the future. Students would then find it even more attractive to attend a high-ranked school at any given belief about schools' past qualities. As a consequence, rankings would become even more informative about schools' past qualities.

[39] Empirical evidence shows that students of higher socio-economic status are more likely to seek out better-performing schools (e.g. Hastings et al. [2005], Allen et al. [2014]) and higher socio-economic status is correlated with more parental investment in education and higher expected attainment (e.g. Machin et al. [2013], Richards et al. [2016], Shaw et al. [2017]).

[40] I assume V^H sufficiently low such that not all high-ability students apply to the most recent winner.

high-ability students 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 than the other school.

8 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 rule leads oversubscribed school to prioritise more able applicants. I also find that such a ranking is more informative if students perceive schools to be less horizontally differentiated. 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 rules. 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.

A Appendix

A.1 Lemma 2 (Steady-State Informativeness given Mobility)

The sequence of signal realisations $\{W_t\}_{t>0}$ follows a time-homogeneous Markov process. If $p(1) < 1$, the process is irreducible and, hence, has a unique stationary distribution characterised by a constant probability π that the better school wins, which solves

$$\pi = p(h(\widehat{m},s))\,\pi + p(1-h(\widehat{m},s))\,(1-\pi)\,. \tag{22}$$

Hence,

$$\pi(h(\widehat{m},s)) = \frac{p(1-h(\widehat{m},s))}{1-p(h(\widehat{m},s))+p(1-h(\widehat{m},s))}. \tag{23}$$

By Bayes' rule and by the symmetry of the setting: for $i, j \in \{0,1\}$ and $j \neq i$,

$$Pr\left(\omega^i|W^i\right) = \frac{\pi\left(h\left(\widehat{m},s\right)\right)Pr\left(\omega^i\right)}{\pi\left(h\left(\widehat{m},s\right)\right)Pr\left(\omega^i\right) + \left(1-\pi\left(h\left(\widehat{m},s\right)\right)\right)Pr\left(\omega^j\right)} = \pi\left(h\left(\widehat{m},s\right)\right). \tag{24}$$

Hence, $\pi(h) > \frac{1}{2}$ for all h given (2). By Definition 2, $I(\widehat{m}) = 2\pi\left(h\left(\widehat{m},s\right)\right) - 1$. Hence, $I(\widehat{m}) > 0$. In addition,

$$\frac{dI}{d\widehat{m}} \geq 0 \Leftrightarrow \frac{d\pi}{dh}\frac{dh}{d\widehat{m}} \geq 0. \tag{25}$$

$\frac{d\pi}{dh}\frac{dh}{d\widehat{m}} \geq 0$ since $\frac{dh}{d\widehat{m}} = \frac{s}{2} \geq 0$ and

$$\frac{d\pi}{dh} \geq 0 \Leftrightarrow \left[1 - p\left(h\left(\widehat{m},s\right)\right) + p\left(1-h\left(\widehat{m},s\right)\right)\right]\frac{dp\left(h\left(\widehat{m},s\right)\right)}{dh} \geq 0, \tag{26}$$

where $p(\cdot) \in [0,1]$ and $\frac{dp(h)}{dh} \geq 0$ by (1).

A.2 Proposition 2 (Convergence)

The sequence of levels of informativeness in non-steady-state equilibrium satisfies (13) because students in period 0 do not observe a ranking and, hence, their posterior beliefs equal their prior beliefs. In addition, the equilibrium sequence satisfies (14), since for $i, j \in \{0,1\}$ and $j \neq i$ by Bayes' rule it follows that

$$I_t \equiv Pr\left(\omega^i|W^i_{t-1}\right) - Pr\left(\omega^j|W^i_{t-1}\right) = 2Pr\left(W^i_{t-1}|\omega^i\right) - 1, \tag{27}$$

and since the stationary distribution must satisfy

$$Pr\left(W^i_{t-1}|\omega^i\right) = p\left(h\left(m_{t-1},s\right)\right)Pr\left(W^i_{t-2}|\omega^i\right) + p\left(1-h\left(m_{t-1},s\right)\right)\left(1 - Pr\left(W^i_{t-2}|\omega^i\right)\right).$$

and since (6) holds.

The sequence $\{I_t\}_{t\geq 0}$ is increasing by the following induction argument. Define

$$\begin{aligned} Z\left(I_{t-1}\right) = &\left[p\left(h\left(m_{t-1},s\right)\right) - p\left(1-h\left(m_{t-1},s\right)\right)\right]I_{t-1} \\ &+ p\left(h\left(m_{t-1},s\right)\right) - \left(1 - p\left(1-h\left(m_{t-1},s\right)\right)\right), \end{aligned} \tag{28}$$

where $m_{t-1} = F\left(V \cdot I_{t-1}\right)$. Then

$$\frac{d}{dI_{t-1}}Z\left(I_{t-1}\right) = \left[p\left(h\left(m_{t-1},s\right)\right) - p\left(1-h\left(m_{t-1},s\right)\right)\right] + 2\left[\frac{d}{dI_{t-1}}p\left(h\left(m_{t-1},s\right)\right)\right]. \tag{29}$$

Since $m_{t-1} \geq 0$, we have $h(m_{t-1}, s) \geq \frac{1}{2}$. Given $h(m_{t-1}, s) \geq \frac{1}{2}$ and (1):

$$p(h(m_{t-1}, s)) - p(1 - h(m_{t-1}, s)) \geq 0. \tag{30}$$

In addition,

$$\frac{d}{dI_{t-1}} p(h(m_{t-1}, s)) = \frac{dp(h)}{dh} \frac{dh(m_{t-1}, s)}{dm_{t-1}} \frac{dm_{t-1}}{dI_{t-1}} \geq 0, \tag{31}$$

since $F(\cdot)$ is positive and increasing, $V > 0$, $I_{t-1} \geq 0$, $\frac{dh(m_{t-1},s)}{dm_{t-1}} = \frac{s}{2} \geq 0$, and (1) holds. Hence, it follows that $Z(I_{t-1})$ increases in I_{t-1}. In addition, by (2),

$$I_1 = Z(I_0) = p\left(\frac{1}{2}\right) - \left(1 - p\left(\frac{1}{2}\right)\right) \geq I_0 = 0, \tag{32}$$

and due to $Z(I_{t-1})$ being increasing in I_{t-1}, it follows that

$$I_t = Z(I_{t-1}) \geq Z(I_{t-2}) = I_{t-1}. \tag{33}$$

Since $\{I_t\}_{t \geq 0}$ is increasing, it converges to its least upper bound. This least upper bound is given by the smallest fixed point I^{1*} for the following reason: 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 (6), $V > 0$ and $F(\cdot)$ being 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.3 Theorem 1 (Comparative Statics)

Define

$$\Gamma(I, s, F(\cdot), p(\cdot), V) = \frac{p(h(F(V \cdot I), s)) - (1 - p(1 - h(F(V \cdot I), s)))}{1 - p(h(F(V \cdot I), s)) + p(1 - h(F(V \cdot I), s))}. \tag{34}$$

For any $I \in [0,1]$, $\frac{d\Gamma}{dI} = \frac{d\Gamma}{dF}\frac{dF}{dI} \geq 0$, since $\frac{d\Gamma}{dF} \geq 0$ holds given (9), and since $\frac{dF}{dI} \geq 0$ holds given $I \geq 0$, $V > 0$ and $F(\cdot)$ being increasing.

