CESifo WORKING PAPERS

6759
2017

November 2017

Competition in Public Service Provision: The Role of Not-for-profit Providers

Timothy Besley, James M. Malcomson

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6759*
Category 1: Public Finance

Competition in Public Service Provision: The Role of Not-for-profit Providers

Abstract

With public services such as health and education, it is not straightforward for consumers to assess the quality of provision. Many such services are provided by monopoly not-for-profit providers and there is concern that for-profit providers may increase profit at the expense of quality. This paper explores whether entry by for-profit providers is good for consumers despite the problem of unobserved quality. The model generates three key policy-relevant insights. First, by developing a novel approach to competition between different organizational forms, it frames the relevant trade-offs precisely. Second, it shows the value of keeping an incumbent not-for-profit as an active provider. Third, it characterizes the optimal payment (or voucher value) to an entrant for each consumer who switches in a way that can be applied empirically.

JEL-Codes: H110, H440, L210, L310.

Keywords: public services, competition, not-for-profit providers.

Timothy Besley
London School of Economics and Political Science, Department of Economics
United Kingdom - WC2A 2AE London
T.Besley@lse.ac.uk

James M. Malcomson
University of Oxford
All Souls College
United Kingdom - Oxford OX1 4AL
james.malcomson@economics.ox.ac.uk

November 3, 2017
We are grateful to Nemanja Antic, Tomer Blumkin, Kurt Brekke, Martin Chalkley, Steve Coate, Marty Gaynor, Julian Le Grand, Carol Propper, Luigi Siciliani and participants in the Economic Theory Workshop at Nuffield College, Oxford, the Economics of Social Sector Organizations conference at Chicago Booth, and the CESifo Area Conference on Public Sector Economics for helpful comments and discussion.

1 Introduction

When it comes to public services such as education and health care, much of the economy is run by not-for-profit providers. There is widespread suspicion that quality of public services (some aspects of which are not easily observed) will suffer if supplied by for-profit providers, even if they are more cost efficient, which can be avoided by having those services supplied by monopoly state-funded not-for-profit providers. This would be fine except that such providers can be problematic not least because, as monopoly providers, they have little incentive to be responsive to customer needs. There may, moreover, be potential entrants who can provide the service at lower cost. These concerns notwithstanding, allowing a greater role for for-profit provision is among the most controversial proposals in public service reform.

This paper provides a window on this policy debate by exploring whether entry by for-profit providers is good for consumers in a world where the principal drawback from for-profit provision is a failure to provide an unobserved dimension of quality. This drawback is mitigated by using a not-for-profit provider even if the latter does not necessarily act in the best interest of consumers. Entry by a for-profit provider nevertheless guarantees that consumers are better off, despite the unobserved quality, provided that the not-for-profit incumbent is retained as an active provider — the for-profit provider supplies markedly (not just marginally) higher observed quality to offset lower unobserved quality. Keeping the not-for-profit incumbent active also ensures that consumers who do not switch to an entrant do not lose out. But a for-profit entrant competing with a not-for-profit incumbent needs a greater cost advantage for entry to be worthwhile than if the incumbent were for-profit. This creates a trade-off from retaining the not-for-profit incumbent: it ensures greater benefit to consumers if entry occurs but with a lower probability of benefit-increasing entry. Entry by another not-for-profit provider, however, can occur with a smaller cost advantage relative to the not-for-profit incumbent, which may be why much competition in education and health services is by not-for-profit providers.

Having explored the potential for for-profit provision to benefit consumers, we explore whether it is optimal for the government to set a capitation fee (in effect a voucher) that discriminates for or against the incumbent. In general, it is not optimal for there to be a "level playing field" with the same capitation fee for both because the probability of entry is endogenous to the fee that is set. We give conditions for the capitation fee for the entrant to be above or below that for the incumbent and also show that the factors that go into this formula can in principal be measured.

The analysis of this paper generates three key policy-relevant insights. First, by developing a novel approach to competition between different organizational forms, in particular what happens when a for-profit and not-for-profit compete, it frames the relevant trade-offs precisely. Second, it shows that the value of keeping an incumbent not-for-profit active as a competitor is relevant for policy discussions about opening up public services to competition. Third, it characterizes the optimal level of the fee or voucher value to an entrant for each switching consumer in a way that can be applied empirically.

The rest of the paper is organized as follows. In the next section, we discuss related literature. Section 3 introduces the core modeling framework. It also sets up the monopoly benchmark and motivates the role for not-for-profit provision in that framework. Section 4 allows entry and studies competition with different provider objectives. Section 5 develops the analysis of optimal funding, including the optimal capitation fee or voucher that should be offered for consumers who move to an entrant. Section 6 concludes. Appendix A contains proofs of propositions. Appendix B shows that the main results are robust to allowing for a more general objective function for not-for-profit firms, a continuous distribution of switching costs/benefits for consumers and multiple quality dimensions.

2 Related Literature

The model of not-for-profit provision we use draws on two established approaches. From Newhouse (1970), we use the idea that not-for-profit providers have a bias towards quality relative to for-profit providers and, following Hansmann (1980), we acknowledge the importance of the difficulties in measuring quality in understanding why firms choose not-for-profit status. Non-contractibility of quality is used in Glaeser and Shleifer (2001) and lies behind the core trade-offs uncovered in Hart et al (1997). The key point is that there is a potential cost-quality trade-off. The literature has observed that the trade-off can be mitigated by employing motivated agents who care directly about quality, as in Besley and Ghatak (2001), Francois (2007), Francois and Vlassopoulos (2008) and Ghatak and Mueller (2011).

There are obvious difficulties in comparing the extent to which for-profit and not-for-profit providers differ in delivery of quality dimensions unobserved by customers — such quality dimensions are typically unobserved by researchers too. Sloan (2000) assesses the reasons, including difficulties in measuring quality among others, why not-for-profit provision is dominant in US healthcare. More recently, Herrera et al (2014) provide an overview of the findings of systematic reviews of differences between for-profit and not-for-profit hospitals in a variety of settings. While expressing concern about the methodological quality of many of the systematic reviews, they still conclude that, among private providers, for-profits have significantly higher mortality rates than not-for-profits. While differences in mortality rates are observable *ex post*, they may well be something patients are unaware of at the time of choosing where to go for treatment.

Here we study the impact of competition when not-for-profit providers have the characteristics emphasized by Newhouse (1970) and Hansmann (1980). The role of competition in public service provision has been discussed in Le Grand (2007). Hoxby (1999) has discussed some formal models of how competition can matter. Lakdawalla and Philipson (2006) also discusses competition with a not-for-profit provider. In that model, a not-for-profit differs from a for-profit only in having the quantity it provides as an argument in its objective function in addition to, and separate from, its role in generating profit. Only because charitable donations enable it to operate at a loss can it indulge its own preferences relative to a for-profit provider with the same cost function. Quality of service does not enter the model. More recently, Laine and Ma (2016) include quality of service in their model of competition between public and private firms. Their public firms, however, are assumed to maximize social surplus, which makes them very different from the not-for-profit providers in Newhouse (1970) that have their own self interests.

The analysis of competition and entry in education is extensive. In its early incarnation, the focus was on competition between jurisdictions with population mobility. However, in recent years interest has been fuelled in large measure by the US charter school experiment allowing entry of schools to compete against public providers. The latter has been taken up in a range of countries including Sweden and the UK. There is now a large theoretical and empirical literature on the role of competition in improving the performance of schools. From the theoretical side, there are contributions by Barseghyan et al (2014), Epple and Romano (1998) and McMillan (2005). Empirical studies of the impact of school competition include Card et al (2010), Hoxby (2003), Lavy (2008) and Gibbons et al (2008). However, as yet there is no canonical theoretical approach to entry in competition with public providers that takes into account of the possibility of strategic interaction between them.

The paper is also related to the large literature on school vouchers (see Ladd (2002) and Neil (2002) for reviews) following the early advocacy of the idea by Friedman (1962). Standard models, such as Nechyba (2000), look at the possibility that a citizen can carry their public funding to another provider. Böhlmark and Lindahl (2015) evaluate Sweden's school voucher system arguing that increased school competition enhanced standards. The debate about the value of voucher systems has typically centred on changes in quality and/or the gains from competition. Here we raise an additional issue — whether vouchers should be more or less generous than the capitation fee given to incumbents — and show that, because quality may not be optimal in the first place, there may be a case for either more or less generous funding of entrants relative to incumbents.

How to ensure service quality is also a major focus of the literature on health care, with significant implications for public provision of health services, see Chalkley and Malcomson (2000). The growing literature on the effects of competition on quality in provision of health services is reviewed in Gaynor et al (2015). The models of quality determination by providers reviewed there focus on a single quality dimension observed by customers, so again there is not the underlying rationale for not-for-profit providers emphasized in Hansmann (1980), and on monopolistic competition, in which there is no strategic interaction between providers. Absence of strategic interaction is appropriate when there is a large number of competitors, none of which impact more on one rival than on another. In our setting, which begins with a status quo of a monopoly state-funded incumbent, taking account of strategic interaction is unavoidable. Brekke et al (2011) and Brekke et al (2012) do that in studying the effect of competition on quality with not-for-profit providers modeled as caring about consumer benefits in addition to profit. Brekke et al (2014) consider how patient mobility affects provision of health care when governments make quality investment decisions to maximize welfare and study the important question of how this depends on transfer payments when patients shop around. In these models, however, quality has a single dimension observable by consumers, so here also there is not the underlying rationale for not-for-profit providers emphasized in Hansmann (1980).

The health literature also provides evidence that significant numbers of patients really do switch providers in response to competition, see Chandra et al (2016) for the US and Gaynor et al (2012) for the UK. Bloom et al (2015) argue that competition between UK public hospitals increases the quality of management practices but do not study competition between for-profit and not-for-profit hospitals.

3 The Model

Set-up The basic model considers provision of a public service for which there is a single incumbent provider, denoted by I, and a single potential entrant, denoted by E. The service has two dimensions of quality in amounts $q, Q \geq 0$, neither of which is contractible. The difference between them is that providers can commit only to q before consumers, having observed q, choose which provider to use. In contrast, providers choose Q only *after* consumers have chosen where to consume. (An equally good alternative would be that consumers are unable to observe Q before experiencing it.) We refer to Q as *unobservable* quality. That dimensions of quality are unobserved motivates the value of not-for-profit provision in this setting since for-profit firms have no incentive to provide such quality.

