Optimal application behavior with incomplete information

Éva Nagypál

May 2004

Abstract

This paper studies the application behavior of agents in an environment where agents with different underlying qualities but with identical preferences are competing for placements (in colleges, on sports teams, etc) and are allowed to submit multiple applications. Two aspects of incomplete information are considered that make the application behavior in these environments an interesting decision problem to study: uncertainty about one’s own quality and noise in the evaluation of applications. First, I consider what happens when agents face uncertainty about their own quality (interpreted as their ranking among the pool of applicants), need to form beliefs about their quality and make applications based on these beliefs. I show that with a fixed number of applications, given that the distribution characterizing the belief of the agents satisfies the monotone likelihood ratio property (MLRP), there is assortative matching on the expected quality of the agent in the sense that all applications are increasing in the expected quality of the agent. When agents are allowed to decide how many applications to submit given some application cost, however, this result breaks down. It is possible to characterize the optimal application behavior locally and in the limit, but the optimal number of applications shows no regularity globally. I show that, even if MLRP holds, it is possible for an agent with a better underlying quality and a better mean belief to secure a worse placement, something that is not possible with a fixed number of applications. I also show that the number of applications — for high enough dispersion — is decreasing in the dispersion of beliefs, and for sufficiently high dispersion, agents submit no application or submit a single application for the best placement.

Second, I study what happens when in the environment studied a random element is added to the evaluation of applications. I show that agents with low dispersion in their beliefs are the ones to submit multiple applications. The results thus show that both aspects of incomplete information imply that agents facing more uncertainty about their own quality submit fewer applications, and hence secure worse placements. In a companion paper I embed the application decision of the agents into a general equilibrium model of college applications where colleges optimally chose which students to admit, and study the properties of its equilibrium.

\footnote{Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2600, USA. Email: nagypal@northwestern.edu.}
1 Introduction

The decision to apply to college is the first major economic decision most young adults undertake as they embark upon their adult life. The stakes are perceived to be large, as is manifested by the tremendous attention paid to the admission policies of the most selective colleges by students, their parents, and commentators at large. The evidence available on labor market performance also points in the direction of considerable labor market payoffs to attending a selective college (Brewer, Eide, and Ehrenberg (1999)).

Despite the importance that this decision has in the early stage of an individual’s economic life, the theoretical literature studying the college application decision is not very large. The most prominent strand of this literature is the game theory and mechanism design literature following Gale and Shapley’s seminal work on the so-called “college admission problem” (see Gale and Shapley (1962)). In the college admission problem students and colleges have diverse preferences over which colleges to attend and which students to admit and the task is to design a mechanism that results in an individually rational and stable outcome.

While the heterogeneity of preferences in a complete information environment is certainly an important aspect of the college application decision, issues relating to incomplete information are equally important. In this paper I study two important sources of incomplete information in the college application process.

First — and this is the type of uncertainty that is the main focus of this paper — high school graduates generally do not fully know how well they rank in the overall pool of potential college applicants. Of course, they might know their quality in terms of their high school ranking and if they take the SAT (which is itself already a cost of applying to college and is something that is done by a select group of high-school graduates) they know their SAT scores, but these pieces of information do not fully reveal the ranking of a college applicant within the pool of all potential applicants. Moreover, I not only study how the presence of uncertainty affects application decisions, but I also study how the “extent” of uncertainty faced by a student affects application decisions. In other words, I explore
how being “informationally disadvantaged” in the sense of facing more uncertainty about one’s own quality affects application decisions and the subsequent chance of admission. I thereby explore a new channel through which students from disadvantaged backgrounds can be affected in the college admission process. I find the assumption that students from disadvantaged backgrounds have more diffuse information about their own quality a natural one, since they tend to have fewer peers and family members who have themselves applied to or attended college, and they tend to attend high schools where going onto college is not the norm. In a companion paper I embed the application decision of the students in this paper into a general equilibrium model of college applications where colleges optimally chose their application policies, study the properties of its equilibrium (which is generally not socially optimal), and show that colleges would optimally set different admission standards for students from “informationally disadvantaged” groups.

The second source of incomplete information that I explore is that students’ face some uncertainty in the evaluation process. This means that even if students were fully aware of their own ranking within the pool of college applicants and had the same preferences over colleges, one would still expect them not to sort perfectly on quality to the different quality colleges, due to the presence of “noise” in the admission process. In other words, while the assessment of a student’s ranking within the college applicant pool across colleges is certainly correlated, this correlation is far from perfect. In this paper I show that this kind of uncertainty interacts in a non-trivial way with uncertainty about one’s own type in students’ application decision.

Of course, while the empirical phenomenon that I am interested in studying is that of college applications, the results of this paper apply more generally in that they characterize the application behavior of agents in an environment where agents with different underlying qualities but with identical preferences compete for placements (in colleges, on sports teams, etc) and are allowed to submit multiple applications. In these environments when agents face uncertainty about their own quality (interpreted as their ranking among the pool of applicants), they need to form beliefs about their own quality and make applications based on these beliefs. I show that with a fixed number of applications, given that the distribu-
tion characterizing the belief of the agents satisfies the monotone likelihood ratio property (MLRP), there is assortative matching on the expected quality of the agent in the sense that all applications are increasing in the expected quality of the agent. When agents are allowed to decide how many applications to submit given some application cost, however, this result breaks down. It is possible to characterize the optimal application behavior locally and in the limit, but the optimal number of applications shows no regularity globally. I show that, even if MLRP holds, it is possible for an agent with a better underlying quality and a better mean belief to secure a worse placement, something that is not possible with a fixed number of applications. I also show that the number of applications — for high enough dispersion — is decreasing in the dispersion of beliefs, and for sufficiently high dispersion, agents submit no application or take a long shot and submit a single application for the best placement.

When a random element is added to the evaluation of applications, I show that agents with low dispersion in their beliefs are again the ones to submit multiple applications. This is because agents with high dispersion in their belief rationally interpret a rejection as a bad signal about their quality, making subsequent applications less profitable in expectation, while agents with low dispersion rationally interpret a rejection as bad luck without much affecting the expectation about their chance of securing a placement elsewhere. The results thus show that both aspects of incomplete information imply that agents facing more uncertainty about their own quality submit fewer applications, and hence secure worse placements.

Beyond addressing the issue of college applications, this paper is closely related to the literature on matching since it considers the decision whom to match with in a matching model with two-sided heterogeneity when meetings between partners are directed (through applications). Most matching models assume that agents are homogeneous and meetings between agents are random which means that the agents in the model cannot influence whom they will have an opportunity to match with in the subsequent period (see the search and matching literature reviewed in Pissarides (2000)). More recently, however, a literature has developed that focuses on the effects of heterogeneity on matching and search outcomes. While most papers in this literature still assume random search (for example, Burdett and Coles (1997)), there are a handful of papers that consider directed search or directed meet-
ings with heterogeneous agents. The starting point for these papers and the literature on matching with heterogeneous agents is Becker’s famous result (Becker (1973)) in which he establishes that with transferable utility, given that traits of the partners are complements (i.e., given a supermodular production function), perfectly assortative matching (PAM) is the optimal matching pattern and it is achieved in equilibrium. Similarly, perfectly assortative matching is the optimal matching pattern in an environment with heterogeneous agents, directed meetings, and non-transferable utility. Perfectly assortative matching means that, in the case of the labor market, the highest quality worker matches with the highest quality firm, the next highest quality worker matches with the next highest quality firm, and so on.

This very stark prediction of perfect assortative matching is in contrast with empirical observations and intuition. There has been considerable effort in the literature to establish more general conditions under which the PAM result holds or breaks down (for example, Shimer and Smith (2000)). A recent paper emphasizing the role of coordination frictions in breaking down PAM is Shimer (2001).