1. For any $I \in [0,1]$, $\frac{d\Gamma}{ds} = \frac{d\Gamma}{dh}\frac{dh}{ds} \geq 0$, since $\frac{d\Gamma}{dh} \geq 0 \Leftrightarrow \frac{d\pi}{dh} \geq 0$ and $\frac{d\pi}{dh} \geq 0$ given (26), and since $\frac{dh}{ds} = \frac{F(V \cdot I)}{2} \geq 0$. After Corollary 1, (p. 446), in Milgrom and Roberts [1994], this implies that the smallest fixed point of $\Gamma(I, s, F(\cdot), p(\cdot), V)$, denoted by $I^{1*}(s, F(\cdot), p(\cdot), V)$, is increasing in s. Since $m^{1*} = F\left(V \cdot I^{1*}\right)$, $V > 0$ and $F(\cdot)$ is increasing, m^{1*} also increases in s.

2. For any $I \in [0,1]$, and any $F(\cdot)$ and $\widetilde{F}(\cdot)$, such that $F(\cdot)$ first-order stochastically dominates

$\widetilde{F}(\cdot)$, it holds that $F(V\cdot I)\leq\widetilde{F}(V\cdot I)$. By (9), this implies that for any $I\in[0,1]$,

$$\Gamma(I,s,F(\cdot),p(\cdot),V)\leq\Gamma\left(I,s,\widetilde{F}(\cdot),p(\cdot),V\right). \tag{35}$$

After Corollary 1, (p. 446), in Milgrom and Roberts [1994], this implies that the smallest fixed point satisfies $I^{1*}(s,F(\cdot),p(\cdot),V)\leq I^{1*}\left(s,\widetilde{F}(\cdot),p(\cdot),V\right)$. In addition, since $m^{1*}=F\left(V\cdot I^{1*}\right)$, $V>0$ and $F(\cdot)$ is increasing: $m^{1*}(s,F(\cdot),p(\cdot),V)\leq m^{1*}\left(s,\widetilde{F}(\cdot),p(\cdot),V\right)$.

A.4 Proposition 3 (Welfare)

By the proof of Lemma 2, $\frac{\partial p(h)}{\partial h}>0\Rightarrow\frac{\partial\pi(h)}{\partial h}>0$. Consider each range of δ in turn.

Suppose $\delta=1$. Then

$$U(h,\pi,\delta)=V-\left[2\int_0^{F^{-1}\left(|h-\frac{1}{2}|\right)}cdF(c)\right]. \tag{36}$$

The second term of (36) is minimised at $h=\frac{1}{2}$. Hence, $\text{argmax}_h U(h,\pi,\delta)=\left\{\frac{1}{2}\right\}$ independent of π and $h_F=\sup\left\{\text{argmax}_h U\left(h,\pi\left(\frac{1}{2}\right),\delta\right)\right\}=\frac{1}{2}$.

Suppose $\delta<1$. 1) Fix any $h_F\in\text{argmax}_h U(h,\pi(h),\delta)$. Then $\text{argmax}_h U(h,\pi(h_F),\delta)\subseteq\left[\frac{1}{2},1\right]$: Suppose $h'_M\in\text{argmax}_h U(h,\pi(h_F),\delta)$ and $h'_M<\frac{1}{2}$. For $h\neq\frac{1}{2}$,

$$\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)=V(2\pi(h_F)-1)(1-\delta)-\frac{\partial}{\partial h}\left[2\int_0^{F^{-1}\left(|h-\frac{1}{2}|\right)}cdF(c)\right]. \tag{37}$$

Then $U\left(\frac{1}{2},\pi(h_F),\delta\right)>U(h'_M,\pi(h_F),\delta)$ since i) the expenditure on transport costs is 0 at $h=\frac{1}{2}$ and positive at h'_M and ii) the expected benefit strictly increases in h for $h\leq\frac{1}{2}$ since the first term of (37) is strictly positive given $\delta<1$ and $\pi(h_F)>\frac{1}{2}$ by the proof of Lemma 2. This is a contradiction to $h'_M\in\text{argmax}_h U(h,\pi(h_F),\delta)$.

2a) If $h_F\in\left(\frac{1}{2},1\right)$, then $h_F>\sup\{\text{argmax}_h U(h,\pi(h_F),\delta)\}$: for $h\neq\frac{1}{2}$,

$$\begin{aligned}\frac{\partial}{\partial h}U(h,\pi(h),\delta)=&V(2\pi(h)-1)(1-\delta)+V(2h-1)(1-\delta)\frac{\partial}{\partial h}\pi(h)\\&-\frac{\partial}{\partial h}\left[2\int_0^{F^{-1}\left(|h-\frac{1}{2}|\right)}cdF(c)\right].\end{aligned} \tag{38}$$

Since h_F is an interior optimum, $\frac{\partial}{\partial h}U(h,\pi(h),\delta)|_{h=h_F}=0$. In addition, given $h_F\in\left(\frac{1}{2},1\right)$, $\frac{\partial}{\partial h}\pi(h)>0$ and $\delta<1$,

$$\left.\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)\right|_{h=h_F}-\left.\frac{\partial}{\partial h}U(h,\pi(h),\delta)\right|_{h=h_F}=-V(2h_F-1)(1-\delta)\frac{\partial}{\partial h}\pi(h_F)<0. \tag{39}$$

Therefore, $\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)|_{h=h_F}<0$. Further, for $h\neq\frac{1}{2}$ and any h_F,

$$\frac{\partial^2}{\partial h^2}U(h,\pi(h_F),\delta)=-\frac{\partial^2}{\partial h^2}\left[2\int_0^{F^{-1}\left(|h-\frac{1}{2}|\right)}cdF(c)\right]\leq 0.^{41} \tag{40}$$