Revenue per customer is fixed at $p_i \geq 0$, for $i \in \{I, E\}$, and is funded from taxation. In the case of entry in education, it can be thought of as a voucher which a consumer can use to spend the per capita cost of provision with an entrant instead of with the incumbent. In the case of health, it corresponds to the payment under Medicare per patient in a given *diagnosis related group* and to the *payment by results* for specific treatments under the British National Health Service. Thus, the model is one of decentralized service provision with centralized finance.[1]

Unobserved dimensions of quality are a characteristic feature of many public services. While a parent may be able to see what is on the curriculum that they choose for their child, whether the teachers are enthusiastic and/or knowledgeable in the subject that they teach cannot be observed ex ante. Similarly, a patient choosing a hospital may observe the level of cleanliness and even the track-record of the surgeons but will find it difficult to assess what efforts are put into patient aftercare and "softer" aspects of care such as bedside manner. Finally, someone who receives legal counsel funded by the state can see what the qualifications of the lawyer are but not how much time is set aside for such activities and whether it is simply viewed as a chore by those assigned to such work.

There is a continuum of consumers of the public service, each of whom consumes at most one unit and from that receives utility $Q + q$.[2] In the basic model, a proportion $1 - \gamma$ of consumers are rigid in the sense of always choosing the incumbent provider as long as it offers utility of at least zero, whereas the remainder are flexible, choosing whichever of the available providers' quality bundles yields the higher utility.

Allowing for rigid consumers is a non-standard, but realistic, feature of the set-up. Many markets for public services opened up to competition have seen quite limited take-up. While this could be interpreted as consumers being content with the service they are provided, it is also interpreted as inertia. Inertia could be due to real costs, as when a patient must travel to receive medical treatment. It could also be psychological, with consumers simply unwilling to explore alternatives even when it is in their interest to do so. Having two groups of consumers in our core model simplifies the exposition. However, in Appendix B, we show that the insights hold in a more general model with a continuum of switching costs and multiple dimensions of quality.

The cost of providing a unit of the service is $[c(Q) + c(q)]/\theta_i$, where $c(\cdot)$ is strictly increasing and strictly convex with $c(0) = c'(0) = 0$ and $\theta_i \in [\underline{\theta}, \bar{\theta}]$ for $i \in \{I, E\}$ an efficiency parameter that can differ between providers.[3]

We assume that θ_I is known to policy makers and to any potential entrant. The entrant's cost θ_E is drawn from a distribution $G(\theta_E)$ with support $[\underline{\theta}, \bar{\theta}]$ and continuous density $g(\theta_E)$. This distribution captures uncertainty about the costs of potential entrants. We let $x_i \in [0, 1]$ denote provider i's market share.

For-Profit Provision A for-profit provider's objective is to maximize

$$\{p_i - [c(Q) + c(q)]/\theta_i\}\, x_i,$$

[1]The model could straightforwardly be extended to allow for a regulated user fee.

[2]Having consumers homogeneous in their tastes for quality serves to highlight the role of competition *per se*, as distinct from more providers increasing welfare simply by offering greater variety.

[3]A fixed cost per consumer that is independent of quality can be deducted from revenue in specifying p_i. Appendix B shows that that the core results hold without making the cost function additive and identical for the two kinds of quality.

i.e. the revenue per customer served less the cost of provision, multiplied by its market share. Consider consumers with best outside option of U (which could be not consuming at all in which case $U = 0$). With Q set only after consumers choose a provider, a for-profit provider will set $Q = 0$, i.e. it always provides the lowest level of unobserved quality. Then observed quality is set equal to utility, i.e.

$$q = U. \tag{1}$$

Profit per consumer at this utility level is therefore

$$v_i^{FP}(U, \theta_i, p_i) = p_i - \frac{c(U)}{\theta_i}. \tag{2}$$

For the analysis that follows, it is useful to use (2) to define $\tilde{U}_i^{FP}(\theta_i, p_i)$ by

$$p_i = \frac{c\left(\tilde{U}_i^{FP}(\theta_i, p_i)\right)}{\theta_i} \tag{3}$$

as the highest utility a for-profit provider with efficiency parameter θ_i is able to deliver without making a loss. This plays a key role in the entry analysis below.

Not-for-profit Provision We focus on a specific model of not-for-profit behavior where a provider cares about quality, its objective being a weighted sum of consumer and provider preferences, i.e.

$$[\lambda(\beta Q + q) + (1 - \lambda)(Q + q)]x_i = (\alpha Q + q)\, x_i, \tag{4}$$

where $\alpha = \lambda\beta + (1 - \lambda)$. The parameter λ is the weight a not-for-profit provider puts on its own preferences relative to those of consumers and β reflects the weight it puts on unobservable, relative to observable, quality which may differ from that of consumers. For $\lambda = 0$, $\alpha = 1$ and the provider is fully benevolent in the sense of maximizing consumer utility.

We focus throughout on the case where $\alpha > 1$ which is implied by setting $\beta > 1$. This approach captures the spirit of the classic contributions to the study of not-for-profit providers such as Newhouse (1970) and Hansmann (1980) where the provision of (unobserved) quality is the *sine qua non* of not-for-profit status. With $\beta > 1$, an increase in λ (and thus α) leads to a larger divergence between the provider's and the consumers' objectives. The model captures a key delegation problem that typifies public service provision where provider interests (for better or worse) play a key role in the way that services are provided. As we shall see below, competition can reduce the power of provider interests.[4]

[4]The model can be extended to not-for-profits with some pure managerial slack.

A not-for-profit provider must cover its costs, which gives the breakeven constraint

$$\{p_i - [c(Q) + c(q)]/\theta_i\} x_i \geq 0. \tag{5}$$

This rules out the possibility that it receives donations to support its activities over and above the publicly funded capitation fee.[5] Since it cares directly about both kinds of quality, it chooses values of $\{Q, q\}$ to maximize (4) subject to the breakeven constraint (5) and to offering sufficient utility to attract consumers. The first-order conditions for its quality choices for given market share $x_i > 0$ are

$$\alpha\theta_i - \mu c'(Q_i) = 0 \qquad \text{and} \qquad \theta_i - \mu c'(q_i) = 0, \tag{6}$$

where μ is a Lagrange multiplier on the breakeven constraint (5). Denote the solution by $\{Q^*(\alpha, \theta_i, p_i), q^*(\alpha, \theta_i, p_i)\}$ and let

$$U^*(\alpha, \theta_i, p_i) = Q^*(\alpha, \theta_i, p_i) + q^*(\alpha, \theta_i, p_i) \tag{7}$$

denote the resulting level of consumer utility.[6] When the best outside option for consumers U satisfies $U \leq U^*(\alpha, \theta_i, p_i)$, as is the case when there is no entry so $U = 0$, this is the optimal solution. Otherwise, the optimal solution is fully determined by the solution with $Q \geq q$ to the binding utility and breakeven constraints and, hence, the following pair of conditions

$$\hat{Q}^*(U, \theta_i, p_i) = U - \hat{q}^*(U, \theta_i, p_i) \tag{8}$$

$$p_i = \frac{c(U - \hat{q}^*(U, \theta_i, p_i)) + c(\hat{q}^*(U, \theta_i, p_i))}{\theta_i}. \tag{9}$$

Note that $\hat{Q}^*$ is strictly positive and depends on U but is independent of α.

Analogous to what we had for a for-profit, define $\tilde{U}_i^{NP}(\theta_i, p_i)$ by

$$p_i = \frac{2c\left(\tilde{U}_i^{NP}(\theta_i, p_i)/2\right)}{\theta_i} \tag{10}$$

as the highest utility a not-for-profit provider with efficiency parameter θ_i can feasibly deliver given the breakeven constraint, i.e. where $q = Q$ as desired by consumers. For the same efficiency θ_i, it is immediate that $\tilde{U}_i^{NP}(\theta_i, p_i) > \tilde{U}_i^{FP}(\theta_i, p_i)$ because the cost function is strictly convex and the not-for-profit provider provides both types of quality. Because a not-for-profit provider's preferences ensure that it delivers positive unobservable quality, it enjoys an effective cost advantage.

[5]It is straightforward to allow this possibility which is considered by Lakdawalla and Philipson (2006).

[6]Both $Q^*(\alpha, \theta_i, p_i)$ and $q^*(\alpha, \theta_i, p_i)$ are unique because $c(.)$ is strictly convex and are strictly positive because $c'(0) = 0$.

A not-for-profit provider's payoff per consumer served is

$$v_i^{NP}\left(U,\theta_i,p_i\right) = \begin{cases} \alpha Q^*\left(\alpha,\theta_i,p_i\right)+q^*\left(\alpha,\theta_i,p_i\right)\ , & \text{if } U\in[0,U^*(\alpha,\theta_i,p_i)]; \\ \alpha \hat{Q}^*\left(U,\theta_i,p_i\right)+\hat{q}^*\left(U,\theta_i,p_i\right)\ , & \text{if } U\in\left[U^*\left(\alpha,\theta_i,p_i\right),\tilde{U}_i^{NP}(\theta_i,p_i)\right]; \\ 0\ , & \text{otherwise.} \end{cases} \quad (11)$$

It is straightforward to check that, for $U\in\left[U^*\left(\alpha,\theta_i,p_i\right),\tilde{U}_i^{NP}(\theta_i,p_i)\right]$, v_i^{NP} is decreasing in U and everywhere non-negative, which implies that a not-for-profit provider will always wish to be active in the market.

Monopoly Benchmark The benchmark for our exercise is an incumbent monopoly provider. Then the only outside option for consumers is not consuming, for which $U=0$. In practice, monopoly public service provision uses not-for-profit provision and the following result makes clear why this is the case in our model.[7]

Proposition 1 *With a monopoly provider, the utility it offers consumers is $u^{FP}\left(\theta_I,p_I\right)=0$ for all $\left(\theta_I,p_I\right)$ if it is for-profit and $u^{NP}\left(\theta_I,p_I\right)=U^*\left(\alpha,\theta_I,p_I\right)>0$ for all $\left(\theta_I,p_I\right)$ if it is not-for-profit. The utility $U^*\left(\alpha,\theta_I,p_I\right)$ is increasing in θ_I and p_I and decreasing in α.*

A monopoly for-profit provider's only interest is to minimize the cost of provision, so it offers only the lowest utility for which consumers will seek provision, normalized as zero, whatever θ_I and p_I are. Thus consumers receive only their reservation payoff and so get no gain from having the service provided. In contrast, a monopoly not-for-profit provider offers consumer utility $U^*\left(\alpha,\theta_I,p_I\right)$ defined in (7) which is strictly greater than zero for all θ_I and p_I.

An implication of Proposition 1 is that consumers are always better off with a monopoly not-for-profit provider, no matter how inefficient, than with a monopoly for-profit provider, no matter how efficient.[8]

Proposition 1 also implies that neither increasing funding (higher p_I) nor having a more efficient provider (higher θ_I) makes a difference to the quality of service supplied by a monopoly for-profit provider. In contrast, both are unambiguously better for consumers with a monopoly not-for-profit provider because they allow more of both kinds of quality to be provided. But not-for-profit provision does not maximize consumer utility for given funding because, with $\alpha>1$, there is non-alignment between the provider's and consumers' objectives.

[7] All proofs are in Appendix A.