The role of incomplete information has received practically no attention in this discussion. There are only a handful of other extension of the matching model with heterogeneous agents that consider an incomplete information framework. Anderson and Smith (2004) focus on an aspect of incomplete information about own quality that is different from the focus of this paper. They show that in a dynamic setting, PAM may fail even with a supermodular production function when matches yield not only output but also information about participating agents’ types. They assume a much simpler form of belief heterogeneity than the one assumed in this paper (beliefs are binary in their model) and do not consider multiple applications, hence the questions asked in this paper about the effects of incomplete information cannot be studied within their framework. Chade (2003) addresses the issue of matching when agents have incomplete information about the quality of their partner, which is again different from the issues studied here. Finally, Chade, Lewis, and Smith (2004)

---

The expression “directed search” is used in the literature to describe a situation in which agents can influence which agents they meet with and hence have the possibility to form a match with. “Directed meetings” is a term that I use to describe an extreme case of “directed search” where agents can exactly determine whom they have the possibility to form a match with. In a sense, there is no search nor are there search frictions in the case of “directed meetings”.

4
consider a setup similar to the one studied in this paper, but they study the application
problem of students who know their own quality but have to apply to colleges knowing that
their quality will be evaluated with noise. In this sense, their work is closest to the model
studied in Section 4, and while they consider a general equilibrium setup not considered in
this paper, they do not study how incomplete information about oneself might interact with
noisy evaluation.

An important issue in matching models is whether utility between the parties in a match is
transferable, for example via wages, or not. In the model studied I assume that utility is
not transferable between the partners in a match\(^3\), hence the model is more appropriate to
study markets where this assumption is more plausible, such as the market for college admis-
sions. This is the reason why the model is phrased in terms of college application decisions,
even though, as mentioned above, its implications are more broadly applicable to matching
markets with incomplete information and costly applications. While the assumption of
non-transferable utility does not completely characterize the college admission process, since
colleges actively price discriminate by using financial aid policies, most admission policies
still aim to be “need-blind”. The extent of price discrimination is also limited by the fact
that colleges have set tuition rates, and at their set tuition rates the demand for selective
colleges still far exceeds supply, implying that price discrimination and hence transferability
of utility is far from perfect between colleges and their students.

In the model I consider agents (henceforth high school graduates or students) are heteroge-
enous in their belief about their own quality. This quality can be thought of as the ranking
of the agent among the pool of college applicants. Students’ beliefs are characterized by a
distribution function parameterized by its mean and variance. These students then decide
which colleges, if any, to apply to. Each application has a fixed cost and gains the possibility
for the student of being admitted by the college she applied to. There are a continuum of
colleges with different reputations. Students have the same preferences over colleges, i.e.,
they all prefer colleges with better reputations, and this is the only attribute of colleges that

\(^3\)A good comparison of models with non-transferable utility to those with transferable utility can be found
in Burdett and Coles (1999).
they care about. Better reputation colleges have higher admission standards, meaning that they have a higher quality cutoff for accepting students. I study two versions of the model. In the first, colleges observe the actual quality (absolute ranking) of an applicant, while in the second they do not observe the actual quality (absolute ranking) of an applicant, instead they observe a noisy realization of it.

2 A simple model with uncertainty about own quality

Most of the intuition behind the results in the general model can be well demonstrated using a simple example when there are only two colleges in the economy. I will review this case in detail in this section before stating the results for the general case in Section 3. The two colleges differ in terms of the payoff a student gets from attending them, i.e., their desirability (which I will call reputation from now on). The high-reputation college has reputation \( \eta_h \), while the low-reputation college has reputation \( \eta_l < \eta_h \). The threshold quality requirement of the high-reputation college is \( q_h \), while the threshold quality requirement of the low-reputation college is \( q_l < q_h \).

Each student in the economy has a distinct quality, \( \mu \), that is drawn from a normal distribution with mean 0 and variance \( \sigma^2_\mu \). The student does not know her own quality. Instead, she gets a signal about her quality, \( \hat{\mu} = \mu + \varepsilon \), where \( \varepsilon \) has a distribution \( N(0, \sigma^2_\varepsilon) \). Once the student receives this signal, she updates her belief about her quality, and her posterior belief will be normal with mean

\[
\tilde{\mu} = \frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2_\varepsilon} (\mu + \varepsilon),
\]

and variance

\[
\tilde{\sigma}^2 = \frac{\sigma^2_\varepsilon \sigma^2_\mu}{\sigma^2_\mu + \sigma^2_\varepsilon}.
\]

Since there is a one-to-one correspondence between \( \sigma^2_\varepsilon \) and \( \tilde{\sigma}^2 \), I can treat the latter as the
primitive of the model, which will turn out to be convenient.

Once the student updates her belief about her quality, she decides whether to apply to any of the two colleges in the economy. When making this decision, the student is maximizing her expected payoff. There is a cost of applying to college, \( c \geq 0 \) per application. Since there are two colleges in the economy, the student submits at most two applications. (There is no benefit from submitting multiple applications to the same college.) If the student’s quality is above the threshold quality requirement of a college she applied to, then she is accepted by that college. Finally, the student decides which college to enroll of the ones that accepted her, and if the student enrolls college \( j \), then she gets a payoff of \( \eta_j > 0 \). If the student does not enroll college, her payoff is 0.

Consider the application policy of the students given thresholds \( q_l < q_h \) of the two colleges. (These threshold qualities are considered exogenous in this paper. For an equilibrium derivation of the threshold qualities, see the companion paper.) A student with posterior belief \((\tilde{\mu}, \tilde{\sigma})\) has four options: she can choose not to apply to any college, she can apply to the low-reputation college, to the high-reputation college, or to both colleges. The optimal policy of a student with any belief can be described by comparing the above alternatives, two at a time, and establishing the values of \((\tilde{\mu}, \tilde{\sigma})\) for which each is preferred.

### 2.1 Pairwise comparison of alternative policies

#### 2.1.1 Applying to a single college versus not applying

If the student with belief \((\tilde{\mu}, \tilde{\sigma})\) applies to the low-reputation college, then her expected payoff is

\[
\bar{F}\left(\frac{q_l - \tilde{\mu}}{\tilde{\sigma}}\right) \eta_l - c,
\]

where \(\bar{F}\) is the survival function of the standard normal distribution, i.e., \(\bar{F} = 1 - F\), where \(F\) is the c.d.f. of the standard normal distribution. One necessary condition for this expected
payoff to be higher than 0, the payoff from not applying, is that \( c < \eta_l \). This assumption is maintained throughout, since it simply means that the cost of applying to the low-reputation college is outweighed by the payoff from attending that college. Then, given this assumption, the expected payoff is higher than 0 if and only if

\[
\tilde{\mu} \geq q_l - F^{-1} \left( 1 - \frac{c}{\eta_l} \right) \tilde{\sigma}.
\]  

(4)

This relationship defines a curve in \((\tilde{\mu}, \tilde{\sigma})\) space that separates the region where it is better to apply to the low-reputation college from the one where it is better not to apply. (Notice that the \((\tilde{\mu}, \tilde{\sigma})\) space is defined in this case for \( \tilde{\mu} \in \bar{\mathbb{R}} \) and \( \tilde{\sigma} \in [0, \sigma^2_l) \).

This curve is either downward- or upward-sloping, depending on the sign of \( F^{-1} \left( 1 - \frac{c}{\eta_l} \right) \), which depends on whether \( c < \frac{\eta_l}{2} \). This curve is denoted \( m_{nl} \) in Figure 2 and is clearly linear.

Similarly, if the student with belief \((\tilde{\mu}, \tilde{\sigma})\) applies to the high-reputation college, then her expected payoff is

\[
\bar{F} \left( \frac{q_h - \tilde{\mu}}{\tilde{\sigma}} \right) \eta_h - c.
\]  

(5)

Given that \( c < \eta_l < \eta_h \), this expected payoff is higher than 0 if and only if

\[
\tilde{\mu} \geq q_h - F^{-1} \left( 1 - \frac{c}{\eta_h} \right) \tilde{\sigma}.
\]  

(6)

Similarly to the \( m_{nl} \) curve, this relationship defines the \( m_{nh} \) curve in \((\tilde{\mu}, \tilde{\sigma})\) space, as demonstrated in Figure 2. The slope of the \( m_{nh} \) curve is lower than the slope of the \( m_{nl} \) curve, since \( \eta_h > \eta_l \).