Hence, if $h'_M\in\text{argmax}_h U(h,\pi(h_F),\delta)$ then either i) $\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)|_{h=h'_M}=0$ and $h'_M<h_F$, or ii) $\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)\leq 0$ for all $h\in\left(\frac{1}{2},1\right]$ and $h'_M=\frac{1}{2}$ given $h'_M\in\left[\frac{1}{2},1\right]$.
2b) If $h_F=\frac{1}{2}$ then $h_F=\sup\{\text{argmax}_h U(h,\pi(h_F),\delta)\}$: If $h_F=\frac{1}{2}$ then $\lim_{h\to\frac{1}{2}^+}\frac{\partial}{\partial h}U(h,\pi(h),\delta)\leq 0$ by continuity of $U(h,\pi(h),\delta)$. Since (39) holds, $\lim_{h\to\frac{1}{2}^+}\frac{\partial}{\partial h}U\left(h,\pi\left(\frac{1}{2}\right),\delta\right)\leq 0$.
Since $\frac{\partial^2}{\partial h^2}U\left(h,\pi\left(\frac{1}{2}\right),\delta\right)\leq 0$ for all $h\in\left(\frac{1}{2},1\right]$, this implies $\frac{\partial}{\partial h}U\left(h,\pi\left(\frac{1}{2}\right),\delta\right)\leq 0$ for all $h\in\left(\frac{1}{2},1\right]$. Hence, $\sup\left\{\text{argmax}_h U\left(h,\pi\left(\frac{1}{2}\right),\delta\right)\right\}=\frac{1}{2}$.
2c) $\text{argmax}_h U(h,\pi(h),\delta)\subseteq\left[\frac{1}{2},1\right]$: Suppose $h'_F\in\text{argmax}_h U(h,\pi(h),\delta)$ and $h'_F<\frac{1}{2}$. Then $U\left(\frac{1}{2},\pi\left(\frac{1}{2}\right),\delta\right)>U(h'_F,\pi(h'_F),\delta)$ since i) the expenditure on transport cost is 0 at $h=\frac{1}{2}$ and positive at h'_F and ii) the expected benefit strictly increases in h for $h<\frac{1}{2}$ since the first two terms of (38) are strictly positive given $\delta<1$ and given that, for all h, $\frac{\partial}{\partial h}\pi(h)>0$ and $\pi(h)>\frac{1}{2}$ by the proof of Lemma 2. This is a contradiction to $h'_F\in\text{argmax}_h U(h,\pi(h),\delta)$. Given 2a) and 2b), this implies (20).

Suppose $\delta>1$. 1) Fix any $h_F\in\text{argmax}_h U(h,\pi(h),\delta)$. Then $\text{argmax}_h U(h,\pi(h_F),\delta)\subseteq\left[0,\frac{1}{2}\right]$: Suppose $h'_M\in\text{argmax}_h U(h,\pi(h_F),\delta)$ and $h'_M>\frac{1}{2}$. Then $U\left(\frac{1}{2},\pi(h_F),\delta\right)>U(h'_M,\pi(h_F),\delta)$ since i) the expenditure on transport costs is 0 at $h=\frac{1}{2}$ and positive at h'_M and ii) the expected benefit strictly decreases in h for $h\in\left[0,\frac{1}{2}\right)$ since the first term of (37) is negative given $\delta>1$ and $\pi(h_F)>\frac{1}{2}$ by the proof of Lemma 2. This is a contradiction to $h'_M\in\text{argmax}_h U(h,\pi(h_F),\delta)$.
2a) From 1), it follows that if $h_F\geq\frac{1}{2}$, then $h_F\geq\sup\{\text{argmax}_h U(h,\pi(h_F),\delta)\}$ and if $h_F>\frac{1}{2}$, then $h_F>\sup\{\text{argmax}_h U(h,\pi(h_F),\delta)\}$.
2b) If $h_F\in\left(0,\frac{1}{2}\right)$, then $h_F>\sup\{\text{argmax}_h U(h,\pi(h_F),\delta)\}$: since h_F is an interior optimum, $\frac{\partial}{\partial h}U(h,\pi(h),\delta)|_{h=h_F}=0$. Hence, $\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)|_{h=h_F}<0$ since (39) holds given $\frac{\partial}{\partial h}\pi(h)>0$ and $\delta<1$. In addition, (40) holds. Hence, if $h'_M\in\text{argmax}_h U(h,\pi(h_F),\delta)$ then either i) $\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)|_{h=h'_M}=0$ and $h'_M<h_F$, or ii) $\frac{\partial}{\partial h}U(h,\pi(h_F),\delta)\leq 0$ for all $h\in\left[0,\frac{1}{2}\right)$ and $h'_M=0$.
2c) If $h_F=0$, $h_F=\sup\{\text{argmax}_h U(h,\pi(h_F),\delta)\}$: if $h_F=0$ then $\frac{\partial}{\partial h}U(h,\pi(h),\delta)|_{h=0}\leq 0$. Since (39) holds given $\frac{\partial}{\partial h}\pi(h)>0$ and $\delta<1$, $\frac{\partial}{\partial h}U(h,\pi(0),\delta)|_{h=0}\leq 0$. In addition, since (40) holds, this implies $\frac{\partial}{\partial h}U(h,\pi(0),\delta)\leq 0$ for all $h\in\left[0,\frac{1}{2}\right)$. Hence, by 1), $\sup\{\text{argmax}_h U(h,\pi(0),\delta)\}=0$. Together with 2a and 2b, this implies (20).

[41] The marginal benefit is constant in h. For $h\in\left[0,\frac{1}{2}\right)$, the cost decreases in h, and it decreases less the higher h because those who no longer travel then live closer to their non-local school, and for $h\in\left(\frac{1}{2},1\right]$ the cost increases in h and it increases more the higher h because those who now travel live further away from their non-local school.

B Online Appendix

B.1 Extension: Quasi-Market Reform

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

$$m^* = F\left(\frac{V}{2} \cdot \frac{p\left(h\left(m^*, s\right)\right) - \left(1 - p\left(1 - h\left(m^*, s\right)\right)\right)}{p\left(1 - h\left(m^*, s\right)\right) + 1 - p\left(h\left(m^*, s\right)\right)}\right) \tag{41}$$

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

$$I\left(m^*\right) = \frac{1}{2} \cdot \frac{p\left(h\left(m^*, s\right)\right) - \left(1 - p\left(1 - h\left(m^*, s\right)\right)\right)}{p\left(1 - h\left(m^*, s\right)\right) + 1 - p\left(h\left(m^*, s\right)\right)}. \tag{42}$$

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

Proof: A school's quality can be either good (G) or bad (B). I will first derive the stationary joint distribution of the pair of school qualities and of the ranking in steady state with mobility $\widehat{m}$. Since I focus on symmetric equilibria, I do not need to keep track of schools' identities. Let $\overline{Q^i}Q^j$ be the event that school i is the winning school and has quality Q^i and school j is the losing school and has quality Q^j where $Q^i, Q^j \in \{G, B\}$ and $i \in \{0, 1\}$. Therefore, I denote the stationary joint distribution by the vector ρ where

$$\rho \equiv \left(Pr\left(\overline{G}G\right), Pr\left(\overline{G}B\right), Pr\left(\overline{B}G\right), Pr\left(\overline{B}B\right)\right). \tag{43}$$