[8] Although not included formally in our model, this result carries over straightforwardly to the case in which there is a fixed cost of provision independent of the number of consumers served. If that fixed cost is sufficiently large that the market can sustain only one provider, it is always better for consumers that this is a not-for-profit provider. This has a direct policy implication. If a community is too small to sustain more than one school or hospital, it is better for consumers that the school or hospital is not-for-profit, as historically the case in many places.

Finally, Proposition 1 shows that consumers will actually be *worse off* if an incumbent not-for-profit either cares more about provider objectives (higher λ) and/or its bias towards quality Q is greater (higher β), either of which implies higher α. This is an important distortion that motivates a role for competition beyond achieving cost-efficiency. The rents earned by monopoly not-for-profit providers are decision rents due to their ability to determine the mix of qualities they prefer.

4 Entry

In this section, we explore entry when the starting point is not-for-profit provision. This is motivated by our observation that, with monopoly, not-for-provision is always best. However, the possibility of paternalistic preferences means that the status quo is not necessarily best from the consumers' point of view. Entry therefore serves two possible roles. First, an entrant may be more efficient (have high θ). Second, an entrant may deliver an outcome that is closer to what (flexible) consumers want.

The timing is as follows:

1. Nature determines the efficiency of the potential entrant $\theta_E \in [\underline{\theta}, \bar{\theta}]$ which, as well as being revealed to the potential entrant, is revealed to consumers and the incumbent along with the entrant's type (for-profit or not-for-profit).
2. The potential entrant decides whether to enter and, if it decides to do so, chooses q_E, which is observed by consumers and the incumbent. (If the entrant anticipates the same equilibrium payoff from entering as from not entering, it chooses to enter if and only if it actually attracts some consumers.)
3. The incumbent chooses q_I, which is observed by consumers.
4. Consumers choose whether to consume and if so where with, for simplicity, indifferent flexible consumers choosing the entrant.
5. Provider I chooses Q_I and provider E, if entered, chooses Q_E.

We solve for a subgame perfect equilibrium.[9]

4.1 The Logic of Entry

For each organizational form $j \in \{FP, NP\}$, $u^j(\theta_i, p_i)$ specified in Proposition 1 for $i \in \{I, E\}$ is the utility to consumers delivered by a type j provider if not constrained

[9]Formally, we solve for a subgame perfect ε-equilibrium because there are subgames for which there does not exist a strictly *best* response. However, the only subgames for which that is the case are ones not reached along the equilibrium path, so the equilibrium payoffs in Proposition 2 are exact. We discuss below the alternative timing in which the incumbent decides quality before the entrant.

by competition. When deciding whether to accommodate entry, the incumbent must decide whether to allow the entrant to serve the flexible consumers. Whether it does so depends on the difference in its payoff, whether from profits or as a not-for-profit, from serving the whole market compared to serving only the proportion $(1-\gamma)$ consisting of rigid consumers. To serve the whole market, it must offer all consumers utility that matches the utility offered by the entrant. But if it seeks to retain only the rigid consumers, it can do that by offering just $u^j(\theta_I, p_I)$ and thus receive a higher payoff per consumer served. There is thus a critical proportion of flexible consumers that makes it optimal to compete for them.

Formally, the incumbent's payoff is $(1-\gamma)\, v_I^j(u^j(\theta_I, p_I), \theta_I, p_I)$ when serving only the rigid consumers. Its payoff when serving the whole market at utility level U determined by the entrant's offer is $v_I^j(U, \theta_I, p_I)$. The critical value of γ below which the incumbent prefers to serve only the rigid consumers is $\widehat{\gamma}^j(U, \theta_I, p_I)$ defined by

$$\widehat{\gamma}^j(U, \theta_I, p_I) = \begin{cases} 1, & \text{if } \; v_I^j(U, \theta_I, p_I) < 0; \\ 1 - \frac{v_I^j(U,\theta_I,p_I)}{v_I^j(u^j(\theta_I,p_I),\theta_I,p_I)}, & \text{if } \; 0 \leq v_I^j(U, \theta_I, p_I) < v_I^j\left(u^j(\theta_I, p_I), \theta_I, p_I\right); \\ 0, & \text{if } \; v_I^j(U, \theta_I, p_I) \geq v_I^j(u^j(\theta_I, p_I), \theta_I, p_I). \end{cases} \tag{12}$$

That is, if $\gamma < \widehat{\gamma}^j(U, \theta_I, p_I)$, there are too few flexible consumers for it to be worth the incumbent competing for them by offering the payoff U. The top and bottom cases in (12) are corner solutions where either the incumbent never finds it worthwhile to compete (top case) or always retains the flexible consumers (bottom case). As U increases, the critical value of $\widehat{\gamma}^j(U, \theta_I, p_I)$ increases and the incumbent is in a weaker position to compete. Define $\bar{U}^j(\gamma, \theta_I, p_I)$ by

$$\gamma = \widehat{\gamma}^j(\bar{U}^j(\gamma, \theta_I, p_I), \theta_I, p_I) \tag{13}$$

as the highest utility the incumbent is willing to offer to retain the flexible consumers. Note that $\bar{U}^j(\gamma, \theta_I, p_I) > u^j(\theta_I, p_I)$ because the incumbent is always willing to give up a small amount of payoff per consumer served to acquire the discrete proportion γ of flexible consumers.

The next proposition gives a necessary and sufficient condition for entry and specifies how consumers fare with and without entry. Recall that $\tilde{U}_i^j(\theta_i, p_i)$, for $i \in \{I, E\}$, is the highest utility provider type j can provide without making a loss.

Proposition 2 *Entry by type k occurs with incumbent type j, for $j, k \in \{FP, NP\}$, if and only if*

$$\gamma \leq \widehat{\gamma}^j(\tilde{U}_E^k(\theta_E, p_E), \theta_I, p_I). \tag{14}$$

If no entry occurs, payoffs for both rigid and flexible consumers are $u^j(\theta_I, p_I)$. If entry occurs, *rigid consumer payoffs are $u^j(\theta_I, p_I)$ and flexible consumer payoffs are as follows:*

$$\begin{array}{ll} \max\left\{\tilde{U}_I^j(\theta_I, p_I), u^k(\theta_E, p_E)\right\}, & \text{if } \gamma \geq \widehat{\gamma}^j\left(\tilde{U}_I^j(\theta_I, p_I), \theta_I, p_I\right); \\ \max\left\{\bar{U}^j(\gamma, \theta_I, p_I), u^k(\theta_E, p_E)\right\}, & \text{otherwise.} \end{array} \tag{15}$$

Entry strictly increases the utility of flexible consumers while leaving the utility of rigid consumers unchanged.[10]

This result applies for all possible organizational forms and efficiency levels for the incumbent and entrant. To understand it, note that $\tilde{U}_E^k(\theta_E, p_E)$ determines how hard the potential entrant can compete for flexible consumers since it is the highest level of utility that it can offer them and still be worth entering. The key issue is whether the proportion of flexible consumers γ is greater than $\widehat{\gamma}^j(\tilde{U}_E^k(\theta_E, p_E), \theta_I, p_I)$. If it is, there is no entry because it is worthwhile for the incumbent to compete and retain the flexible consumers by offering them more than the highest utility the potential entrant can afford to offer. In this case, the potential entrant would be unable to capture any of the market and would not enter. If γ is below $\widehat{\gamma}^j(\tilde{U}_E^k(\theta_E, p_E), \theta_I, p_I)$ (condition (14)), the entrant can attract the flexible consumers. But it is worth entering to do that only if it has a positive payoff. This is the case for $U \leq \tilde{U}_E^k(\theta_E, p_E)$. Hence this condition is also sufficient for entry. The second part of the proposition shows how consumers of different types fare with entry. Rigid consumers neither gain nor lose with entry because entry occurs only if the entrant can successfully attract the flexible consumers and, in that case, the incumbent has no reason to respond by offering the rigid consumers anything other than what it would offer in the absence of entry. However, flexible consumers gain whenever there is entry, despite the unobservable quality dimension, because the entrant has to offer a higher utility to them to make it unattractive for the incumbent to more than match that offer.[11]

4.2 For-profit Entry

Proposition 2 shows that, whatever the organizational forms of the incumbent and the entrant, flexible consumers gain, and rigid consumers do not lose, with entry despite the unobserved quality dimension. But organizational form has an impact both on when entry occurs and on the size of the gains when it does occur. We consider first the politically controversial case in which a for-profit entrant is permitted to compete

[10] If the ordering of Stages 2 and 3 in the timing of the entry game are reversed so that the incumbent moves before the entrant, the necessary and sufficient condition for entry (14) is unchanged. The threat of entry (without actual entry) can then increase the payoffs to both rigid and flexible consumers. But the equilibrium has the unappealing characteristic that neither type of consumer necessarily gains when entry actually occurs.

[11] The welfare results do not go through if there is either an internality whereby flexible consumers do not know their own true welfare or an externality from flexible consumers to rigid consumers as in the case of peer group effects.

with a not-for-profit incumbent. We do this for the case in which the funding level is the same for both incumbent and entrant, which applies, for example, to payment by diagnosis related group under US Medicare or payment by results in the British National Health Service. For this section, therefore, $p_E = p_I = p$.[12]

Proposition 3 *When a not-for-profit incumbent competes with a for-profit potential entrant, a necessary condition for entry is that $\theta_E > \theta_I$. For flexible consumers, entry increases observed quality but reduces unobserved quality to the minimal level.*

This says that $\theta_E > \theta_I$ is necessary for entry. For $\theta_E \leq \theta_I$, the utility provided by a not-for-profit incumbent in the absence of entry, $U^* (\alpha, \theta_I, p)$, is greater than the highest utility a for-profit entrant with the same efficiency parameter can profitably provide, i.e. $\tilde{U}_E^{FP} (\theta_E, p)$. That reflects the not-for-profit incumbent's provision of unobserved quality which gives it an implicit cost advantage, the strict convexity of the cost function, when competing with a for-profit entrant. A for-profit entrant can provide utility only by spending on observable quality.[13] This suggests a reason for why it is difficult to obtain effective for-profit competition in some contexts, such as US school districts or the British National Health Service, which have established not-for-profit incumbents.

For a sufficiently high entrant efficiency level, θ_E, it is infeasible for the incumbent to compete with the entrant because the incumbent cannot feasibly offer $\tilde{U}_E^{FP} (\theta_E, p)$ due to the breakeven constraint. At this point $\widehat{\gamma}^{NP}(\tilde{U}_E^{FP} (\theta_E, p), \theta_I, p) = 1$ and there is entry for all $\gamma \in [0, 1]$. Thus, a large enough entrant efficiency advantage is sufficient for entry. Because the not-for-profit incumbent sets unobservable quality above the minimal level, whereas the for-profit entrant sets it at zero, unobservable quality for flexible consumers falls with entry. However, as Proposition 2 showed, their utility increases with entry because, to attract them, the entrant must offer observed quality sufficiently high to compensate for the loss in unobserved quality. Indeed, observed quality provided by the for-profit entrant must not only be higher than that offered by the not-for-profit incumbent but discretely higher because the drop in unobserved quality is discrete.