### 2.1.2 Applying to the low-reputation college versus applying to both colleges

If the student with belief \((\tilde{\mu}, \tilde{\sigma})\) applies to both colleges, then her expected payoff is

\[
\bar{F} \left( \frac{q_h - \tilde{\mu}}{\tilde{\sigma}} \right) \eta_h + \left[ \bar{F} \left( \frac{q_l - \tilde{\mu}}{\tilde{\sigma}} \right) - \bar{F} \left( \frac{q_h - \tilde{\mu}}{\tilde{\sigma}} \right) \right] \eta_l - 2c.
\]  

(7)
Comparing this with the payoff in Equation (3) from applying to just the low-reputation college, we can see that it is worth applying to both colleges instead of applying to just the low-reputation college if and only if
\[
\bar{F}\left(\frac{q_h - \tilde{\mu}}{\tilde{\sigma}}\right) (\eta_h - \eta_l) - c \geq 0.
\]
(8)

This holds if \(c < \eta_h - \eta_l\) and
\[
\tilde{\mu} \geq q_h - \bar{F}^{-1}\left(1 - \frac{c}{\eta_h - \eta_l}\right) \tilde{\sigma}.
\]
(9)

This relationship defines the \(m_{hl}\) curve in \((\tilde{\mu}, \tilde{\sigma})\) space. Its slope is clearly higher than that of \(m_{nh}\) (as in Figure 2).

### 2.1.3 Applying to the high-reputation college versus applying to both colleges

Comparing the payoff from applying to both colleges in Equation (7) to that from applying to the high-reputation college in Equation (5), we can see that it is worth applying to both colleges instead of applying to just the high-reputation college if
\[
\bar{F}\left(\frac{q_l - \tilde{\mu}}{\tilde{\sigma}}\right) - \bar{F}\left(\frac{q_h - \tilde{\mu}}{\tilde{\sigma}}\right) \geq \frac{c}{\eta_l}.
\]
(10)

This relationship defines the \(m_{bh}\) curve in \((\tilde{\mu}, \tilde{\sigma})\) space in Figure 2. That the \(m_{bh}\) curve has the depicted shape follows from the properties of the normal distribution. In particular, that it does not include a region beyond some maximum \(\tilde{\sigma}\) is the result of the atomless property of the normal distribution, since it implies that \(\bar{F}\left(\frac{q_l - \tilde{\mu}}{\tilde{\sigma}}\right) - \bar{F}\left(\frac{q_h - \tilde{\mu}}{\tilde{\sigma}}\right)\) is declining in \(\tilde{\sigma}\) for large enough \(\tilde{\sigma}\).

### 2.1.4 Applying to the high-reputation versus the low-reputation college

Comparing the payoff from applying to the high-reputation college in Equation (5) versus that from applying to the low-reputation college in Equation (3), we can see that it is worth
applying to the high-reputation college instead of applying to the low-reputation college if and only if

\[
\frac{F\left(\frac{q_h - \tilde{\mu}}{\tilde{\sigma}}\right)}{F\left(\frac{q_l - \tilde{\mu}}{\tilde{\sigma}}\right)} \geq \frac{\eta_l}{\eta_h}.
\] (11)

This relationship defines the \( m_{th} \) curve in \((\tilde{\mu}, \tilde{\sigma})\) space in Figure 2. That the \( m_{th} \) curve has the depicted shape follows from the properties of the normal distribution, in particular from the fact that the normal distribution satisfies the monotone likelihood ratio property.

### 2.2 The role of the application cost

Based on the value of the application cost, \( c \), we can distinguish two cases based on the value of the model’s exogenous parameters (notably, not based on the value of \( q_l \) and \( q_h \).) First, if \((\eta_h - \eta_l) \frac{\mu_l}{\eta_h} \leq c\), then the \( m_{bl} \) and \( m_{bh} \) curves do not intersect (since that would imply \( F\left(\frac{q_l - \tilde{\mu}}{\tilde{\sigma}}\right) \geq 1\), which is not possible for \( \tilde{\sigma} > 0 \).) This case is represented in Figure 1. Note that, when \( \eta_h - \eta_l \leq c \), then there is no part of the \((\tilde{\mu}, \tilde{\sigma})\) space in which applying to both colleges is preferred to applying to just the low-reputation college (i.e., there is no \( m_{bl} \) curve), but this is qualitatively the same as when there is an \( m_{bl} \) curve that does not intersect the \( m_{bh} \) curve. The empty circles represent \((\tilde{\mu}, \tilde{\sigma})\) combinations for which it is preferable to apply to the low-reputation college, while the small dots represent posterior beliefs for which it is preferable to apply to the high-reputation college. (Circles with dots in them represent posterior beliefs for which it is preferable to apply to both colleges, though there are no such beliefs in this case.)

This is the case when the cost of the second application is prohibitively expensive so that a second application is never submitted. While this case features no multiple applications, it demonstrates several interesting results. First, for a given posterior standard deviation, the PAM result holds in the sense that there is PAM along the posterior means of the agents, those with high posterior means (good signals) match with the high-reputation college, while those with low posterior means (bad signals) match with the low-reputation college. (An
equivalent result holds in the static version of the model of Anderson and Smith (2004).

Figure 1: Optimal application policy of students as a function of posterior mean and standard deviation in the case when \((\eta_h - \eta_l) \frac{\mu}{\eta_h} \leq c\).

More interestingly, however, we can see from Figure 1 that students with the same posterior mean but different posterior standard deviation will generally follow different application policies. For very high values of the posterior standard deviation, students will “take long shots” and apply to the high-reputation college or not apply at all, without considering to apply to the low-reputation college for any value of the posterior mean. For lower values of the posterior standard deviation, students with intermediate values of the posterior mean will “play it safe” and apply to the low-reputation college. This means that some students with the same posterior mean will follow different application policies and potentially will enroll different colleges. Thus Becker’s result no longer holds in the simple sense of PAM on posterior means. It is also the case that taking a group of students with the same quality \(\mu\) and different precision signals which imply different posterior variances \(\tilde{\sigma}\), both the value of the signal and the precision of the signal will influence the application policy of students. This is because for a given \(\mu\), the posterior mean \(\tilde{\mu}\) has distribution \(N\left(\frac{\sigma^2 - \tilde{\sigma}^2}{\sigma_{\tilde{\mu}}^2} \mu, \frac{\tilde{\sigma}^2}{\sigma_{\tilde{\mu}}^2} \left(\sigma^2 - \tilde{\sigma}^2\right)\right)\). For a given \(\mu\) then, the average posterior mean starts out at \(\mu\) when \(\tilde{\sigma} = 0\) and approaches 0 (the mean quality) as \(\tilde{\sigma}\) approaches \(\sigma_{\tilde{\mu}}\).
The second case, when \( c < (\eta_h - \eta_l) \frac{m_l}{\eta_h} \), is displayed in Figure 2. In this case the \( m_{bl} \) and \( m_{bh} \) curves intersect. Once again, the empty circles represent \((\tilde{\mu}, \tilde{\sigma})\) combinations for which it is preferable to apply to the low-reputation college, the small dots represent posterior beliefs for which it is preferable to apply to the high-reputation college and the circles with dots in them represent posterior beliefs for which it is preferable to apply to both colleges.