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

$$T = \begin{pmatrix} Pr\left(\overline{G}G_t|\overline{G}G_{t-1}\right) & Pr\left(\overline{G}B_t|\overline{G}G_{t-1}\right) & Pr\left(\overline{B}G_t|\overline{G}G_{t-1}\right) & Pr\left(\overline{B}B_t|\overline{G}G_{t-1}\right) \\ Pr\left(\overline{G}G_t|\overline{G}B_{t-1}\right) & Pr\left(\overline{G}B_t|\overline{G}B_{t-1}\right) & Pr\left(\overline{B}G_t|\overline{G}B_{t-1}\right) & Pr\left(\overline{B}B_t|\overline{G}B_{t-1}\right) \\ Pr\left(\overline{G}G_t|\overline{B}G_{t-1}\right) & Pr\left(\overline{G}B_t|\overline{B}G_{t-1}\right) & Pr\left(\overline{B}G_t|\overline{B}G_{t-1}\right) & Pr\left(\overline{B}B_t|\overline{B}G_{t-1}\right) \\ Pr\left(\overline{G}G_t|\overline{B}B_{t-1}\right) & Pr\left(\overline{G}B_t|\overline{B}B_{t-1}\right) & Pr\left(\overline{B}G_t|\overline{B}B_{t-1}\right) & Pr\left(\overline{B}B_t|\overline{B}B_{t-1}\right) \end{pmatrix}. \tag{44}$$

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 $h\left(\widehat{m}, s\right)$ high-ability students. Hence, the transition matrix T has the following entries:

$$\begin{aligned}
Pr\left(\overline{G}B_t|\overline{G}G_{t-1}\right) &= \frac{\gamma}{2}\left[p\left(h\left(\widehat{m},s\right)\right)\right] \\
Pr\left(\overline{B}G_t|\overline{G}G_{t-1}\right) &= \frac{\gamma}{2}\left[1-p\left(h\left(\widehat{m},s\right)\right)\right] \\
Pr\left(\overline{G}B_t|\overline{G}B_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]p\left(h\left(\widehat{m},s\right)\right) \\
Pr\left(\overline{B}G_t|\overline{G}B_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]\left[1-p\left(h\left(\widehat{m},s\right)\right)\right] \\
Pr\left(\overline{G}B_t|\overline{B}G_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]p\left(1-h\left(\widehat{m},s\right)\right) \\
Pr\left(\overline{B}G_t|\overline{B}G_{t-1}\right) &= \left[1-\frac{\gamma}{2}\right]\left(1-p\left(1-h\left(\widehat{m},s\right)\right)\right) \\
Pr\left(\overline{G}B_t|\overline{B}B_{t-1}\right) &= \frac{\gamma}{2}p\left(1-h\left(\widehat{m},s\right)\right) \\
Pr\left(\overline{B}G_t|\overline{B}B_{t-1}\right) &= \frac{\gamma}{2}\left[1-p\left(1-h\left(\widehat{m},s\right)\right)\right] \\
Pr\left(\overline{G}G_t|\overline{G}G_{t-1}\right) = Pr\left(\overline{B}B_t|\overline{B}B_{t-1}\right) &= 1-\frac{\gamma}{2} \\
Pr\left(\overline{G}G_t|\overline{G}B_{t-1}\right) = Pr\left(\overline{B}B_t|\overline{B}G_{t-1}\right) &= \frac{\gamma}{2}
\end{aligned}$$

and the remaining entries are equal to 0.

If $\gamma > 0$, 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 the following two equations

$$\rho = \rho T \tag{45}$$

$$Pr\left(\overline{G}G\right)+Pr\left(\overline{G}B\right)+Pr\left(\overline{B}G\right)+Pr\left(\overline{B}B\right) = 1. \tag{46}$$

Hence, the stationary distribution is defined by

$$Pr\left(\overline{G}G\right) = Pr\left(\overline{G}B\right) = \frac{p\left(1-h\left(\widehat{m},s\right)\right)}{2\left(1+p\left(1-h\left(\widehat{m},s\right)\right)-p\left(h\left(\widehat{m},s\right)\right)\right)} \tag{47}$$

$$Pr\left(\overline{B}G\right) = Pr\left(\overline{B}B\right) = \frac{1-p\left(h\left(\widehat{m},s\right)\right)}{2\left(1+p\left(1-h\left(\widehat{m},s\right)\right)-p\left(h\left(\widehat{m},s\right)\right)\right)}. \tag{48}$$

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

$$Pr\left(\overline{G}G\right) = Pr\left(\overline{B}B\right) = \frac{1}{4} \tag{49}$$

and $Pr\left(\overline{G}B\right)$ and $Pr\left(\overline{B}G\right)$ are given by (47) and (48) respectively.

An ordered pair of school qualities in period t is $\left(Q^0,Q^1\right)_t$ where the quality of school $i\in\{0,1\}$

is $Q^i \in \{G,B\}$. Informativeness in period t is defined as

$$\begin{aligned} I_t &\equiv Pr\left((G,B)_{t-1} | W^0_{t-1}\right) - Pr\left((B,G)_{t-1} | W^0_{t-1}\right) \\ &= Pr\left((B,G)_{t-1} | W^1_{t-1}\right) - Pr\left((G,B)_{t-1} | W^1_{t-1}\right). \end{aligned} \tag{50}$$

Students' optimal application strategy is again given by $m_t = F(V \cdot I_t)$, where informativeness is given by (50). By Bayes' rule and by the symmetry of the setting,

$$\begin{aligned} Pr\left((G,B)_{t-1} | W^0_{t-1}\right) &= Pr\left((B,G)_{t-1} | W^1_{t-1}\right) = Pr\left(\overline{G}B\right), \\ Pr\left((B,G)_{t-1} | W^0_{t-1}\right) &= Pr\left((G,B)_{t-1} | W^1_{t-1}\right) = Pr\left(\overline{B}G\right), \\ Pr\left((G,G)_{t-1} | W^0_{t-1}\right) &= Pr\left((G,G)_{t-1} | W^1_{t-1}\right) = Pr\left(\overline{G}G\right), \\ Pr\left((B,B)_{t-1} | W^0_{t-1}\right) &= Pr\left((B,B)_{t-1} | W^1_{t-1}\right) = Pr\left(\overline{B}B\right). \end{aligned}$$

Hence, for $\gamma \geq 0$, the steady-state equilibrium mobility solves (10) and is given by (41). Existence follows by Tarski's fixed point theorem for the same reasons given in the proof of Proposition 1.