The basic logic of the model is illustrated in Figure 1. The curve 'CC' is the zero-profit line for the incumbent that gives the combinations of (Q, q) just attainable for given θ_I and p without the incumbent incurring a loss. It moves out with higher θ_I and p. The indifference curves for consumers are downward sloping straight lines, with higher lines corresponding to higher utility. The first best qualities for consumers given θ_I and p are at point A. The incumbent prefers a higher ratio of Q to q so, when a monopolist, offers a point such as B. A for-profit entrant produces only on the q-axis. If

[12]We consider below whether differentiating the payment between the incumbent and entrant is optimal.

[13]This is different from Lakdawalla and Philipson (2006) where the cost advantage of a not-for-profit comes from its access to donations.

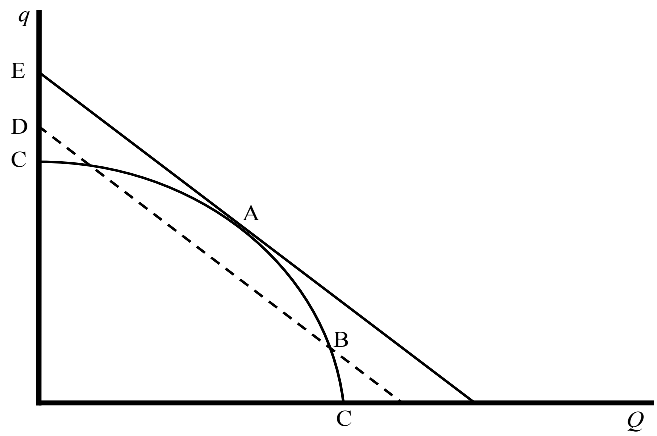

Figure 1: Illustration of Entry Condition

its zero-profit amount of q is below D, it cannot match the utility the monopoly incumbent offers at B without making a loss and so will not enter. If its zero-profit amount of q is above E, it can offer higher utility than the incumbent can afford at A and still make a profit, and so will certainly enter. If its zero-profit amount of q is between D and E, whether it enters depends on whether it is better for the incumbent to compete for the flexible consumers or to serve just the rigid ones at B. It is clear from this diagram why a for-profit entrant needs θ_E strictly greater than θ_I — it must be able to reach D without making a loss, whereas the incumbent can reach only C on the q-axis. Figure 1 also illustrates that this basic logic does not depend on the particular functional forms we have used for consumer utility and for the cost of providing quality. It applies as long as, for given efficiency, it is less costly to provide utility with strictly higher unobserved quality than a for-profit provides.

We now look at how the cost difference and the extent of the incumbent's paternalism affect the likelihood of entry.

Proposition 4 *When a not-for-profit incumbent competes with a for-profit potential entrant, the critical value of γ at which entry occurs is increasing in α and θ_E whenever it is less than one. Moreover, if there is entry when $\tilde{U}_E^{FP}(\theta_E, p) < \tilde{U}_I^{NP}(\theta_I, p)$, the utility of flexible consumers is increasing in α.*

A higher critical value of γ increases the range of γ for which entry takes place. Proposition 4 thus implies that a more efficient entrant (higher θ_E) and a more paternalistic incumbent (higher α) increase the probability of entry. These are intuitive. When θ_E is higher, the entrant can afford a more aggressive offer to the flexible consumers in order to attract them. When the incumbent is more paternalistic, it would lean towards serving only the rigid consumers rather than compromising and serving the flexible consumers when the entrant tries to attract them.

4.3 Varying Incumbent and Entrant Motives

We next consider what happens when the incumbent and entrant are both for-profit or both not-for-profit. Not only do these widen the comparisons, they also allow us to appreciate better why having not-for-profit provision makes a difference when there is competition. We again do this for equal funding to incumbent and entrant so, for this section also, $p_E = p_I = p$.

Here we have the following result:

Proposition 5 1. ***Incumbent and potential entrant both for-profit.***

$$\hat{\gamma}^{FP}\left(\tilde{U}_E^{FP}(\theta_E, p), \theta_I, p\right) = \min\left\{\frac{\theta_E}{\theta_I}, 1\right\}. \tag{16}$$

A sufficient condition for entry is that $\theta_E \geq \theta_I$. For flexible consumers, entry increases observed quality but leaves unobserved quality at the minimal level.

2. ***Incumbent and potential entrant both not-for-profit.*** *A sufficient condition for entry is that $\theta_E > \theta_I$.*

To understand the implications of Proposition 5, consider first the case in which both incumbent and potential entrant are for-profit providers (Case 1). An efficiency advantage for the entrant ($\theta_E \geq \theta_I$) is then sufficient for entry. But it is not necessary. Even if $\theta_E < \theta_I$, entry is still possible if $\gamma \leq \theta_E/\theta_I$, since the incumbent may prefer to make a higher profit per consumer on just the rigid consumers than a lower profit per consumer on all consumers.[14]

The case in which both incumbent and potential entrant are not-for-profit providers (Case 2 in Proposition 5) has similarities with the case of competing for-profit providers (Case 1).[15] Specifically, entry occurs for sure if $\theta_E > \theta_I$. This is because a not-for-profit entrant's choice of observable quality effectively commits it to providing the quality bundle that maximizes consumer utility subject to its breakeven constraint if that is required to attract flexible consumers. Thus, in contrast to entry by a for-profit provider, the incumbent no longer has an implicit cost advantage from its provision of unobserved quality. As a result, for $\theta_E > \theta_I$, then $\hat{\gamma}^{NP}(\tilde{U}_E^{NP}(\theta_E, p), \theta_I, p) = 1$ and entry occurs for all γ. In effect the incumbent cannot then offer the utility that a more efficient entrant offers to flexible consumers.

[14]With $\theta_E < \theta_I$, then $\tilde{U}_E^{FP}(\theta_E, p) < \tilde{U}_I^{FP}(\theta_I, p)$, so offering the consumer utility in the top line in (15) in Proposition 2 would impose a loss on the entrant. Hence, if entry occurs, it must be that the bottom line in (15) in Proposition 2 applies. Moreover, with a for-profit entrant, $u^{FP}(\theta_E, p) = 0$. Thus the utility of the flexible consumers is $\bar{U}^{FP}(\gamma, \theta_I, p)$ defined in (13). This can be evaluated by equating $1-\gamma$ times the incumbent payoff in (2) for $U = 0$ to the incumbent payoff in (2) for $U = \bar{U}^{FP}(\gamma, \theta_I, p)$ to give $\bar{U}^{FP}(\gamma, \theta_I, p) = c^{-1}(\gamma p \theta_I)$.

[15]It is straightforward to derive expressions for $\hat{\gamma}^{NP}(\tilde{U}_E^k(\theta_E, p), \theta_I, p)$ for $k \in \{FP, NP\}$ for the case in which the incumbent is a not-for-profit provider but this offers little additional insight and so is not included in the proposition.

In Case 2 in Proposition 5, entry is also possible with $\theta_E \leq \theta_I$ if γ is low enough. This can be seen from Figure 1 where a not-for-profit entrant, even with θ_E slightly lower than θ_I, can offer a quality combination above the indifference curve through point B, which is why it does not have to have higher $\theta_E > \theta_I$ to enter. Comparing this with Proposition 3 shows that there is a range of θ_E for which there is no entry with a for-profit entrant while there is entry with a not-for-profit entrant. Thus entry can occur with a lower entrant efficiency advantage if the entrant is not-for-profit than if it is for-profit. Another way to think about this is that the range of γ for which there is entry when the entrant, as well as the incumbent, is not-for-profit is strictly wider than when the entrant is for-profit. This stems from the not-for-profit entrant's similar implicit cost advantage to the incumbent because its preferences ensure that it delivers positive unobservable quality. It is consistent with much competition in education and health services in practice being by not-for-profit providers.

When, with competition between two not-for-profit providers (Case 2 in Proposition 5), entry occurs with $\theta_E \leq \theta_I$, it is purely *paternalism induced*, that is, it occurs only because of preference divergence between the incumbent provider and consumers and not because of any cost advantage. To illustrate this formally, consider the case of $\theta_E = \theta_I$, in which case neither incumbent nor entrant has an inherent cost advantage. We know from Proposition 3 that a for-profit provider never enters in this case, so entry is possible only if there is a not-for-profit entrant. If α were equal to 1, the incumbent would, even without entry, always make the choices optimal for consumers given the breakeven constraint, so we would have $u^{NP}(\theta_I, p) = \tilde{U}_I^{NP}(\theta_I, p)$. Since $\theta_E = \theta_I$ the entrant could not offer utility greater than this to attract flexible consumers, so entry would not occur. Thus, entry can occur only if $\alpha > 1$ since, even without a cost advantage, a not-for-profit entrant can provide higher utility to flexible consumers by offering a quality mix that is closer to what consumers prefer. Provided the proportion of flexible consumers is small enough, the incumbent will prefer to serve only the rigid consumers with the monopoly quality mix than to compete to retain the flexible consumers. Moreover, $\tilde{U}_E^{NP}(\theta_E, p) = \tilde{U}_I^{NP}(\theta_I, p)$ when $\theta_E = \theta_I$. So, when flexible consumers receive payoff $\tilde{U}_I^{NP}(\theta_I, p)$, as in the upper line of (15), the paternalism of the entrant is completely undone, with flexible consumers getting their maximal utility given the productive efficiency of the provider.

The next result compares the payoffs to consumers with competition when the incumbent is a not-for-profit provider, rather than a for-profit provider. Proposition 1 showed that, in the absence of competition, consumers are always better off with a not-for-profit provider no matter how much more efficient a for-profit provider is. The following result applies when there is competition.

Proposition 6 *With competition between providers, rigid consumers receive higher utility with a not-for-profit than with a for-profit incumbent. When entry occurs, flexible consumers receive higher utility with a not-for-profit incumbent than with a for-profit incumbent of the* *same efficiency; specifically,* $\tilde{U}_I^{NP}(\theta_I, p) \geq \bar{U}^{NP}(\gamma, \theta_I, p) > \tilde{U}_I^{FP}(\theta_I, p) \geq \bar{U}^{FP}(\gamma, \theta_I, p)$ *for all* (γ, θ_I, p).

This result establishes that, provided entry occurs, all consumers have higher utility with a not-for-profit incumbent than with a for-profit incumbent of the same efficiency. Of particular interest is the result that a not-for-profit incumbent is always willing to offer higher utility to attract flexible consumers than the highest utility a for-profit incumbent of the same efficiency can afford (that is, $\bar{U}^{NP}(\gamma, \theta_I, p) > \tilde{U}_I^{FP}(\theta_I, p)$). This is because the cost function is strictly convex, so a not-for-profit (which values unobservable quality) can provide given consumer utility at lower cost than a for-profit with the same efficiency parameter.