This case is more interesting than the previous one in the sense that it features multiple applications for some posterior beliefs. The \((\tilde{\mu}, \tilde{\sigma})\) combinations for which students will choose to apply to both colleges are relatively low values of the posterior standard deviation coupled with posterior means around the cutoff of the high-reputation college. These students have a quite precise idea about their own quality and realize that being so close to the cutoff of the high-reputation college, they run the risk of “just not making it”, i.e., ending up with a quality that is just below the cutoff.
3 The general model

There are a continuum of colleges in the economy that differ in their reputation, \( \eta \in [\underline{\eta}, \overline{\eta}] \), which determines the payoff a student gets from attending each college. Moreover, the colleges differ in terms of their threshold quality requirement, denoted by \( q(\eta) \). It is assumed that, \( q(\eta) \) is increasing, meaning that colleges that provide a higher payoff to students attending them are more stringent in their quality requirement.\(^4\) I also introduce the inverse of \( q(\eta) \), denoted by \( \eta(q) \), since in what follows, it will prove easier to work with the inverse notation.

**Assumption 1.** \( q(\eta) : [\underline{\eta}, \overline{\eta}] \rightarrow \mathbb{R} \) is continuous, differentiable and strictly increasing on \([\underline{\eta}, \overline{\eta}]\). Hence it is invertible with a continuous, differentiable, and strictly increasing inverse function \( \eta(q) : [\underline{q}, \overline{q}] \rightarrow [\underline{\eta}, \overline{\eta}] \).

Each student has a belief about her own quality characterized by the distribution function \( F(\mu) \). Instead of specifying how a student arrives to having the beliefs she has, I make more general assumptions on the distribution function \( F(\mu) \) that allow me to derive results about the application behavior of students. \( F(\mu) \) can, of course, be the result of Bayesian updating as in the simple model in Section 2, but this need not be the case.

**Assumption 2.** \( F(\mu) \) is a probability distribution function with support on \((-\infty, \infty)\). \( F(\mu) \) is continuous, strictly increasing and continuously differentiable.

**Assumption 3.** \( F(\mu) \) belongs to a class of distribution functions parameterized by \( \tilde{\mu} \) and \( \tilde{\sigma} \) such that

\[
F(\mu | \tilde{\mu}, \tilde{\sigma}) = F \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \mid 0, 1 \right).
\]

**Assumption 4.** \( F(\mu | \tilde{\mu}, \tilde{\sigma}) \) satisfies the strict monotone likelihood ratio property, i.e., for any \( \mu_1 < \mu_2 \), \( \frac{F(\mu_1 | \tilde{\mu}, \tilde{\sigma})}{F(\mu_2 | \tilde{\mu}, \tilde{\sigma})} \) is strictly increasing in \( \tilde{\mu} \).

\(^4\)In a companion paper, I consider the optimizing behavior of colleges when setting their admission standards and the resulting equilibrium. I show there that, taking the payoff of the colleges — the distribution of \( \eta \) — as the primitive of the model, the equilibrium admission standard, \( q(\eta) \), is in fact increasing in \( \eta \).
Next, I consider two alternative assumptions about the application behavior of a student. In Section 3.1, I consider the case when students can submit a fixed number of applications, while in 3.2 I endogenize the application intensity by allowing students to submit any number of applications given some application cost.

### 3.1 Fixed number of applications

**Assumption 5A.** *A student submits a fixed $N$ number of applications.*

Let us denote the ordered optimal application set of a student with belief $F(\tilde{\mu}, \tilde{\sigma})$ by $A(\tilde{\mu}, \tilde{\sigma})$ and its $i^{th}$ element by $\eta^*_i(\tilde{\mu}, \tilde{\sigma})$.

**Theorem 1** (Assortative matching given fixed number of applications). *Given Assumptions 1 through 4 and 5A, for a given $\tilde{\sigma}$, $\eta^*_i(\tilde{\mu}, \tilde{\sigma})$ is strictly increasing in $\tilde{\mu}$ if $\eta^*_i(\tilde{\mu}, \tilde{\sigma}) < \overline{\eta}$.

**Proof:** See Appendix.

If the assumption of strict MLRP were to be substituted with the assumption of MLRP, then the only the change would be that $\eta^*_i(\tilde{\mu}, \tilde{\sigma})$ would be increasing in $\tilde{\mu}$ as opposed to strictly increasing.

Theorem 1 states that students with better beliefs about themselves will apply to a better set of schools, in the sense that their top choice will be higher than the top choice of the students with a worse belief about themselves, their second top choice will be higher than the second top choice of the students with a worse belief, and so on. Theorem 1 is a generalization of the case depicted in Figure 1, since there, too, with a fixed number of applications (assuming that one application is always made), students with higher signals apply to a (weakly) better school. With a continuum of schools, they would be applying to a strictly better school, as long as strict MLRP holds (which does for the normal distribution, for example). As I show in the next section, assortative matching completely breaks down once the assumption of a fixed number applications is relaxed.
3.2 Endogenous number of applications

Now let us assume that instead of submitting a fixed number of applications, a student can submit any number of applications given some application cost.

Assumption 5B. A student can submit any number of applications at a cost of $c$ per application, where $0 < c < \eta$.

The characterization of the optimal policy proceeds by first considering what happens in the limit as $\tilde{\mu}$ becomes very small or very large.

Lemma 2. For a given $\tilde{\sigma}$, there exists a finite $\tilde{\mu}(\tilde{\sigma})$, such that for any $\tilde{\mu} < \tilde{\mu}(\tilde{\sigma})$, $A(\tilde{\mu}, \tilde{\sigma}) = \emptyset$.

Proof. Note that, for any $i$, $\bar{F}(\frac{q_i - \tilde{\mu}}{\tilde{\sigma}}) \eta(q_i)$ is continuous and decreasing in $\tilde{\mu}$ and approaches 0 as $\tilde{\mu}$ approaches $-\infty$. Therefore, at a finite $\tilde{\mu}$ it becomes less than $c$, therefore for low enough $\tilde{\mu}$, the payoff from applying does not justify paying the application cost. One value for $\tilde{\mu}(\tilde{\sigma})$ that would certainly work would be $q - \tilde{\sigma} \bar{F}^{-1}\left(\frac{c}{\eta}\right)$. For any $\tilde{\mu} < \tilde{\mu}(\tilde{\sigma})$ an agent will not submit an application even if the chance of getting into a school is the highest possible, that of getting into the lowest quality school, and the payoff from getting into a school is the highest possible, that from attending the best school.

Simply put, this lemma states that for unfavorable enough beliefs, agents will not find it profitable to submit any applications, since their perception of their chance of getting into any college is too small.

Lemma 3. For a given $\tilde{\sigma}$, there exists a finite $\tilde{\mu}(\tilde{\sigma})$, such that for any $\tilde{\mu} > \tilde{\mu}(\tilde{\sigma})$, $A(\tilde{\mu}, \tilde{\sigma}) = \{\tilde{\eta}\}$.

Proof. This is because for any $q_i < \tilde{q}$, $\left[\bar{F}(q_i - \tilde{\mu}) - \bar{F}(\tilde{q} - \tilde{\mu})\right] \eta(q_i)$ goes to 0 as $\tilde{\mu}$ goes to $\infty$, therefore at a finite $\tilde{\mu}$ it becomes less than $c$, therefore for high enough $\tilde{\mu}$, the payoff from applying to a second school does not justify paying the application cost, while the payoff from applying to the best school is approaching $\eta(\tilde{q}) > c$.

This lemma simply states that for favorable enough beliefs, an agent will submit a single application to the best school.
Theorem 4. Given Assumptions 1 through 4 and 5B, for a given $\tilde{\sigma}$, if the application set at $\tilde{\mu}$ is $A(\tilde{\mu}, \tilde{\sigma}) = \{\eta_1^*, \eta_2^*, ..., \eta_N^*\}$ then there exists $\tilde{\epsilon} > 0$ such that for any $0 < \epsilon < \tilde{\epsilon}$ the optimal application set at $\tilde{\mu} + \epsilon$ takes on one of the following forms:

1. (ordered shift in the optimal policy with no change in the number of applications) $A(\tilde{\mu} + \epsilon, \tilde{\sigma}) = \{\lambda_1^*, \lambda_2^*, ..., \lambda_N^*\}$ such that $\eta_i < \lambda_i$.