B.1.1 Proposition 4 (Quasi-Market Reforms)

Part 1: Informativeness and mobility are unaffected by γ as shown in Lemma 3.
Part 2: The average fraction of good schools in equilibrium, denoted by Θ, is given by:

$$\Theta = \begin{cases} Pr\left(\overline{G}G\right) + \frac{1}{2} Pr\left(\overline{G}B\right) + \frac{1}{2} Pr\left(\overline{B}G\right) & \text{if } \gamma > 0 \\ \frac{1}{2} & \text{if } \gamma = 0 \end{cases} \tag{51}$$

Given (47)-(49), then by (2) and since $p(\cdot) \in [0,1]$, for any $m^* \in [0,1]$,

$$\begin{aligned} \Theta(\gamma > 0) - \Theta(\gamma = 0) &= \frac{p(1-h(m^*,s))}{2(1+p(1-h(m^*,s)) - p(h(m^*,s)))} - \frac{1}{4} \\ &= \frac{p(h(m^*,s)) - (1-p(1-h(m^*,s)))}{4(1+p(1-h(m^*,s)) - p(h(m^*,s)))} \geq 0. \end{aligned} \tag{52}$$

Part 3: By Part 1, mobility is unaffected by γ and, hence, the share of high-ability students at the most recent winner is independent of γ. In steady state, this implies that, conditional on schools differing in quality, the share of high-ability students at the better school is independent of γ. By Part 2, if $\gamma > 0$ rather than $\gamma = 0$, the average fraction of good schools increases and, hence, the fraction of students of each ability type at a good school increases.

B.1.2 Proposition 5 (Quasi-Market Reforms - Comparative Statics)

Throughout this proof, I take as given the stationary distribution given by (47) and (48).

Bullet Point 1: The equilibrium levels of mobility and informativeness 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 in s and with a negative shift in the sense of FOSD of F.

Bullet Point 2: 1. The fraction Θ of good schools is given by (51). It holds that

$$\begin{aligned}\frac{\partial}{\partial h}Pr\left(\overline{G}G\right) = \frac{\partial}{\partial h}Pr\left(\overline{G}B\right) = -\frac{\partial}{\partial h}Pr\left(\overline{B}G\right) = -\frac{\partial}{\partial h}Pr\left(\overline{B}B\right) \\ = \frac{\frac{\partial p}{\partial h}\left(p\left(h\left(m^{*},s\right)\right)-\left(1-p\left(1-h\left(m^{*},s\right)\right)\right)\right)}{2\left[1-p\left(h\left(m^{*},s\right)\right)+p\left(1-h\left(m^{*},s\right)\right)\right]^{2}} \geq 0,\end{aligned} \tag{53}$$

given (1) and (2). Given (53), $\frac{\partial \Theta}{\partial h} \geq 0$ and, hence,

$$\frac{d\Theta}{ds} = \frac{\partial \Theta}{\partial h}\left[\frac{\partial h}{\partial s} + \frac{\partial h}{\partial m^{*}}\frac{dm^{*}}{ds}\right] \geq 0, \tag{54}$$

since $\frac{\partial h}{\partial m^{*}} = \frac{s}{2} \geq 0$, $\frac{\partial h}{\partial s} = \frac{m^{*}}{2} \geq 0$ and $\frac{dm^{*}}{ds} \geq 0$ by Bullet Point 1.

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^{*}, and by Part 1 of Bullet Point 2,

$$\frac{\partial \Theta}{\partial m^{*}} = \frac{\partial \Theta}{\partial h}\frac{\partial h}{\partial m^{*}} \geq 0, \tag{55}$$

and, by Bullet Point 1, m^{*} weakly increases with a negative shift in the sense of FOSD of $F(\cdot)$.

Bullet point 3 and last statement: Denote the share of high-ability (low-ability) students enrolled at a good school in equilibrium by H_G (L_G). Then

$$H_G = Pr\left(\overline{G}G\right) + h\left(m^{*},s\right)Pr\left(\overline{G}B\right) + \left(1-h\left(m^{*},s\right)\right)Pr\left(\overline{B}G\right), \tag{56}$$

$$L_G = Pr\left(\overline{G}G\right) + \left(1-h\left(m^{*},s\right)\right)Pr\left(\overline{G}B\right) + h\left(m^{*},s\right)Pr\left(\overline{B}G\right). \tag{57}$$

1. H_G increases in s for the following reason. By (2),

$$Pr\left(\overline{G}B\right) - Pr\left(\overline{B}G\right) = \frac{1}{2}\frac{p\left(h\left(m^{*},s\right)\right)-\left(1-p\left(1-h\left(m^{*},s\right)\right)\right)}{1-p\left(h\left(m^{*},s\right)\right)+p\left(1-h\left(m^{*},s\right)\right)} \geq 0. \tag{58}$$

Given (53) and (58) and $h \geq 0$, it follows that

$$\frac{\partial H_G}{\partial h} = \frac{\partial}{\partial h}Pr\left(\overline{G}G\right)\left[2h\left(m^{*},s\right)\right] + Pr\left(\overline{G}B\right) - Pr\left(\overline{B}G\right) \geq 0. \tag{59}$$

Given (59) and since $\frac{\partial h}{\partial m^*}=\frac{s}{2}\geq 0$, $\frac{\partial h}{\partial s}=\frac{m^*}{2}\geq 0$ and $\frac{dm^*}{ds}\geq 0$ by Bullet Point 1, it holds that

$$\frac{dH_G}{ds}=\frac{\partial H_G}{\partial h}\left[\frac{\partial h}{\partial s}+\frac{\partial h}{\partial m^*}\frac{dm^*}{ds}\right]\geq 0. \tag{60}$$

L_G decreases in s if (21) holds for the following reason. Given (53),

$$\begin{aligned}\frac{\partial L_G}{\partial h}&=\left[\frac{\partial}{\partial h}Pr\left(\overline{G}G\right)[2-2h]-\left[Pr\left(\overline{G}B\right)-Pr\left(\overline{B}G\right)\right]\right]\\&=\left[p(h)-(1-p(1-h))\right]\frac{\frac{\partial p}{\partial h}[2-2h]-[1-p(h)+p(1-h)]}{2[1-p(h)+p(1-h)]^2}\\&=\frac{\partial p}{\partial h}[2-2h]+[-1+p(h)-p(1-h)].\end{aligned} \tag{61}$$

If (21) holds, then $\frac{\partial L_G}{\partial h}\leq 0$ for any $h\in\left[\frac{1}{2},1\right]$, and by Lemma 3, $h(m^*,s)\in\left[\frac{1}{2},1\right]$. Given $\frac{\partial L_G}{\partial h}\leq 0$ and since $\frac{\partial h}{\partial m^*}=\frac{s}{2}\geq 0$, $\frac{\partial h}{\partial s}=\frac{m^*}{2}\geq 0$ and $\frac{dm^*}{ds}\geq 0$ by Bullet Point 1, it holds that

$$\frac{dL_G}{ds}=\frac{\partial L_G}{\partial h}\left[\frac{\partial h}{\partial s}+\frac{\partial h}{\partial m^*}\frac{dm^*}{ds}\right]\leq 0. \tag{62}$$

2. H_G increases with a negative shift in the sense of FOSD of $F(\cdot)$ because i) such a shift affects H_G only through its effect on m^*, ii) m^* increases with such a shift by Bullet Point 1 and iii) (59) holds by Part 1 of Bullet Point 3. Hence,

$$\frac{\partial H_G}{\partial m^*}=\frac{\partial H_G}{\partial h}\frac{\partial h}{\partial m^*}\geq 0. \tag{63}$$

L_G decreases with a negative shift in the sense of FOSD of $F(\cdot)$ by the same reasoning, i.e. i) and ii) are as above and by Part 1 of Bullet Point 3, $\frac{\partial L_G}{\partial h}\leq 0$ if (21) holds.