The payoff gains in Proposition 6 from having a not-for-profit, rather than a for-profit, incumbent are conditional on entry occurring. However, it follows from Propositions 3 and 5 that a higher efficiency entrant is required for entry to occur with a not-for-profit incumbent than with a for-profit incumbent of the same efficiency. Thus entry may not occur with a not-for-profit incumbent even though it would have occurred with a for-profit incumbent of the same efficiency. In that case, flexible consumers may have lower expected utility with a not-for-profit incumbent than with a for-profit one when the probability of entry is taken into account. Rigid consumers, though, always do better with a not-for-profit incumbent, so the two types of consumers may have conflicting interests.

Proposition 6 compares consumer payoffs with for-profit and not-for-profit incumbents. Flexible consumers may also receive higher utility from having a not-for-profit entrant than from having a for-profit entrant with the same efficiency. This can happen in two ways. One way is that a sufficiently productive not-for-profit entrant may choose to provide utility higher than the minimum required to attract flexible consumers, as a result of which the second term in the maximum expressions in (15) exceeds the first. In contrast, a for-profit entrant never offers utility higher than the minimum required to attract flexible consumers because to do so would lower profit. The other way is because, given the strictly convex cost function and that the not-for-profit entrant provides both types of quality, the highest utility a not-for-profit entrant can afford is always strictly greater than the highest utility a for-profit entrant with the same efficiency can afford, that is, $\tilde{U}_E^{NP}(\theta_E, p) > \tilde{U}_E^{FP}(\theta_E, p)$. Thus, from (14), a not-for-profit potential entrant may enter when a for-profit potential entrant with the same efficiency would not, which is consistent with much competition in education and health services being by not-for-profit providers. In this case, there is no conflict between rigid and flexible consumers.

5 Pricing Policy

We have so far assumed that an entrant receives the same capitation fee as the incumbent. We now explore optimal payment for the public service, including whether it is optimal to pay a per capita amount to an entrant different from that to the incumbent in order to encourage or discourage entry. With standard voucher schemes for education, such as that introduced in Sweden in 1992, a consumer can transfer the public funding to the entrant.[16] However, the value of a voucher could be different from the public funding for the incumbent. Here we consider payment that is optimal from the perspective of consumers who pay taxes to fund the service with a constant marginal cost of public funds $\xi \geq 1$ and show that, in general, it is optimal to treat entrants and incumbents differently.[17] This in turn affects the probability of entry.

We look at a regulator's optimal choice of p_I and p_E from an ex ante perspective, i.e. before the efficiency of the potential entrant is known and start with the case of a not-for-profit incumbent facing a for-profit potential entrant. For given (p_I, p_E), there will, by Proposition 2, be entry if θ_E is large enough. Specifically, let $\hat{\theta}_E\left(p_E, p_I\right)$ denote the entrant efficiency level that makes the incumbent just unwilling to offer the highest consumer utility the entrant is prepared to offer to retain the flexible consumers. ($\hat{\theta}_E\left(p_E, p_I\right)$ also depends on θ_I but that is taken as given for this analysis.) This efficiency level is defined by equality in (14) for $k = FP$ with $\tilde{U}_E^{FP}(\theta_E, p_E)$ defined by (3), that is by

$$\gamma = \hat{\gamma}^j(c^{-1}(p_E\hat{\theta}_E\left(p_E, p_I\right)), \theta_I, p_I). \tag{17}$$

The probability of entry is then $1 - G\left(\hat{\theta}_E\left(p_E, p_I\right)\right)$. Also let $\hat{U}(\theta_I, \gamma, p_I)$ be the utility of a flexible consumer who switches to the entrant as given by Proposition 2. It does not depend on either θ_E or p_E because $u^k\left(\theta_E, p_E\right) = 0$ for all (θ_E, p_E) for $k = FP$, so the consumer utilities in (15) do not in this case depend on (θ_E, p_E). Ex ante expected consumer welfare for given (p_I, p_E) is then

$$\left[(1-\gamma) + \gamma G\left(\hat{\theta}_E\left(p_E, p_I\right)\right)\right]\left[U^*\left(\alpha, \theta_I, p_I\right) - \xi p_I\right] \\ + \gamma\left[1 - G\left(\hat{\theta}_E\left(p_E, p_I\right)\right)\right]\left[\hat{U}\left(\theta_I, \gamma, p_I\right) - \xi p_E\right]. \tag{18}$$

[16]The kind of voucher that we have in mind here is like that used in Sweden where no consumer-financed "top-up" is allowed.

[17]We could, as in standard models of regulation, introduce a welfare weight that values providers' payoffs, though possibly somewhat less than consumer utility. Our framework is, however, somewhat non-standard because, in the case of not-for-profit provision, provider payoffs take the form of "decision rents" rather than monetary profits. Moreover, the question of how the welfare of teachers and doctors should count in the provision of the services is moot. In political economy models, it is common to ignore the welfare of providers (politicians and bureaucrats) and simply count the welfare of voters. In the case of for-profit providers, the policy debate often proceeds as if there should be a negative weight on profit in public service provision. For example, in the UK, there is a campaign called "Public Services Not Private Profit" supported by around 14 major trade unions whose objective could be interpreted in this way, as could the objective of a lobby group such as "We Own it" https://weownit.org.uk/ whose strap line is "Public Services for People not Profit".

The regulator chooses (p_I, p_E) to maximize this. The first term in (18) is the welfare of rigid consumers plus that of flexible consumers for the entrant efficiency levels for which there is no entry, i.e. $\theta_E < \hat{\theta}_E(p_E, p_I)$, the second term the welfare of flexible consumers when $\theta_E \geq \hat{\theta}_E(p_E, p_I)$ and hence entry occurs. Changing the payments to providers has three main effects on consumer welfare in (18). Increasing funding to either the incumbent or the entrant necessitates higher taxes which reduce consumers' welfare. Counteracting this is an increase in quality. For rigid or flexible consumers who remain with the not-for-profit incumbent, this effect is direct. However, increasing p_I also affects the utility of flexible consumers who switch since their utility level is set by what the incumbent would be prepared to offer to retain them. Finally, funding arrangements change the probability of entry, i.e. the critical efficiency level at which an entrant finds it worthwhile to enter.

An important policy question is whether the per capita payment to the entrant should be the same as that to the incumbent, that is $p_E = p_I$, the so-called "level playing field". To state the results on this, let $\Delta U(p_I) = \hat{U}(\theta_I, \gamma, p_I) - U^*(\alpha, \theta_I, p_I)$ be the utility gain to a flexible consumer of switching to the entrant.[18]

Proposition 7 *Suppose that a not-for-profit incumbent faces a for-profit entrant and a policy-maker sets the per capita payment to the entrant, p_E, to maximize expected consumer welfare (18) for given per capita payment to the incumbent, p_I. Then for $g(\theta_E)$ log-concave and optimal $\hat{\theta}_E(p_E, p_I) \in [\underline{\theta}, \bar{\theta}]$, the optimal per capita payment to the entrant is the unique p_E^* that satisfies*

$$\left[\frac{\Delta U(p_I)}{\xi} + (p_I - p_E^*)\right] \frac{\hat{\theta}_E(p_E, p_I)}{p_E^*} = \frac{1 - G\left(\hat{\theta}_E(p_E^*, p_I)\right)}{g\left(\hat{\theta}_E(p_E^*, p_I)\right)}. \tag{19}$$

Equation (19) applies for any p_I, including the optimal value that maximizes expected consumer welfare (18) when the density function is log-concave.[19] In general, it implies $p_E^* \neq p_I$. To understand the implications of Proposition 7, define

$$\eta\left(\hat{\theta}_E(p_E, p_I)\right) = \frac{g\left(\hat{\theta}_E(p_E, p_I)\right) \hat{\theta}_E(p_E, p_I)}{1 - G\left(\hat{\theta}_E(p_E, p_I)\right)}. \tag{20}$$

This corresponds to the elasticity of the entry probability with respect to the payment to the entrant given the definition of $\hat{\theta}_E(p_E, p_I)$ in (17). It depends on the shape of the distribution of the potential entrant's efficiency parameter, θ_E. Rearranging (19), we

[18] Arguments other than p_I are suppressed for notational simplicity because they are given for the analysis of this section.

[19] Log-concavity of the density is satisfied by many standard probability distributions and is widely used in economic models, see Bagnoli and Bergstrom (2005). Its role in Proposition 7 is to ensure that (19) has only one solution for p_E^*. Without it, the optimal entrant payment for $\hat{\theta}_E(p_E, p_I) \in (\underline{\theta}, \bar{\theta})$ will still satisfy (19) but there may be multiple solutions to (19), so one would have to check which corresponds to a maximum of (18).

have the following formula for the optimal payment to the entrant

$$p_E^* = \frac{\eta\left(\hat{\theta}_E\left(p_E^*, p_I\right)\right)\left[\Delta U\left(p_I\right) + \xi p_I\right]}{\left[1 + \eta\left(\hat{\theta}_E\left(p_E^*, p_I\right)\right)\right]\xi}, \tag{21}$$

which also holds for *any* value of p_I. The value of the payment p_E^* is thus increasing in $\eta(\cdot)$, i.e. the more responsive is entry to a higher payment then the larger it is all else equal. The payment should also be more generous when the marginal gain to the flexible consumers from switching to the entrant, $\Delta U\left(p_I\right)$, is larger. This makes sense as entry is better for flexible consumers in this case.

An attractive feature of (21) is that it depends on magnitudes that can be specified in applications. For example suppose that the entrant efficiency parameter θ_E follows a Pareto distribution with shape parameter ζ, i.e. $G\left(\theta_E\right) = 1 - \left(\underline{\theta}/\theta_E\right)^{\zeta}$, then $\eta\left(\hat{\theta}_E\left(p_E, p_I\right)\right) = \zeta$.[20] This is motivated by noting from Axtell (2001) that the size distribution of firms suggests that productivity follows a Zipf distribution, i.e. a Pareto distribution with $\zeta = 1$. A value of $\xi = 1.5$ is a reasonable figure in line with many estimates of the cost of public funds and p_I would be known from the funding levels currently used in the market. The only additional element of (21) needed to apply the formula for policy purposes would be $\Delta U\left(p_I\right)$, i.e. the "willingness to pay" by flexible consumers to switch to the entrant.

To illustrate how to apply this formula, consider the case of hip replacement surgery in the UK. A National Health Service (NHS) provider is paid around £5000 per operation while the cost of private treatment is around £10,000. If the latter is all out of pocket, we could use it as a rough estimate $\Delta U\left(p_I\right)$ because it measures consumers' willingness to pay for the additional benefit of the private treatment. Then if $\eta = 1$ and $\xi = 1.5$, the optimal amount that the NHS should pay for a hip replacement from a private provider should an NHS patient wish to switch, is

$$p_E^* = \frac{£10,000 + £7,500}{3} \simeq £5,833.$$

So this is a case where the per capita payment to the entrant should be larger than the current per capita payment to the incumbent but less than the standard private treatment fee. These specific numbers are, of course, only illustrative but they show how Proposition 7 can be applied to real-world cases.[21]

[20]The results in Proposition 7 actually hold for $G\left(\theta\right)$ a Pareto distribution even though the density function for that distribution is not log-concave.