2. (ordered expansion in the optimal policy by one application) $A(\tilde{\mu} + \epsilon, \tilde{\sigma}) = \{\lambda_1^*, \lambda_2^*, ..., \lambda_{N+1}^*\}$ such that $\lambda_i < \eta_i < \lambda_{i+1}$, $i = 1, ..., N$.

3. (ordered contraction in the optimal policy by one application) $A(\tilde{\mu} + \epsilon, \tilde{\sigma}) = \{\lambda_1^*, \lambda_2^*, ..., \lambda_{N-1}^*\}$ such that $\eta_i < \lambda_i < \eta_{i+1}$, $i = 1, ..., N - 1$.

Proof: See Appendix.

Theorem 3 states that as $\tilde{\mu}$ changes for a given $\tilde{\sigma}$, the application set changes gradually, by shifting up all applications by a small amount and then either not changing their number, adding one more application at the lowest level, or dropping the highest application.

It is worthwhile to highlight one of the implication of Theorem 3. Jointly with Lemma 3, it implies that, for a given $\tilde{\sigma}$, if multiple applications are submitted for some value of $\tilde{\mu}$, then it must be the case that there exists a $\tilde{\mu}$, let us call it $\tilde{\mu}^a$, such that there is a $\epsilon$ such that at $\tilde{\mu}^a + \epsilon$ the third case from Theorem 3 applies. This means that between $\tilde{\mu}^a$ and $\tilde{\mu}^a + \epsilon$ there is an ordered contraction in the optimal policy by one application in a way such that the highest element of $A(\tilde{\mu}^a, \tilde{\sigma})$ is larger than the highest element of $A(\tilde{\mu}^a + \epsilon, \tilde{\sigma})$. This means that there is a positive probability that an agent with a better belief gets into a worse school. In other words, in the Bayesian updating framework a student of better underlying quality and with a better signal can still end up enrolling a worse quality school than a student of worse underlying quality and with a worse signal. An example of this is given below. This implies lack of assortative matching even on the beliefs or the signals.

Of course, this does not mean that the expected value of having higher belief is lower, indeed the expected value of applications is monotone in $\tilde{\mu}$ for a given $\tilde{\sigma}$. 
Theorem 3 gives a local characterization of the optimal application policy for a given level of uncertainty. To derive a global characterization, we need to examine how the number of applications changes as an agent’s expected quality changes.

Given the above observations, it might seem natural to think that the optimal number of applications is increasing up to some \( \tilde{\mu} \) and then declines. This is not true, however, as the following example demonstrates.

**Example 1.** Let the application cost be equal to \( c = .117 \), let the distribution function be normal with mean \( \tilde{\mu} \) and variance 1 and let the payoff function be piece-wise linear.

\[
\eta(q) = \begin{cases} 
0 & q \in (-\infty, -4/3] \\
1 + .75q & q \in (-4/3, 0] \\
1 + .35q & q \in (0, \infty)
\end{cases}
\]  

Figure 3 plots difference between the value of optimally applying to college given that the student submits one application and the value applying to college given that the student submits two applications. (In this example, it is never beneficial for the student to submit more than two applications.)

Figure 4 plots the optimal application behavior of a student as a function of \( \tilde{\mu} \).

This example is also useful to demonstrate the above discussed implication of Theorem 3 that — in a Bayesian updating environment — it is possible that student A is of better underlying quality than student B and student A gets a better signal about quality than student B, nonetheless student B ends up enrolling a better quality college than student A.

**Example 1. continued** Now let us assume that the underlying quality distribution in the population is \( N(0, 2) \), and the student receive signals that are contaminated by an error with a distribution \( N(0, 2) \). Now let us assume that student A has inherent quality \( \mu_A = .30 \), while student B has inherent quality \( \mu_B = .20 < \mu_A \). Student A receives a signal \( \mu_A + \epsilon_A = .30 + .30 = .60 \) and — using the simple Bayesian updating formulas above — updates her belief to \( N(.30, 1) \), while student B receives a signal \( \mu_B + \epsilon_A = .20 + .20 = .40 \) and updates
Figure 3: Value of submitting a second application as a function of the mean of the student’s belief.
Figure 4: Optimal choice of school(s) as a function of the mean of the student’s belief.
his belief to $N(0.20, 1)$. This implies that student A will optimally apply and get into school $q_A = 0$, while student B will optimally apply to schools $q_{B1} = -0.51$ and $q_{B2} = 0.174$, get into both and enroll school $q_{B2} = 0.174$, a better school than the one attended by student A.

Next let us turn to the characterization of the optimal application set as a function of the dispersion of beliefs, $\tilde{\sigma}$. The difficulty in giving a complete characterization is already demonstrated in Figure 2, which shows that once the number of applications is allowed to vary, there is no clear monotone relationship between $\tilde{\sigma}$ and the application set. In fact, by considering an upward-sloping $m_b$ curve, we can construct an example where, as $\tilde{\sigma}$ increases, the agent first applies to the better school, then applies to both schools, then applies to the worse school, and then as $\tilde{\sigma}$ increases further starts applying to the better school again. It turns out, however, that the very last part of the constructed example is the result of a general pattern, namely that, as $\tilde{\sigma}$ increases, there comes a point where students only take long shots or no shots at all.

**Theorem 5** (Long shot theorem). There exists $\hat{\sigma} < \infty$ such that for all $\tilde{\mu} \in \mathbb{R}$ and $\tilde{\sigma} > \hat{\sigma}$ the acceptance set is $A(\tilde{\mu}, \tilde{\sigma}) = \{\eta\}$ or $A(\tilde{\mu}, \tilde{\sigma}) = \emptyset$.

**Proof:** See Appendix.

This means that no multiple applications will exist for high enough $\tilde{\sigma}$, and if there is an application submitted, it is submitted to the college where it is most likely to be turned down, which leads to the expected probability of getting into college decrease with $\tilde{\sigma}$. The intuition for this result is the following. Whenever a student decides to apply to multiple schools, the second and third and so on applications are submitted in case the student “falls between the lines”, meaning that she is good enough for one school, but not good enough for a better school. This means that multiple applications are a way to “fine-tune” the payoff of a student. When $\tilde{\sigma}$ is large, then such fine-tuning has very little payoff, since the perceived probability that the student “falls between the lines” is very small (goes to zero as $\tilde{\sigma}$ goes to $\infty$). Therefore such fine-tuning is only worth paying the application cost for if the student has a relatively precise idea about her own quality.
4 Noise in the evaluation of applications

Now consider the following alternative scenario. For simplicity, there are again two colleges in the economy. These two colleges now, however, are exactly the same in terms of the payoff a student gets from attending them. The threshold quality requirement of each college is $q$. The information structure on the side of the student is the same as in Section 2, hence the student’s belief about her quality is normal with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$.

Once the student updates her belief about her quality, she decides how many of the colleges to apply to, if any. If a student applies to a college, then that college receives a signal about the student $\hat{\mu} = \mu + \nu_j$, where $\nu_j$ is independent across colleges and students and is distributed $N(0,1)$.\footnote{Assuming that the noise has a variance of 1 is without loss of generality, since what is important in the inference process is its relative magnitude compared to $\sigma_\mu$, which is allowed to take on any value.} If this signal is above the threshold $q$, then the student is accepted, otherwise she is rejected.

Once again, the student is maximizing her expected payoff and again the cost per application is $c$. Since there are two colleges in the economy, the student submits at most two applications. A student with posterior belief $(\tilde{\mu}, \tilde{\sigma})$ has three options: she can choose not to apply to any college, she can apply one college or to two colleges. The optimal policy of a student with any belief can be described by comparing the above alternatives, two at a time, and establishing the values of $(\tilde{\mu}, \tilde{\sigma})$ for which each is preferred.

Intuitively, the decision of students to make multiple applications is driven by the fact that the extra payoff from applying to an additional college of the same quality is higher if the correlation between the signals that the colleges receive is lower. In the extreme, if the signals are perfectly correlated, then there is no reason for students to submit multiple applications.