B.2 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 $i,j\in\{0,1\}$. Given $\left(W^i_{t-1},W^j_{t-2}\right)$, denote by m_{ij} the share of non-local students who apply to school i and denote the level of informativeness by

$$I_{ij}\equiv Pr\left(\omega^i|W^i_{t-1},W^j_{t-2}\right)-Pr\left(\omega^j|W^i_{t-1},W^j_{t-2}\right). \tag{64}$$

Proposition 6 *1. A steady-state equilibrium vector of mobility $m^*=\left(m^*_{00},m^*_{01},m^*_{10},m^*_{11}\right)$ is characterised by*

$$m_{00}^{*}=m_{11}^{*}=F\left[V\cdot\frac{p\left(\frac{1}{2}\right)p\left(1-h\left(m_{00}^{*},s\right)\right)-\left[1-p\left(h\left(m_{00}^{*},s\right)\right)\right]\left[1-p\left(\frac{1}{2}\right)\right]}{\left[1-p\left(h\left(m_{00}^{*},s\right)\right)\right]\left[1-p\left(\frac{1}{2}\right)\right]+p\left(1-h\left(m_{00}^{*},s\right)\right)p\left(\frac{1}{2}\right)}\right], \tag{65}$$

and

$$m_{10}^{*}=m_{01}^{*}=0, \tag{66}$$

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

$$I_{00}\left(m^{*}\right)=I_{11}\left(m^{*}\right)=F\left[V\cdot\frac{p\left(\frac{1}{2}\right)p\left(1-h\left(m_{00}^{*},s\right)\right)-\left[1-p\left(h\left(m_{00}^{*},s\right)\right)\right]\left[1-p\left(\frac{1}{2}\right)\right]}{\left[1-p\left(h\left(m_{00}^{*},s\right)\right)\right]\left[1-p\left(\frac{1}{2}\right)\right]+p\left(1-h\left(m_{00}^{*},s\right)\right)p\left(\frac{1}{2}\right)}\right], \tag{67}$$

and

$$I_{01}\left(m^{*}\right)=I_{10}\left(m^{*}\right)=0. \tag{68}$$

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

2. The equilibrium vectors of mobility and informativeness increase with an increase in the share s of non-local high-ability applicants admitted by an oversubscribed school, or with a negative shift in the sense of FOSD of the distribution $F(c)$ of transport costs.

Proof: Part 1. Optimal mobility is given by $m_{ij}=F\left(V\cdot I_{ij}\right)$ for the same reasons as in the proof of Lemma 1. Next, I derive steady-state informativeness at time-invariant vector of mobility levels $\widehat{m}=\left(\widehat{m}_{00},\widehat{m}_{01},\widehat{m}_{10},\widehat{m}_{11}\right)$. Since I restrict attention to symmetric strategies, $\widehat{m}_{00}=\widehat{m}_{11}$ and $\widehat{m}_{01}=\widehat{m}_{10}$. Suppose the state is ω^{i}. The stationary distribution of $\left(W_{t-1},W_{t-2}\right)$ is defined by the vector ρ, where $i,j\in\{0,1\}$ and $j\neq i$ and

$$\rho\equiv\left(Pr\left(W_{t-1}^{i},W_{t-2}^{i}\right),Pr\left(W_{t-1}^{i},W_{t-2}^{j}\right),Pr\left(W_{t-1}^{j},W_{t-2}^{i}\right),Pr\left(W_{t-1}^{j},W_{t-2}^{j}\right)\right). \tag{69}$$

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

$$T=\begin{pmatrix} p\left(h\left(\widehat{m}_{00},s\right)\right) & 0 & 1-p\left(h\left(\widehat{m}_{00},s\right)\right) & 0\\ p\left(h\left(\widehat{m}_{01},s\right)\right) & 0 & 1-p\left(h\left(\widehat{m}_{01},s\right)\right) & 0\\ 0 & p\left(1-h\left(\widehat{m}_{01},s\right)\right) & 0 & 1-p\left(1-h\left(\widehat{m}_{01},s\right)\right)\\ 0 & p\left(1-h\left(\widehat{m}_{00},s\right)\right) & 0 & 1-p\left(1-h\left(\widehat{m}_{00},s\right)\right)\end{pmatrix}. \tag{70}$$

The Markov process of $\left(W_{t-1},W_{t-2}\right)$ has a unique stationary distribution defined by the row vector ρ that satisfies the following two equations,

$$\rho = \rho T, \tag{71}$$

$$Pr\left(W_{t-1}^{i}, W_{t-2}^{i}\right) + Pr\left(W_{t-1}^{i}, W_{t-2}^{j}\right) + Pr\left(W_{t-1}^{j}, W_{t-2}^{i}\right) + Pr\left(W_{t-1}^{j}, W_{t-2}^{j}\right) = 1. \tag{72}$$

By Bayes' rule and the symmetry of the setting, posterior beliefs in steady state satisfy for any $i, j \in \{0,1\}$ and $j \neq i$ are given by

$$\begin{aligned} Pr\left(\omega^{i}|W_{t-1}^{i}, W_{t-2}^{i}\right) =& \frac{Pr\left(W_{t-1}^{i}, W_{t-2}^{i}|\omega^{i}\right) Pr\left(\omega^{i}\right)}{Pr\left(W_{t-1}^{i}, W_{t-2}^{i}|\omega^{i}\right) Pr\left(\omega^{i}\right) + Pr\left(W_{t-1}^{j}, W_{t-2}^{j}|\omega^{i}\right) Pr\left(\omega^{i}\right)} \\ =& \frac{p\left(1-h\left(\widehat{m}_{00}, s\right)\right) p\left(h\left(\widehat{m}_{01}, s\right)\right)}{\left[1-p\left(h\left(\widehat{m}_{00}, s\right)\right)\right]\left[1-p\left(1-h\left(\widehat{m}_{01}, s\right)\right)\right] + p\left(1-h\left(\widehat{m}_{00}, s\right)\right) p\left(h\left(\widehat{m}_{01}, s\right)\right)}, \end{aligned} \tag{73}$$

$$Pr\left(\omega^{i}|W_{t-1}^{i}, W_{t-2}^{j}\right) = \frac{Pr\left(W_{t-1}^{i}, W_{t-2}^{j}|\omega^{i}\right) Pr\left(\omega^{i}\right)}{Pr\left(W_{t-1}^{i}, W_{t-2}^{j}|\omega^{i}\right) Pr\left(\omega^{i}\right) + Pr\left(W_{t-1}^{j}, W_{t-2}^{i}|\omega^{i}\right) Pr\left(\omega^{i}\right)} = \frac{1}{2}. \tag{74}$$