[21]The argument presented here can be extended to cover the case where the entrant is not-for-profit. This affects the critical $\hat{\theta}\left(p_E, p_I\right)$ but the core factors which shape optimal funding for entrants remain the same.

6 Concluding Comments

This paper argues that, while a case can be made for allowing competition by for-profit providers in public service provision despite difficulties in assessing quality, there is a benefit to retaining a not-for-profit incumbent. This is an important principle for policy design which does not seem to be widely recognized. To explain the logic of this, we have used an approach which combines the insight of Hansmann (1980) that not-for-profit providers can be valuable when there is an unobservable dimension to their output with the recognition by Newhouse (1970) that many such providers have a bias towards quality that is not solely paternalistic but also reflects producer interests.

The view that provider interests matter fits a range of services where physicians, lawyers and teachers run public services according to their views of what is good for consumers, and implies that providers earn decision rents even if with not-for-profit status. Monopoly provision with public funding does not then guarantee that consumers get what they want from public services even if incumbents provide some unobserved quality that would not be provided by profit-maximizing firms. This motivates why competition is valuable over and above considerations of cost efficiency.

Our model captures an important trade-off: while retaining a not-for-profit incumbent benefits consumers conditional on entry occurring, it reduces the probability of beneficial entry. If entry does not occur, it is better for consumers to have a not-for-profit provider. Moreover, a not-for-profit provider can enter with a lower cost advantage over the incumbent than a for-profit provider. These are consistent with many schools and hospitals in areas where there is no competition being set up as not-for-profit institutions and, where there is competition, much of it being among not-for-profit providers.

As well as exploring entry conditions, the paper has shown that offering entrants a "level playing field" (that is, the same capitation fee, or voucher of the same value, as the incumbent) is not generally optimal. Depending on ex ante market conditions, it may be optimal to pay the entrant either less or more than the incumbent. The model offers an insight into the factors determining payment that can be applied in specific situations.

There are potential downsides to competition where there are externalities or internalities that arise through consumers not knowing their own true welfare. We have not included these in our model because they are well known. But Appendix B shows that our main results are robust to allowing for a more general objective function for not-for-profit firms, a continuous distribution of switching costs/benefits for consumers and multiple quality dimensions. Our model could also be developed in a range of ways other ways. In future work, it would be interesting to enrich the analysis by looking at the endogenous choice of not-for-profit status. One could also allow for differential selection of providers' employees by competence and motivation, as in Barigozzi and Burani (2016), and how competition affects this. There is, in addition, scope to explore a range of wider contractual possibilities and regulatory approaches. But our basic insights rely essentially only on not-for-profit providers offering a level of some unobserved dimension of quality different from for-profit providers that makes it less costly, for given production efficiency, to provide given consumer utility. As long as this characteristic is retained, we would expect our fundamental insights to hold.

References

[1] Axtell, Robert L. [2001], "Zipf distribution of U.S. firm sizes," *Science*, Vol. 293, Issue 5536, 1818-1820.

[2] Bagnoli, Mark and Ted Bergstrom [2005], "Log-concave probability and its applications," *Economic Theory*, 26, 445-469.

[3] Barigozzi, Francesca and Nadia Burani [2016], "Competition and screening with motivated health professionals," *Journal of Health Economics*, 50, 358-371.

[4] Barseghyan, Levon, Damon Clark and Stephen Coate [2014], "Public school choice: an economic analysis," unpublished typescript.

[5] Besley, Timothy and Maitreesh Ghatak [2001], "Government versus private ownership of public goods," *Quarterly Journal of Economics*, 116(4), 1343-1372.

[6] Besley, Timothy and Maitreesh Ghatak [2005], "Competition and incentives with motivated agents," *American Economic Review*, 95(3), 616-636.

[7] Bloom, Nicholas, Carol Propper, Stephan Seiler and John Van Reenen [2015] "The impact of competition on management quality: Evidence from public hospitals," *Review of Economic Studies*, 82(2), 457-489.

[8] Böhlmark, Anders and Mikael Lindahl [2015], "Independent schools and long-run educational outcomes: evidence from Sweden's large-scale voucher reform," *Economica*, 82, 508-551.

[9] Brekke, Kurt R., Luigi Siciliani and Odd Rune Straume [2011], "Hospital competition and quality with regulated prices," *Scandinavian Journal of Economics*, 113(2), 444-469.

[10] Brekke, Kurt R., Luigi Siciliani and Odd Rune Straume [2012], "Quality competition with profit constraints," *Journal of Economic Behavior and Organization*, 84, 642-659.

[11] Brekke, Kurt R., Rosella Levaggi, Luigi Siciliani and Odd Rune Straume [2014], "Patient mobility, health care quality and welfare," *Journal of Economic Behavior and Organization*, 105, 140-157.

[12] Card, David, Martin D. Dooley and A. Abigail Payne [2010], "School competition and efficiency with publicly funded Catholic schools," *American Economic Journal: Applied Economics*, 2(4), 150-176.

[13] Chalkley, Martin and James M. Malcomson [2000], "Government purchasing of health services," in Anthony J. Culyer, and Joseph P. Newhouse (eds), *Handbook of Health Economics*, Amsterdam: Elsevier.

[14] Chandra, Amitabh, Amy Finkelstein, Adam Sacarny and Chad Syverson [2016], "Health care exceptionalism? Performance and allocation in the US health care sector," *American Economic Review*, 106(8), 2110-2144.

[15] Epple, Dennis and Richard Romano [1998], "Competition between private and public schools, vouchers, and peer-group effects," *American Economic Review*, 88(1), 33-62.

[16] Francois, Patrick [2000], "Public service motivation as an argument for government provision," *Journal of Public Economics*, 78, 275-299.

[17] Francois, Patrick and Michael Vlassopoulos [2008], "Pro-social motivation and the delivery of social services," *CESifo Economic Studies*, 54, 22-54.

[18] Friedman, Milton [1962], *Capitalism and Freedom*, Chicago: University of Chicago Press.

[19] Gaynor, Martin, Kate Ho and Robert J. Town [2015], "The industrial organization of health-care markets," *Journal of Economic Literature*, 53(2), 235-284.

[20] Gaynor, Martin, Carol Propper and Stephen Seiler [2012], "Free to choose: reform and demand response in the English National Health Service," NBER Working Paper 18574.

[21] Ghatak, Maitreesh and Hannes Mueller [2011], "Thanks for nothing? Not-for-profits and motivated agents," *Journal of Public Economics*, 95 (1), 94-105.

[22] Gibbons, Stephen, Stephen Machin and O. Silva [2008], "Choice, competition, and pupil achievement," *Journal of the European Economic Association*, 6(4), 912-947.

[23] Glaeser, Edward and Andrei Shleifer [2001], "Not-for-profit entrepreneurs," *Journal of Public Economics*, 81, 99-115.

[24] Hansmann, Henry [1980], "The role of nonprofit enterprise," *Yale Law Journal*, 89, 835-901.

[25] Hart, Oliver, Andrei Shleifer and Robert Vishny [1997], "The proper scope of government: theory and an application to prisons," *Quarterly Journal of Economics*, 108, 1119-1158.

[26] Herrera, C. A., G. Rada, L. Kuhn-Barrientos, and X. Barrios [2014], "Does Ownership Matter? An Overview of Systematic Reviews of the Performance of Private For-Profit, Private Not-For-Profit and Public Healthcare Providers," *PLoS ONE*, 9(12), e93456. http://doi.org/10.1371/journal.pone.0093456.

[27] Hoxby, Caroline M. [1999], "The productivity of schools and other local public goods," *Journal of Public Economics*, 74(1), 1-30.

[28] Hoxby, Caroline M. [2003], "School choice and school productivity: should school choice be a tide that lifts all boats," in Caroline M. Hoxby (ed), *The Economics of School Choice*, NBER, University of Chicago Press.

[29] Jewitt, Ian [1987], "Risk aversion and the choice between risky prospects: the preservation of comparative statics results," *Review of Economic Studies*, 54, 73-86.

[30] Ladd, Helen F. [2002], "School vouchers: a critical view," *Journal of Economic Perspectives*, 16(4), 3-24.

[31] Laffont, Jean-Jacques and Jean Tirole [1993], *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press.

[32] Laine, Liisa and Ching-to Albert Ma [2016], "Quality and competition between public and private firms," unpublished typescript.

[33] Lakdawalla, Darius and Tomas Philipson, [2006], "The nonprofit sector and industry performance," *Journal of Public Economics*, 90(9), 1681-1698.

[34] Lavy, Victor [2010], "Effects of free choice among public schools,"*Review of Economic Studies*, 77(3), 1164-1191.

[35] Le Grand, Julian [2007], *The Other Invisible Hand: Delivering Public Services through Choice and Competition*, Princeton NJ and Oxford: Princeton University Press.

[36] McMillan, Robert [2005], "Competition, incentives, and public school productivity," *Journal of Public Economics*, 89, 1131-1154.

[37] Neal, Derek [2002], "How vouchers could change the market for education," *Journal of Economic Perspectives*, 16(4), 25-44.

[38] Nechyba, Thomas J. [2000], "Mobility, targeting, and private-school vouchers," *American Economic Review*, 90(1), 130-146.

[39] Newhouse, Joseph P. [1970], "Toward a theory of nonprofit institutions: an economic model of a hospital," *American Economic Review*, 60(1), 64-74.

[40] Sloan, Frank A. [2000], "Not-for-profit ownership and hospital behavior," in Anthony J. Culyer, and Joseph P. Newhouse (eds), *Handbook of Health Economics*, Amsterdam: Elsevier.

[41] Stigler, George [1971], "The theory of economic regulation," *Bell Journal of Economics and Management Science*, 2, 3-21.

[42] Tooley, James and Pauline Dixon [2005], "Private education is good for the poor: a study of private schools serving the poor in low-income countries," Cato Institute, available at http://object.cato.org/sites/cato.org/files/pubs/pdf/tooley.pdf.