The unconditional correlation between the signals that the two colleges receive is $\frac{\sigma_\mu^2}{\sigma_\mu^2 + 1}$. The correlation conditional on the information available to the student is lower, however, it is

$$\frac{\tilde{\sigma}_\mu^2}{\tilde{\sigma}_\mu^2 + 1}. \tag{14}$$
which is clearly increasing in $\tilde{\sigma}_\mu^2$. Hence students with higher precision about their own quality (lower posterior variance) will perceive the correlation between the signals about them that the colleges receive to be lower. This means that it is these students that will benefit more from making multiple applications. To put it differently, if a college rejects a high uncertainty student with a given posterior mean, she rationally interprets it as a signal that she is in fact not good enough to get into that quality college, while if a college rejects a low uncertainty student with the same posterior mean, she rationally interprets it as “bad luck”, i.e. a low realization of the noise term $\nu_j$. This is because the admission decision of a college carries lower information content for a well-informed student.

The probability of getting into college if making just one applications is $P_1(\tilde{\mu}, \tilde{\sigma})$, where this is equal to

$$P_1(\mu + \nu_1 \mid \tilde{\mu}, \tilde{\sigma}) = \int_{-\infty}^{\infty} \frac{1}{\sigma} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) P(\nu_1 > q - \mu) d\mu = \int_{-\infty}^{\infty} \frac{1}{\sigma} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \bar{F}(q - \mu) d\mu. \quad (15)$$

**Proposition 6.** $P_1(\tilde{\mu}, \tilde{\sigma})$ is increasing in $\tilde{\mu}$ and is increasing in $\tilde{\sigma}$ if and only if $\tilde{\mu} \leq q$.

**Proof:** See Appendix.

Proposition 6 implies that in $(\tilde{\sigma}, \tilde{\mu})$-space, the indifference curve between applying to a single college versus not applying at all is monotone. If it intersects the $\tilde{\mu}$ axis below $q$, which occurs if $P_1(q, 0) \eta - c \geq 0$, i.e., if $c \leq \frac{\eta}{2}$, then the indifference curve is downward-sloping. On the contrary, if it intersects the $\tilde{\mu}$ axis above $q$, which occurs if $c \geq \frac{\eta}{2}$, then the indifference curve is upward-sloping.

The probability of getting into college if making two applications is $P_2(\tilde{\mu}, \tilde{\sigma})$, where this is
equal to

\[ P_2(\mu + \nu_1 > q \text{ or } \mu + \nu_2 > q \mid \tilde{\mu}, \tilde{\sigma}) = \]

\[ = \int_{-\infty}^{\infty} \frac{1}{\sigma} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ 1 - P(\nu_1 < q - \mu)P(\nu_2 < q - \mu) \right] d\mu = \]

\[ = \int_{-\infty}^{\infty} \frac{1}{\sigma} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ 1 - F(q - \mu)^2 \right] d\mu. \]

It is preferable to apply to two colleges as opposed to just one if the payoff from doing so is higher, i.e., when

\[ P_2(\tilde{\mu}, \tilde{\sigma}) \eta - 2c \geq P_1(\tilde{\mu}, \tilde{\sigma}) \eta - c, \]  \hspace{1cm} (16)

or

\[ P_2(\tilde{\mu}, \tilde{\sigma}) - P_1(\tilde{\mu}, \tilde{\sigma}) = \int_{-\infty}^{\infty} \frac{1}{\sigma} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) F(q - \mu) F(q - \mu) d\mu \geq \frac{c}{\eta}. \]  \hspace{1cm} (17)

**Proposition 7.** \( P_2(\tilde{\mu}, \tilde{\sigma}) - P_1(\tilde{\mu}, \tilde{\sigma}) \) is increasing in \( \tilde{\mu} \) if and only if \( \tilde{\mu} \leq q \). Moreover, for each \( \tilde{\mu} \), there exists a \( 0 \leq \tilde{\sigma}(\tilde{\mu}) \leq 2 | q - \tilde{\mu} | \) such that \( P_2(\tilde{\mu}, \tilde{\sigma}) - P_1(\tilde{\mu}, \tilde{\sigma}) \) is decreasing in \( \tilde{\sigma} \) for any \( \tilde{\sigma} \geq \tilde{\sigma}(\tilde{\mu}) \).

**Proof:** See Appendix.

Proposition 7 implies that if any student applies to two colleges, it will be the ones with posterior means close to the cutoff of the colleges and with low posterior variances. Moreover, if for some student finds it profitable to apply to two colleges, then the student who is certain about her quality and has a quality equal to the cutoff quality of the colleges must find it profitable to apply to two colleges.\(^6\) This student finds two applications profitable if \((P_2(q, 0) - P_1(q, 0)) \eta \geq c, \) or \( c \leq \frac{\eta}{4}. \)

Hence, based on the magnitude of the application costs, we can distinguish between three

\(^6\)This is because from Proposition 7 it follows that \( P_2(\tilde{\mu}, \tilde{\sigma}) - P_1(\tilde{\mu}, \tilde{\sigma}) \leq P_2(q, \tilde{\sigma}) - P_1(q, \tilde{\sigma}) \leq P_2(q, 0) - P_1(q, 0). \)
cases. When $c \geq \frac{\eta}{2}$, the indifference curve between applying to one college and not applying is upward-sloping and the application cost is so high that no student finds it preferable to apply to two colleges as opposed to applying to just one. When $\frac{\eta}{4} \leq c < \frac{\eta}{2}$ the same indifference curve is downward-sloping though there are still no multiple applications. This latter case is depicted in Figure 5.

![Figure 5: Optimal application policy of students as a function of posterior mean and standard deviation when there is noise in the decision process in the case when $\frac{\eta}{4} \leq c < \frac{\eta}{2}$.](image)

In the last case, some students (those with fairly precise knowledge about their own quality and having mean beliefs around the cutoff of the colleges) find it preferable to apply to two colleges as opposed to applying to just one, i.e., $c < \frac{\eta}{4}$. This is the case depicted in Figure 6.
Figure 6: Optimal application policy of students as a function of posterior mean and standard deviation when there is noise in the decision process in the case when $c \leq \frac{3}{4}$. 
References


Appendix

Proof of Theorem 1. Let us fix $\tilde{\sigma} = 1$ (it is easy to show that this simply makes the notation a little less cumbersome and is without loss of generality). Given a fixed $N$ number of applications, the objective of the student is to choose $q_1 < q_2 < ... < q_N$ to maximize

$$\sum_{i=1}^{N} F(q_i - \tilde{\mu}) [\eta(q_i) - \eta(q_{i-1})].$$

(For notational convenience let us use the notation $q_0 = -\infty$ and $q_{N+1} = \infty$, so that $\eta(q_0) = 0$ and $F(q_{N+1} - \tilde{\mu}) = 0$.) The optimal choices $q^* = [q_1^*, ..., q_N^*]$ (the arguments of which are suppressed) satisfy the necessary first-order conditions

$$[F(q_i - \tilde{\mu}) - F(q_{i+1} - \tilde{\mu})] \eta'(q_i) - f(q_i - \tilde{\mu}) [\eta(q_i) - \eta(q_{i-1})] = 0. \quad (A-2)$$

The second-order condition for a maximum requires that the Hessian evaluated at $q^*$ be negative definite:

$$H = \begin{bmatrix}
    d_1 & o_1 & 0 & 0 & . & . & . \\
    o_1 & d_2 & o_2 & 0 & 0 & . & . \\
    . & . & . & . & . & . & . \\
    . & 0 & o_{i-1} & d_i & o_i & 0 & . \\
    . & . & . & . & . & . & . \\
    . & 0 & 0 & o_{n-2} & d_{n-1} & o_{n-1} & . \\
    . & . & . & 0 & 0 & o_{n-1} & d_n
\end{bmatrix}, \quad (A-3)$$

where

$$d_i = [F(q_i^* - \tilde{\mu}) - F(q_{i+1}^* - \tilde{\mu})] \eta''(q_i^*) - f'(q_i^* - \tilde{\mu}) [\eta(q_i^*) - \eta(q_{i-1}^*)] - 2f(q_i^* - \tilde{\mu}) \eta'(q_i^*), \quad (A-4)$$

27
and
\[ o_i = f(q_{i+1}^* - \bar{\mu})\eta'(q_i^*). \]  

Notice that \( o_i \) needs to be strictly positive since \( f(q_i^*) > 0 \) by Assumption 2 and \( \eta'(q_i^*) > 0 \), \( i = 1, ..., n - 1 \) since, given Assumption 1, the optimal school choices feature \( q \leq q_1^* < ... < q_{n-1}^* < q_N^* \leq \bar{q} \).