Solving for informativeness yields that for any $\widehat{m}$,

$$I_{01}\left(\widehat{m}\right) = I_{10}\left(\widehat{m}\right) = 2Pr\left(\omega^{i}|W_{t-1}^{i}, W_{t-2}^{j}\right) - 1 = 0, \tag{75}$$

and

$$\begin{aligned} I_{00}\left(\widehat{m}\right) = I_{11}\left(\widehat{m}\right) =& 2Pr\left(\omega^{i}|W_{t-1}^{i}, W_{t-2}^{i}\right) - 1 \\ =& \frac{p\left(h\left(\widehat{m}_{01}, s\right)\right) p\left(1-h\left(\widehat{m}_{00}, s\right)\right) - \left[1-p\left(h\left(\widehat{m}_{00}, s\right)\right)\right]\left[1-p\left(h\left(\widehat{m}_{01}, s\right)\right)\right]}{\left[1-p\left(h\left(\widehat{m}_{00}, s\right)\right)\right]\left[1-p\left(1-h\left(\widehat{m}_{01}, s\right)\right)\right] + p\left(1-h\left(\widehat{m}_{00}, s\right)\right) p\left(h\left(\widehat{m}_{01}, s\right)\right)}. \end{aligned} \tag{76}$$

Hence, $I_{11}\left(\widehat{m}\right) = I_{00}\left(\widehat{m}\right) \geq 0$, since the denominator of (76) is positive, i.e.

$$X \equiv \left[1-p\left(h\left(\widehat{m}_{00}, s\right)\right)\right]\left[1-p\left(1-h\left(\widehat{m}_{01}, s\right)\right)\right] + p\left(1-h\left(\widehat{m}_{00}, s\right)\right) p\left(h\left(\widehat{m}_{01}, s\right)\right) \geq 0, \tag{77}$$

given $p\left(\cdot\right) \in [0,1]$, and since the numerator of (76) is positive, i.e.

$$Y \equiv p\left(h\left(\widehat{m}_{01}, s\right)\right) p\left(1-h\left(\widehat{m}_{00}, s\right)\right) - \left[1-p\left(h\left(\widehat{m}_{00}, s\right)\right)\right]\left[1-p\left(h\left(\widehat{m}_{01}, s\right)\right)\right] \geq 0, \tag{78}$$

given $p\left(\cdot\right) \in [0,1]$ and since by (2):

$$p\left(1-h\left(\widehat{m}_{00}, s\right)\right) > 1-p\left(h\left(\widehat{m}_{00}, s\right)\right), \tag{79}$$

$$p\left(h\left(\widehat{m}_{01}, s\right)\right) > 1-p\left(1-h\left(\widehat{m}_{01}, s\right)\right). \tag{80}$$

Hence, if $\widehat{m}_{ij}$ increases for any $i,j \in \{0,1\}$, then $I_{01}(\widehat{m})$, and $I_{10}(\widehat{m})$ remain unchanged. If $\widehat{m}_{01}$ or $\widehat{m}_{10}$ increases, $I_{00}(\widehat{m})$ and $I_{11}(\widehat{m})$ increase:

$$\begin{aligned}\frac{\partial I_{00}}{\partial \widehat{m}_{01}} \geq 0 \Leftrightarrow & p'(h(\widehat{m}_{01},s))\left[1-p(h(\widehat{m}_{00},s))+p(1-h(\widehat{m}_{00},s))\right]X \\ & + p'(h(\widehat{m}_{01},s))\left[p(h(\widehat{m}_{00},s))-\left[1-p(1-h(\widehat{m}_{00},s))\right]\right]Y \geq 0, \end{aligned} \tag{81}$$

and $\frac{\partial I_{00}}{\partial \widehat{m}_{01}} \geq 0$ holds since $p(\cdot) \in [0,1]$ and given (77), (78), (1) and (2). In addition, if $\widehat{m}_{00}$ or $\widehat{m}_{11}$ increases, $I_{00}(\widehat{m})$ and $I_{11}(\widehat{m})$ increase:

$$\frac{\partial I_{00}}{\partial \widehat{m}_{00}} \geq 0 \Leftrightarrow p'(h(\widehat{m}_{00},s))X + p'(h(\widehat{m}_{00},s))\left[1-p(1-h(\widehat{m}_{01},s))-p(h(\widehat{m}_{01},s))\right]Y \geq 0, \tag{82}$$

and $\frac{\partial I_{00}}{\partial \widehat{m}_{00}} \geq 0$ holds since $p(\cdot) \in [0,1]$ and given (77), (78), (1) and (2) and since

$$X - Y = 2(1-p(h(\widehat{m}_{00},s)))(1-p(h(\widehat{m}_{01},s))) \geq 0. \tag{83}$$

Hence, the vector of optimal mobilities $\overline{m}$ increases in the vector of levels of informativeness $(I_{00}, I_{01}, I_{10}, I_{11})$, which increases in the vector of steady-state mobilities $\widehat{m}$. In addition, $m \in [0,1]^4$. Therefore, Tarski's fixed point theorem applies and the equilibrium levels of mobility and informativeness must satisfy (65)-(68).

Part 2. Define

$$\begin{aligned}&\Gamma(I,s,F(\cdot),p(\cdot),V) = \\ &F\left[V \cdot \frac{p\left(\frac{1}{2}\right)p(1-h(F(V\cdot I),s))-\left[1-p(h(F(V\cdot I),s))\right]\left[1-p\left(\frac{1}{2}\right)\right]}{\left[1-p(h(F(V\cdot I),s))\right]\left[1-p\left(\frac{1}{2}\right)\right]+p(1-h(F(V\cdot I),s))\,p\left(\frac{1}{2}\right)}\right],\end{aligned} \tag{84}$$

where $\frac{d\Gamma}{dI} = \frac{d\Gamma}{dF}\frac{dF}{dI} \geq 0$, since $\frac{d\Gamma}{dF} \geq 0$ holds given Part 1, and since $\frac{dF}{dI} \geq 0$ holds given $I \geq 0$, $V > 0$ and $F(\cdot)$ is increasing. The result follows by the same steps as in the proof of Theorem 1.