Appendix A Proofs

Proof of Proposition 1 That $u^{FP}(\theta_I, p_I) = 0$ for all (θ_I, p_I) follows directly from maximization of $v_I^{FP}(U, \theta_I, p_I)$ specified in (2) subject to $U \geq 0$. That $u^{NP}(\theta_I, p_I) = U^*(\alpha, \theta_I, p_I) > 0$ for all (θ_I, p_I) follows from the definition of $U^*(\alpha, \theta_I, p_I)$ in (7) and that this is strictly positive because both $Q^*(\alpha, \theta_i, p_i)$ and $q^*(\alpha, \theta_i, p_i)$ are strictly positive. With a monopoly not-for-profit incumbent, consumers' utility is given by (7) with $i = I$. For any parameter $z \in \{\alpha, \theta_I, p_I\}$,

$$\frac{\partial U^*(\alpha, \theta_I, p_I)}{\partial z} = \frac{\partial Q^*(\alpha, \theta_I, p_I)}{\partial z} + \frac{\partial q^*(\alpha, \theta_I, p_I)}{\partial z}. \tag{A.1}$$

From the first-order conditions (6), note that μ must be strictly greater than zero, so the profit constraint (5) holds with equality. From these, for $i = I$ and $x_I > 0$,

$$\alpha c'(q^*(\alpha, \theta_I, p_I)) = c'(Q^*(\alpha, \theta_I, p_I)) \tag{A.2}$$

and

$$\theta_I p_I = c(Q^*(\alpha, \theta_I, p_I)) + c(q^*(\alpha, \theta_I, p_I)). \tag{A.3}$$

Consider first $z = \alpha$. Differentiation of (A.3) with respect to α gives

$$c'(Q^*(\alpha, \theta_I, p_I)) \frac{\partial Q^*(\alpha, \theta_I, p_I)}{\partial \alpha} + c'(q^*(\alpha, \theta_I, p_I)) \frac{\partial q^*(\alpha, \theta_I, p_I)}{\partial \alpha} = 0$$

and, hence,

$$\frac{\partial q^*(\alpha, \theta_I, p_I)}{\partial \alpha} = -\frac{c'(Q^*(\alpha, \theta_I, p_I))}{c'(q^*(\alpha, \theta_I, p_I))} \frac{\partial Q^*(\alpha, \theta_I, p_I)}{\partial \alpha}. \tag{A.4}$$

Substitution for $c'(Q^*(\alpha, \theta_I, p_I))$ in (A.4) from (A.2) gives

$$\frac{\partial q^*(\alpha, \theta_I, p_I)}{\partial \alpha} = -\alpha \frac{\partial Q^*(\alpha, \theta_I, p_I)}{\partial \alpha}$$

which, substituted into (A.1) for $z = \alpha$, gives

$$\frac{\partial U^*(\alpha, \theta_I, p_I)}{\partial \alpha} = (1 - \alpha) \frac{\partial Q^*(\alpha, \theta_I, p_I)}{\partial \alpha}.$$

Differentiation of (A.2) with respect to α gives

$$\alpha c''(q^*(\alpha, \theta_I, p_I)) \frac{\partial q^*(\alpha, \theta_I, p_I)}{\partial \alpha} + c'(q^*(\alpha, \theta_I, p_I)) - c''(Q^*(\alpha, \theta_I, p_I)) \frac{\partial Q^*(\alpha, \theta_I, p_I)}{\partial \alpha} = 0.$$

Substitution for $\partial q^*(\alpha, \theta_I, p_I)/\partial\alpha$ in this from (A.4) gives

$$\frac{\partial Q^*(\alpha,\theta_I,p_I)}{\partial\alpha}\left[\frac{c(Q^*(\alpha,\theta_I,p_I))}{c'(q^*(\alpha,\theta_I,p_I))}\alpha c''(q^*(\alpha,\theta_I,p_I)) + c''(Q^*(\alpha,\theta_I,p_I))\right] = c'(q^*(\alpha,\theta_I,p_I)),$$

which implies $\partial Q^*(\alpha,\theta_I,p_I)/\partial\alpha > 0$ and hence $\partial U^*(\alpha,\theta_I,p_I)/\partial\alpha < 0$ because c is strictly increasing and strictly convex and $\alpha > 1$.

Consider now $z = \theta_I$. Differentiation of (A.2) and (A.3) with respect to θ_I gives

$$\alpha c''(q^*(\alpha,\theta_I,p_I))\frac{\partial q^*(\alpha,\theta_I,p_I)}{\partial\theta_I} - c''(Q^*(\alpha,\theta_I,p_I))\frac{\partial Q^*(\alpha,\theta_I,p_I)}{\partial\theta_I} = 0$$

and

$$c'(Q^*(\alpha,\theta_I,p_I))\frac{\partial Q^*(\alpha,\theta_I,p_I)}{\partial\theta_I} + c'(q^*(\alpha,\theta_I,p_I))\frac{\partial q^*(\alpha,\theta_I,p_I)}{\partial\theta_I} = p_I. \tag{A.5}$$

The former can be solved for $\partial Q^*(\alpha,\theta_I,p_I)/\partial\theta_I$ to give

$$\frac{\partial Q^*(\alpha,\theta_I,p_I)}{\partial\theta_I} = \frac{\alpha c''(q^*(\alpha,\theta_I,p_I))}{c''(Q^*(\alpha,\theta_I,p_I))}\frac{\partial q^*(\alpha,\theta_I,p_I)}{\partial\theta_I}. \tag{A.6}$$

Use of this in (A.1) for $z = \theta_I$ gives

$$\frac{\partial U^*(\alpha,\theta_I,p_I)}{\partial\theta_I} = \left[\frac{\alpha c'(q^*(\alpha,\theta_I,p_I))}{c''(Q^*(\alpha,\theta_I,p_I))} + 1\right]\frac{\partial q^*(\alpha,\theta_I,p_I)}{\partial\theta_I},$$

which is positive if $\partial q^*(\alpha,\theta_I,p_I)/\partial\theta_I > 0$. Use of (A.6) in (A.5) and substitution for $c'(Q^*(\alpha,\theta_I,p_I))$ from (A.2) gives

$$\left[\frac{\alpha^2 c'(q^*(\alpha,\theta_I,p_I))}{c''(Q^*(\alpha,\theta_I,p_I))} + 1\right]c'(q^*(\alpha,\theta_I,p_I))\frac{\partial q^*(\alpha,\theta_I,p_I)}{\partial\theta_I} = p_I,$$

from which $\partial q^*(\alpha,\theta_I,p_I)/\partial\theta_I > 0$ because c is strictly increasing and strictly convex. For $z = p_I$, the argument is essentially identical to that for $z = \theta_I$.

Proof of Proposition 2 Suppose entry were to occur when (14) does not hold. Then, by the definition of $\widehat{\gamma}^j(U,\theta_I,p_I)$ in (12), the incumbent would compete to supply the whole market for even the highest payoff $\tilde{U}_E^k(\theta_E,p_E)$ the entrant would be willing to offer the γ flexible consumers. So the entrant would not succeed in acquiring the flexible consumers and thus no entry would occur, which is a contradiction.[22]

[22]While there exists no *best* response for the incumbent to an offer by the entrant of payoff $\tilde{U}_E^k(\theta_E,p_E)$ in this subgame because indifferent flexible consumers choose the entrant, some offer by the incumbent strictly greater than $\tilde{U}_E^k(\theta_E,p_E)$ is always better than an offer less than or equal to $\tilde{U}_E^k(\theta_E,p_E)$ and

Suppose now (14) holds. Then, by the definition of $\widehat{\gamma}^j(U, \theta_I, p_I)$ in (12), the incumbent would not compete for the γ flexible consumers if the entrant were to offer $\tilde{U}_E^k(\theta_E, p_E)$. By offering $\tilde{U}_E^k(\theta_E, p_E)$, the entrant would make no less payoff than from not entering and would acquire the flexible consumers, so entry occurs.

For determining consumer payoffs, there are two cases to consider.

Case 1: $\gamma \geq \widehat{\gamma}^j(\tilde{U}_I^j(\theta_I, p_I), \theta_I, p_I)$. In this case, there are sufficient flexible consumers for it to be worth the incumbent competing for them at the highest utility it is ever prepared to offer, $\tilde{U}_I^j(\theta_I, p_I)$. If entry occurs (which in this case is only if $\tilde{U}_E^k(\theta_E, p_E) \geq \tilde{U}_I^j(\theta_I, p_I)$ because otherwise (14) is not satisfied), the entrant offers utility of $\tilde{U}_I^j(\theta_I, p_I)$ so that it is not worth the incumbent attracting flexible consumers or, if higher, the payoff $u^k(\theta_E, p_E)$ it would offer in the absence of competition. The incumbent offers utility $u^j(\theta_I, p_I)$ and attracts only the rigid consumers, who thus receive utility $u^j(\theta_I, p_I)$. Flexible consumers choose the entrant and receive payoff $\max\{\tilde{U}_I^j(\theta_I, p_I), u^k(\theta_E, p_E)\}$.

Case 2: $\gamma < \widehat{\gamma}^j(\tilde{U}_I^j(\theta_I, p_I), \theta_I, p_I)$. In this case, there are insufficient flexible consumers for it to be worth the incumbent competing for them at the highest utility it is ever prepared to offer, $\tilde{U}_I^j(\theta_I, p_I)$. If entry occurs, therefore, the entrant offers the lowest consumer payoff, $\bar{U}^j(\gamma, \theta_I, p_I)$ defined in (13), for which it is not worth the incumbent competing for flexible consumers or, if higher, the payoff $u^k(\theta_E, p_E)$ it would offer in the absence of competition. The incumbent then offers $u^j(\theta_I, p_I)$ and serves only the rigid consumers, who thus receive utility $u^j(\theta_I, p_I)$. Flexible consumers choose the entrant and receive payoff $\max\{\bar{U}^j(\gamma, \theta_I, p_I), u^k(\theta_E, p_E)\}$.

Entry increases the utility of flexible consumers because $\tilde{U}_I^j(\theta_I, p_I) \geq \bar{U}^j(\gamma, \theta_I, p_I) > u^j(\theta_I, p_I)$ and leaves utility of rigid consumers unchanged because they receive $u^j(\theta_I, p_I)$ both with and without entry.

Proof of Proposition 3 For a not-for-profit incumbent with $p_I = p$, $u^{NP}(\theta_I, p) = U^*(\alpha, \theta_I, p)$. We first show that, if $\theta_E \leq \theta_I$, then $U^*(\alpha, \theta_I, p) > \tilde{U}_E^{FP}(\theta_E, p)$. Suppose not. Then from (3),

$$\begin{aligned} p\theta_E &= c\left(\tilde{U}_E^{FP}(\theta_E, p)\right) \\ &> c\left(\tilde{U}_E^{FP}(\theta_E, p) - q^*(\alpha, \theta_I, p)\right) + c\left(q^*(\alpha, \theta_I, p)\right) \\ &> c\left(U^*(\alpha, \theta_I, p) - q^*(\alpha, \theta_I, p)\right) + c\left(q^*(\alpha, \theta_I, p)\right) \\ &= p\theta_I. \end{aligned}$$

ensures that the potential entrant does not actually enter. For this reason, the equilibrium described in Proposition 2 is technically a subgame perfect ε-equilibrium. Because, however, this subgame is not reached along the equilibrium path, the equilibrium payoffs in Proposition 2 are exact, not approximate.