Totally differentiating the \( N \) first-order conditions with respect to \( \bar{\mu} \) and suppressing \( \bar{\mu} \) in what follows, we get
\[
[f(q_i^*) - f(q_{i+1}^*)]\eta'(q_i^*) + f'(q_i^*)\eta(q_i^*) - \eta(q_{i-1}^*)] + o_{i-1} \frac{dq_{i-1}}{d\bar{\mu}} + d_i \frac{dq_i}{d\bar{\mu}} + o_i \frac{dq_{i+1}}{d\bar{\mu}} = 0, \quad (A-6)
\]
or in matrix notation
\[
p + Hg = 0, \quad (A-7)
\]
where \( p \) is a vector of length \( N \) with general element
\[
p_i = [f(q_i^*) - f(q_{i+1}^*)]\eta'(q_i^*) + f'(q_i^*)\eta(q_i^*) - \eta(q_{i-1}^*)], \quad (A-8)
\]
and \( g \) is a vector of length \( N \) with general element
\[
g_i = \frac{dq_i}{d\bar{\mu}}. \quad (A-9)
\]
Multiplying by the inverse of \( H \) and rearranging, we get an expression for the gradient vector
\[
g = -H^{-1}p. \quad (A-10)
\]
First notice that every element of \( p \) is positive given a distribution function satisfying strict
MLRP, because

\[ \frac{p_i}{f(q_i)\eta(q_i) - \eta(q_{i-1})} = \frac{[f(q_i) - f(q_{i+1})]f'(q_i)}{[\eta(q_i) - \eta(q_{i-1})]f(q_i)} + \frac{f'(q_i)}{f(q_i)} = \frac{f(q_i) - f(q_{i+1})}{F(q_i) - F(q_{i+1})} + \frac{f'(q_i)}{f(q_i)} = \]

\[ = \frac{-\int_{q_i}^{q_{i+1}} f'(q) dq}{\int_{q_i}^{q_{i+1}} f(q) dq} + \frac{f'(q_i)}{f(q_i)} = -\int_{q_i}^{q_{i+1}} \frac{f'(q)}{f(q)} f(q) dq + \frac{f'(q_i)}{f(q_i)} > \]

\[ > \frac{-\int_{q_i}^{q_{i+1}} f(q) dq}{\int_{q_i}^{q_{i+1}} f(q) dq} + \frac{f'(q_i)}{f(q_i)} = 0, \]

where the second equality follows from the first-order condition, while the inequality follows from the fact that a distribution function satisfies strict MLRP iff \( \frac{f'(q)}{f(q)} \) is strictly declining.

To show that all the elements of \( g \) are positive, it is then sufficient to show that all the elements of \( H^{-1} \) are negative. This, however, immediately follows from the fact that \( H \) is negative definite, and has the tridiagonal form in (A-3).

Proof of Theorem 5. The payoff from applying to school \( \bar{q} \) is

\[ \bar{F} \left( \frac{\bar{q} - \tilde{\mu}}{\tilde{\sigma}} \right) \eta(\bar{q}) - c. \] (A-11)

This payoff is increasing in \( \tilde{\mu} \), hence at any given \( \tilde{\sigma} \) for a high enough value of \( \tilde{\mu} \), a student with belief \( F(\tilde{\mu}, \tilde{\sigma}) \) will prefer to apply to school \( i \) than not to apply at all as long as \( \eta(\bar{q}) > c \), while for low enough value of \( \tilde{\mu} \), a student with belief \( F(\tilde{\mu}, \tilde{\sigma}) \) will prefer to not apply at all to college to applying to school \( \bar{q} \).

The difference in payoff from applying to the best school \( \bar{q} \) and to a lower quality school
\[ \hat{q} < \tilde{q} \text{ is} \]
\[ \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) \eta(\tilde{q}) - \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) \eta(\tilde{q}) = \]
\[ = \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) \left[ \eta(\tilde{q}) - \eta(\hat{q}) \right] - \left[ \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) - \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) \right] \eta(\hat{q}) \]
\[ \geq \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) \left[ \eta(\tilde{q}) - \eta(\hat{q}) \right] - \left[ \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{2\sigma} \right) - \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{2\sigma} \right) \right] \eta(\hat{q}), \]

where the inequality follows from the fact that, for a given \( a \), \( F(x + a) - F(x) \) is maximized at \( x = -a/2 \). Notice that since \( \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{2\sigma} \right) - \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{2\sigma} \right) = 0 \) and \( \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) \) is increasing in \( \tilde{\sigma} \) with \( \lim_{\tilde{\sigma} \to \infty} \tilde{F} \left( \frac{\hat{q} - \tilde{\mu}}{\sigma} \right) = F(0) \), the lower bound on the difference is increasing in \( \tilde{\sigma} \) with a strictly positive limit as \( \tilde{\sigma} \to \infty \), hence there exists \( \tilde{\sigma} < \infty \) such that for all \( \tilde{\sigma} > \tilde{\sigma} \) it is preferable for a student to apply only to the best school than to apply only to one of the other schools. Depending then on whether the value from applying to the best school is positive or not, the acceptance set for any \( \tilde{\mu} \in \mathbb{R} \) is \( A(\tilde{\mu}, \tilde{\sigma}) = \{\tilde{q}\} \) or \( A(\tilde{\mu}, \tilde{\sigma}) = \emptyset \).

**Proof of Proposition 6.** Notice that \( P_1(\tilde{\mu}, \tilde{\sigma}) \) can be rewritten as

\[ P_1(\tilde{\mu}, \tilde{\sigma}) = \int_{-\infty}^{\infty} f(\nu) \tilde{F} \left( \frac{\mu - \tilde{\mu} - \nu}{\tilde{\sigma}} \right) d\nu. \]  \hfill (A-12)

Then

\[ \frac{dP_1(\tilde{\mu}, \tilde{\sigma})}{d\tilde{\mu}} = \frac{1}{\tilde{\sigma}} \int_{-\infty}^{\infty} f(\nu) f \left( \frac{\mu - \tilde{\mu} - \nu}{\tilde{\sigma}} \right) d\nu \geq 0. \]  \hfill (A-13)

To determine the sign of \( \frac{dP_1(\tilde{\mu}, \tilde{\sigma})}{d\tilde{\sigma}} \), notice that

\[ \frac{dP_1(\tilde{\mu}, \tilde{\sigma})}{d\tilde{\sigma}} = \int_{-\infty}^{\infty} f(\nu) f \left( \frac{\mu - \tilde{\mu} - \nu}{\tilde{\sigma}} \right) \frac{d\mu - \tilde{\mu} - \nu}{\tilde{\sigma}^2} d\nu = \]
\[ = \int_{-\infty}^{\mu - \tilde{\mu}} f(\nu) f \left( \frac{\mu - \tilde{\mu} - \nu}{\tilde{\sigma}} \right) \frac{d\mu - \tilde{\mu} - \nu}{\tilde{\sigma}^2} d\nu + \int_{\mu - \tilde{\mu}}^{\infty} f(\nu) f \left( \frac{\mu - \tilde{\mu} - \nu}{\tilde{\sigma}} \right) \frac{d\mu - \tilde{\mu} - \nu}{\tilde{\sigma}^2} d\nu = \]
\[ = \int_{\mu - \tilde{\mu}}^{\infty} (f(\nu) - f(2(\mu - \tilde{\mu}) - \nu)) f \left( \frac{\mu - \tilde{\mu} - \nu}{\tilde{\sigma}} \right) \frac{d\mu - \tilde{\mu} - \nu}{\tilde{\sigma}^2} d\nu, \]  \hfill (A-14)
where the second equality follows from the fact that 
\[ f \left( \frac{\mu - \bar{\mu} - \nu}{\sigma} \right) \frac{\mu - \bar{\mu} - \nu}{\sigma^2} = - f \left( \frac{\nu - (\mu - \bar{\mu})}{\sigma} \right) \frac{\nu - (\mu - \bar{\mu})}{\sigma^2}. \]