B.3 Exogenous Changes in School Qualities

Suppose the state of the world ω changes with fixed probability $g \in \left[0, \frac{1}{2}\right)$ in any period $t \geq 0$ and these changes are unobserved by students. I will show that Theorem 1 still holds. First, optimal mobility is given by $\overline{m} = F(V(1-2g)\cdot I)$ because with probability $1-g$ school qualities remain unchanged and the gain from enrolling at the high-ranked school is $V \cdot I$, and with probability g school qualities swap and the gain from enrolling at the high-ranked school is $-V \cdot I$. Second, the steady-state level of informativeness is positive and increasing in mobility (i.e. the properties described in Lemma 2 still hold). Fix the mobility level to be $\widehat{m}$. The stationary probability that the better school ranks higher, denoted by π, satisfies:

$$\begin{aligned}\pi = & \left[(1-g)\,p\left(h\left(\widehat{m},s\right)\right) + g\cdot p\left(1-h\left(\widehat{m},s\right)\right)\right]\pi \\ & + \left[g\cdot p\left(h\left(\widehat{m},s\right)\right) + (1-g)\cdot p\left(1-h\left(\widehat{m},s\right)\right)\right](1-\pi)\,.\end{aligned} \tag{85}$$

Hence,

$$\pi\left(h\left(\widehat{m},s\right)\right) = \frac{p\left(1-h\left(\widehat{m},s\right)\right) - g\left[p\left(1-h\left(\widehat{m},s\right)\right) - p\left(h\left(\widehat{m},s\right)\right)\right]}{1+p\left(1-h\left(\widehat{m},s\right)\right) - p\left(h\left(\widehat{m},s\right)\right) - 2g\left[p\left(1-h\left(\widehat{m},s\right)\right) - p\left(h\left(\widehat{m},s\right)\right)\right]}. \tag{86}$$

Then $I = 2\pi - 1$. Hence, $I \geq 0 \Leftrightarrow \pi \geq \frac{1}{2}$ and $\pi \geq \frac{1}{2}$ since $p\left(1-h\left(\widehat{m},s\right)\right) > 1 - p\left(h\left(\widehat{m},s\right)\right)$ by (2). In addition,

$$\frac{\partial}{\partial \widehat{m}}\pi\left(h\left(\widehat{m},s\right)\right) = \frac{\partial \pi}{\partial h}\frac{\partial h}{\partial \widehat{m}} \geq 0, \tag{87}$$

since $\frac{\partial h}{\partial \widehat{m}} = \frac{s}{2} \geq 0$ and

$$\frac{\partial \pi}{\partial h} \geq 0 \Leftrightarrow \frac{\partial p\left(h\left(\widehat{m},s\right)\right)}{\partial h}\left[1 - p\left(h\left(\widehat{m},s\right)\right)(1-2g) + p\left(1-h\left(\widehat{m},s\right)\right)(1-2g)\right] \geq 0, \tag{88}$$

given $\frac{\partial p(h(\widehat{m},s))}{\partial h} \geq 0$ by (1), $p(\cdot) \in [0,1]$ and $g \in \left[0, \frac{1}{2}\right)$. Hence, a fixed point always exists by the same reasoning as in the proof of Proposition 1. Define

$$\begin{aligned}&\Gamma\left(I, s, F\left(\cdot\right), p\left(\cdot\right), V, g\right) = \\ &2\frac{p\left(1-h\left(F\left(V\cdot I\right),s\right)\right) - g\left[p\left(1-h\left(F\left(V\cdot I\right),s\right)\right) - p\left(h\left(F\left(V\cdot I\right),s\right)\right)\right]}{1+p\left(1-h\left(F\left(V\cdot I\right),s\right)\right) - p\left(h\left(F\left(V\cdot I\right),s\right)\right) - 2g\left[p\left(1-h\left(F\left(V\cdot I\right),s\right)\right) - p\left(h\left(F\left(V\cdot I\right),s\right)\right)\right]} - 1\end{aligned} \tag{89}$$

where $\frac{d\Gamma}{dI} = \frac{d\Gamma}{dF}\frac{dF}{dI} \geq 0$, since $\frac{d\Gamma}{dF} \geq 0$ holds given (88), and since $\frac{dF}{dI} \geq 0$ holds given $I \geq 0$, $V > 0$ and $F\left(\cdot\right)$ is increasing. The result follows by the same steps as in the proof of Theorem 1.

B.4 Admission not based on Ability

Suppose a student of ability $\alpha \in \{H, L\}$ derives benefit V^{α} 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^{α} denote the share of students of ability α who apply to their non-local winning school. Suppose F is Uniform on $[0,1]$, such that $F\left(c\right) = c$.

Proposition 7 *A steady-state equilibrium level of informativeness is given by*

$$I^* = \frac{p\left(h\left(\left(V^H \cdot I^*, V^L \cdot I^*\right)\right)\right) - \left(1 - p\left(1 - h\left(\left(V^H \cdot I^*, V^L \cdot I^*\right)\right)\right)\right)}{p\left(1 - h\left(\left(V^H \cdot I^*, V^L \cdot I^*\right)\right)\right) + 1 - p\left(h\left(\left(V^H \cdot I^*, V^L \cdot I^*\right)\right)\right)}, \tag{90}$$

where

$$h\left(\left(V^H \cdot I^*, V^L \cdot I^*\right)\right) = \frac{1 + V^H \cdot I^*}{2 + \left(V^H + V^L\right) \cdot I^*}, \tag{91}$$

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

$$\left(m^{H*}, m^{L*}\right) = \left(V_H \cdot I^*, V_L \cdot I^*\right). \tag{92}$$

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

Proof: By Lemma 1, optimal mobility in period t in terms of period-t informativeness is given by $\left(m_t^H, m_t^L\right) = \left(V^H \cdot I_t, V^L \cdot I_t\right)$. By Lemma 2, (8) is satisfied with $h(\widehat{m}) = \frac{1+\widehat{m}_H}{2+\widehat{m}_H+\widehat{m}_L}$ since 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. Hence, in steady-state equilibrium, (90) and (92) hold.

Define

$$\Gamma\left(I, p(\cdot), V^H, V^L\right) \equiv \frac{p\left(h\left(\left(V^H \cdot I, V^L \cdot I\right)\right)\right) - \left(1 - p\left(1 - h\left(\left(V^H \cdot I, V^L \cdot I\right)\right)\right)\right)}{p\left(1 - h\left(\left(V^H \cdot I, V^L \cdot I\right)\right)\right) + 1 - p\left(h\left(\left(V^H \cdot I, V^L \cdot I\right)\right)\right)}. \tag{93}$$

Then,

$$\frac{\partial \Gamma}{\partial I} = \frac{\partial \Gamma}{\partial h}\frac{\partial h}{\partial I} > 0, \tag{94}$$

since $\frac{\partial \Gamma}{\partial h} \geq 0$ by the proof of Lemma 2 and since

$$\frac{\partial h}{\partial I} = \frac{\partial}{\partial I}\left[\frac{1 + V^H \cdot I}{2 + \left(V^H + V^L\right) \cdot I}\right] = V^H - V^L > 0. \tag{95}$$

As I is bounded, by Tarski's fixed point theorem, a fixed point I^* satisfying $I^* = \Gamma(I^*)$ exists.

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.

S. Machin, S. McNally, and G. Wyness. Educational attainment across the UK nations: performance, inequality and evidence. *Educational Research*, 55(2):139–164, 2013.

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.

L. Richards, E. Garratt, A.F. Heath, L. Anderson, and E. Altintas. The childhood origins of social mobility: socio-economic inequalities and changing opportunities. *Social Mobility Commision*, June 2016.

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.

B. Shaw, S. Baars, L. Menzies, M. Parameshwaran, and R. Allen. Low income pupil's progress at secondary school. *Social Mobility Commision*, February 2017.

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.