Notice that \( \mu - \bar{\mu} - \nu \) is always negative over the interval \([q - \bar{\mu}, \infty)\), and if \( \bar{\mu} \leq q \) then \( f(\nu) - f(2(\mu - \bar{\mu}) - \nu) \) is negative over the same interval, otherwise it is positive. This establishes the second part of Proposition 6.

\[ \Box \]

**Proof of Proposition 7.** Notice that

\[
\frac{d(P_2(\mu, \bar{\sigma}) - P_1(\bar{\mu}, \bar{\sigma}))}{d\bar{\mu}} = - \frac{1}{\bar{\sigma}^2} \int_{-\infty}^{\infty} f' \left( \frac{\mu - \bar{\mu}}{\bar{\sigma}} \right) F(q - \mu) \bar{F}(q - \mu) d\mu =
\]

\[
= - \frac{1}{\bar{\sigma}^2} \int_{-\infty}^{\infty} f' \left[ F(q - \mu) F(q - \mu) - F(q - (2\bar{\mu} - \mu)) F(q - (2\bar{\mu} - \mu)) \right] d\mu =
\]

\[
= - \frac{1}{\bar{\sigma}^2} \int_{-\infty}^{\infty} f' \left[ F(q - \mu + 2(\mu - \bar{\mu})) - F(q - \mu) \right] \left[ F(q - \mu + 2(\mu - \bar{\mu})) - F(q - \mu) \right] d\mu,
\]

where the argument of \( f' \) is suppressed and where the second equality follows from the fact that \( f'(x) = -f'(-x) \), and the third equality follows from \( x(1 - x) - y(1 - y) = (y - x)(y - (1 - x)) \). Clearly, for \( \mu \geq \bar{\mu} \), the second term in the integrand is always positive, while the third term is positive if and only if \( \bar{\mu} \leq q \). Hence \( \frac{d(P_2(\mu, \bar{\sigma}) - P_1(\bar{\mu}, \bar{\sigma}))}{d\bar{\mu}} \) is positive if and only if \( \bar{\mu} \leq q \).

To determine the sign of \( \frac{d(P_2(\bar{\mu}, \bar{\sigma}) - P_1(\bar{\mu}, \bar{\sigma}))}{d\bar{\sigma}} \), notice that

\[
\frac{d(P_2(\bar{\mu}, \bar{\sigma}) - P_1(\bar{\mu}, \bar{\sigma}))}{d\bar{\sigma}} = - \frac{1}{\bar{\sigma}^2} \int_{-\infty}^{\infty} \left[ f + f'(\mu - \bar{\mu}) \right] \bar{F}(q - \mu) d\mu =
\]

\[
= \frac{1}{\bar{\sigma}^2} \int_{-\infty}^{\infty} \left[ f \left( \frac{(\mu - \bar{\mu})^2}{\bar{\sigma}^2} - 1 \right) \right] \bar{F}(q - \mu) d\mu,
\]

where once again the arguments of \( f \) and \( f' \) are suppressed, and where the second equality follows from the property of the normal distribution that \( f'(x) = -f(x)x \). Since \( f \left( \frac{\mu - \bar{\mu}}{\bar{\sigma}} \right) \left[ \frac{(\mu - \bar{\mu})^2}{\bar{\sigma}^2} - 1 \right] \) is symmetric around \( \bar{\mu} \), takes on negative values on \( [\bar{\mu} - \bar{\sigma}, \bar{\mu} + \bar{\sigma}] \) and positive values otherwise, and given \( \int_{-\infty}^{\infty} f \left( \frac{\mu - \bar{\mu}}{\bar{\sigma}} \right) \left[ \frac{(\mu - \bar{\mu})^2}{\bar{\sigma}^2} - 1 \right] d\mu = 0 \), we know that

\[
\int_{\bar{\mu}}^{\bar{\mu} + \bar{\sigma}} f \left( \frac{\mu - \bar{\mu}}{\bar{\sigma}} \right) \left[ 1 - \frac{(\mu - \bar{\mu})^2}{\bar{\sigma}^2} \right] d\mu = \int_{\bar{\mu} + \bar{\sigma}}^{\infty} f \left( \frac{\mu - \bar{\mu}}{\bar{\sigma}} \right) \left[ \frac{(\mu - \bar{\mu})^2}{\bar{\sigma}^2} - 1 \right] d\mu, \tag{A-15}
\]

31
and

\[
\int_{\tilde{\mu} - \tilde{\sigma}}^{\tilde{\mu}} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ 1 - \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2} \right] d\mu = \int_{-\infty}^{\tilde{\mu} - \tilde{\sigma}} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2} - 1 \right] d\mu. \tag{A-16}
\]

Now consider the case when \( \tilde{\mu} \leq q \). Then \( F \left( \frac{q - \mu}{\sigma_v} \right) F \left( \frac{q - \mu}{\sigma_v} \right) \) is increasing over the interval \( (-\infty, \tilde{\mu}] \), hence

\[
\int_{-\infty}^{\tilde{\mu} - \tilde{\sigma}} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2} - 1 \right] F \left( \frac{q - \mu}{\sigma_v} \right) F \left( \frac{q - \mu}{\sigma_v} \right) d\mu \leq 0,
\]

which implies that

\[
\int_{-\infty}^{\tilde{\mu}} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2} - 1 \right] F \left( \frac{q - \mu}{\sigma_v} \right) F \left( \frac{q - \mu}{\sigma_v} \right) d\mu \leq 0. \tag{A-17}
\]

Similarly, if \( \tilde{\sigma} \geq 2(q - \tilde{\mu}) \), then \( F \left( \frac{q - \mu}{\sigma_v} \right) F \left( \frac{q - \mu}{\sigma_v} \right) \geq F \left( \frac{q - (\tilde{\mu} + \tilde{\sigma})}{\sigma_v} \right) F \left( \frac{q - (\tilde{\mu} + \tilde{\sigma})}{\sigma_v} \right) \) for any \( \mu \in [\tilde{\mu}, \tilde{\mu} + \tilde{\sigma}] \), and hence

\[
\int_{\tilde{\mu} + \tilde{\sigma}}^{\tilde{\mu}} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ 1 - \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2} \right] F \left( \frac{q - \mu}{\sigma_v} \right) F \left( \frac{q - \mu}{\sigma_v} \right) d\mu \geq 0,
\]

which implies that

\[
\int_{\tilde{\mu}}^{\infty} f \left( \frac{\mu - \tilde{\mu}}{\tilde{\sigma}} \right) \left[ \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2} - 1 \right] F \left( \frac{q - \mu}{\sigma_v} \right) F \left( \frac{q - \mu}{\sigma_v} \right) d\mu \leq 0, \tag{A-18}
\]
which together with Equation (A-17) means that

\[
\int_{-\infty}^{\infty} f \left( \frac{\mu - \bar{\mu}}{\bar{\sigma}} \right) \left[ \frac{(\mu - \bar{\mu})^2}{\bar{\sigma}^2} - 1 \right] F \left( \frac{q - \mu}{\sigma_\nu} \right) \bar{F} \left( \frac{q - \mu}{\sigma_\nu} \right) d\mu \leq 0,
\]

(A-19)
establishing the second part of Proposition 7.

\[\Box\]