The second line of this follows because c is strictly convex, the third line from the supposition that $U^*(\alpha, \theta_I, p) \leq \tilde{U}_E^{FP}(\theta_E, p)$, and the final line because the breakeven constraint for the not-for-profit incumbent (5) holds with equality. This contradicts $\theta_E \leq \theta_I$ so it must be the case that, when that condition holds, $U^*(\alpha, \theta_I, p) > \tilde{U}_E^{FP}(\theta_E, p)$. But then the payoff to consumers that the not-for-profit incumbent would choose to offer even if not competing for flexible consumers is greater than the highest payoff the for-profit potential entrant would offer them. So the entrant would never attract the flexible consumers and so would not enter. Since in this case $Q_I > 0$ and $Q_E = 0$, entry reduces unobservable quality for flexible consumers to the minimal level. But, from Proposition 2, their payoff increases with entry so it must be that $q_E > q_I$. Thus entry increases observable quality for flexible consumers.

Proof of Proposition 4 Let

$$\varphi(U, \alpha, \theta_I, p) = \frac{\alpha \hat{Q}^*(U, \theta_I, p) + \hat{q}^*(U, \theta_I, p)}{\alpha Q^*(\alpha, \theta_I, p) + q^*(\alpha, \theta_I, p)}.$$

Observe that the denominator in this is a maximum value function for the monopoly not-for-profit's maximization problem with $Q^*(\alpha, \theta_I, p)$ and $q^*(\alpha, \theta_I, p)$ the maximizers and that α enters only the objective function and not the constraint. So, by the envelope theorem, its derivative with respect to α is just $Q^*(\alpha, \theta_I, p)$. Moreover, $\hat{Q}^*(U, \theta_I, p)$ and $\hat{q}^*(U, \theta_I, p)$ are independent of α. Thus

$$\operatorname{sgn} \frac{\partial \varphi(U, \alpha, \theta_I, p)}{\partial \alpha} = \operatorname{sgn} \quad \left(\hat{Q}^*(U, \theta_I, p)\, q^*(\alpha, \theta_I, p) - Q^*(\alpha, \theta_I, p)\, \hat{q}^*(U, \theta_I, p)\right) < 0, \qquad \text{(A.7)}$$

the inequality following because $\hat{Q}^*(U, \theta_I, p) < Q^*(\alpha, \theta_I, p)$ and $\hat{q}^*(U, \theta_I, p) > q^*(\alpha, \theta_I, p)$. Note from (12) that, for entry to occur with $\widehat{\gamma}^j(U, \theta_I, p) = 0$, it must be that the utility U offered by the entrant satisfies $U = u^j(\theta_I, p)$ because indifferent flexible consumers choose the entrant and offering higher U would reduce the entrant's profit. It then follows from (11) and (12) that, when less than one,

$$\widehat{\gamma}^{NP}(U, \theta_I, p) = 1 - \varphi(U, \alpha, \theta_I, p),$$

which is thus increasing in α for any U and, in particular, for $U = \tilde{U}_E^{FP}(\theta_E, p)$.

Next note that $\varphi(.)$ is decreasing in U when $U \geq U^*(\alpha, \theta_I, p)$ because the numerator is then the maximum value function of a problem in which an increase in U corresponds to a tighter constraint. To show $\widehat{\gamma}^{NP}(\tilde{U}_E^{FP}(\theta_E, p), \theta_I, p)$ is then increasing in θ_E, it thus suffices to note that $\tilde{U}_E^{FP}(\theta_E, p)$ is increasing in θ_E. From the definition of

$\hat{\gamma}^j(U, \theta_I, p)$ in (12) and $\bar{U}^j(\gamma, \theta_I, p)$ in (13),

$$\gamma = \frac{v_I^{NP}\left(U^*\left(\alpha, \theta_I, p\right), \theta_I, p\right) - v_I^{NP}(\bar{U}^{NP}(\gamma, \theta_I, p), \theta_I, p)}{v_I^{NP}\left(U^*\left(\alpha, \theta_I, p\right), \theta_I, p\right)}.$$

The right-hand side of this is just $1 - \phi\left(U, \alpha, \theta_I, p\right)$ evaluated at $U = \bar{U}^j\left(\gamma, \theta_I, p\right)$. It has already been shown that $1 - \phi\left(U, \alpha, \theta_I, p\right)$ is increasing in α for any U and it was previously shown that $v_I^{NP}\left(U, \theta_I, p\right)$ is decreasing in U, which suffices to complete the result.

Proof of Proposition 5 **Case 1:** For a for-profit incumbent with $p_I = p$, $u^{FP}\left(\theta_I, p\right) = 0$. Use of this and the incumbent payoff, (2) for $i = I$, in (12) gives

$$\hat{\gamma}^{FP}\left(U, \theta_I, p\right) = \min\left\{\frac{c\left(U\right) - c\left(0\right)}{p\theta_I - c\left(0\right)}, 1\right\}, \tag{A.8}$$

which yields $\hat{\gamma}^{FP}\left(U, \theta_I, p\right) = 1$ only if $U \geq \tilde{U}_I^{FP}\left(\theta_I, p\right)$. When the potential entrant is also a for-profit provider, from (3) for $i = E$, $\tilde{U}_E^{FP}(\gamma, \theta_E)$ satisfies

$$c(\tilde{U}_E^{FP}(\theta_E, p)) = p\theta_E. \tag{A.9}$$

Use of this and $c\left(0\right) = 0$ in (A.8) gives (16). By Proposition 2, entry occurs if $\gamma \leq \hat{\gamma}^j(\tilde{U}_E^k(\theta_E, p), \theta_I, p)$. From (16), when $\theta_E \geq \theta_I$, then $\hat{\gamma}^j(\tilde{U}_E^k(\theta_E, p), \theta_I, p) = 1$, so entry occurs for any $\gamma \leq 1$. For-profit providers always set $Q_I = Q_E = 0$, so unobserved quality for flexible consumers is the same minimal level with entry as without. But, from Proposition 2, their utility increases with entry so it must be that $q_E > q_I$. Thus entry increases observable quality for flexible consumers.

Case 2: For $\theta_E > \theta_I$, $\tilde{U}_E^{NP}\left(\theta_E, p\right) > \tilde{U}_I^{NP}\left(\theta_I, p\right)$. The entrant is, therefore, always willing to offer utility higher than the incumbent can afford to attract flexible consumers, so entry always occurs.

Proof of Proposition 6 From Proposition 2, rigid consumers receive utility $u^j\left(\theta_I, p\right)$ when the incumbent is type $j \in \{FP, NP\}$, which is exactly the same as when type j is a monopoly provider, so the result for them follows from Proposition 1. Also from Proposition 2, the result certainly holds for flexible consumers if the utility ranking claimed in the proposition holds. To establish that ranking, note that $\tilde{U}_I^j\left(\theta_I, p\right) \geq \bar{U}^j(\gamma, \theta_I, p)$ for $j \in \{FP, NP\}$ follows from the definition of $\tilde{U}_I^j\left(\theta_I, p\right)$ as the highest utility a type j incumbent with efficiency parameter θ_I can feasibly deliver. So, to establish the proposition, it remains to show only that $\bar{U}^{NP}(\gamma, \theta_I, p) > \tilde{U}_I^{FP}\left(\theta_I, p\right)$.

Suppose contrary to this that $\bar{U}^{NP}(\gamma, \theta_I, p) \leq \tilde{U}_I^{FP}(\theta_I, p)$. Then, from (3),

$$\begin{aligned}
p\theta_I &= c\left(\tilde{U}_I^{FP}(\theta_I, p)\right) \\
&> c\left(\tilde{U}_I^{FP}(\theta_I, p) - \hat{q}^*(\bar{U}^{NP}(\gamma, \theta_I, p), \theta_I, p)\right) + c\left(\hat{q}^*(\bar{U}^{NP}(\gamma, \theta_I, p), \theta_I, p)\right) \\
&\geq c\left(\bar{U}^{NP}(\gamma, \theta_I, p) - \hat{q}^*(\bar{U}^{NP}(\gamma, \theta_I, p), \theta_I, p)\right) + c\left(\hat{q}^*(\bar{U}^{NP}(\gamma, \theta_I, p), \theta_I, p)\right) \\
&= c\left(\hat{Q}^*\left(\bar{U}^{NP}(\gamma, \theta_I, p)\right), \theta_I, p\right) + c\left(\hat{q}^*\left(\bar{U}^{NP}(\gamma, \theta_I, p), \theta_I, p\right)\right) \\
&= p\theta_I.
\end{aligned}$$

The second line of this follows because c is strictly convex, the third line from the supposition that $\bar{U}^{NP}(\gamma, \theta_I, p) \leq \tilde{U}_I^{FP}(\theta_I, p)$, the fourth line from (8) and the final line because the breakeven constraint for a not-for-profit incumbent (5) holds with equality. But this gives a contradiction, so it must be that $\bar{U}^{NP}(\gamma, \theta_I, p) > \tilde{U}_I^{FP}(\theta_I, p)$.

Proof of Proposition 7 First note that the welfare criterion (18) is differentiable with respect to p_E for $\hat{\theta}_E(p_E, p_I) \in [\underline{\theta}, \bar{\theta}]$ with derivative

$$-\gamma g\left(\hat{\theta}_E(p_E, p_I)\right)\left[\Delta U(p_I) + \xi(p_I - p_E)\right]\frac{\partial \hat{\theta}_E(p_E, p_I)}{\partial p_E} - \gamma\left[1 - G\left(\hat{\theta}_E(p_E, p_I)\right)\right]\xi. \tag{A.10}$$

From (17),

$$\frac{\partial \hat{\theta}_E(p_E, p_I)}{\partial p_E} = -\frac{\hat{\theta}_E(p_E, p_I)}{p_E} \tag{A.11}$$

so (A.10) can be written

$$\gamma g\left(\hat{\theta}_E(p_E, p_I)\right)\left[\Delta U(p_I) + \xi(p_I - p_E)\right]\frac{\hat{\theta}_E(p_E, p_I)}{p_E} - \gamma\left[1 - G\left(\hat{\theta}_E(p_E, p_I)\right)\right]\xi. \tag{A.12}$$

This has the same sign as

$$\frac{g\left(\hat{\theta}_E(p_E, p_I)\right)}{1 - G\left(\hat{\theta}_E(p_E, p_I)\right)}\hat{\theta}_E(p_E, p_I)\left[\frac{\Delta U(p_I)}{p_E} + \xi\left(\frac{p_I}{p_E} - 1\right)\right] - \xi. \tag{A.13}$$

Now, by Corollary 2 in Bagnoli and Bergstrom (2005), $g(\theta_E)$ log-concave implies that $g(\theta_E) / [1 - G(\theta_E)]$ is monotone increasing in θ_E and, from (A.11), $\hat{\theta}_E(p_E, p_I)$ is decreasing in p_E. From this it follows that the whole expression in (A.13) is decreasing in p_E. Thus the derivative (A.12) can be zero for at most one value of p_E — call it p_E^*. Moreover, at p_E^* it is passing from positive to negative, so p_E^* corresponds to a maximum. Furthermore, for p_E low enough, $\hat{\theta}_E(p_E, p_I) = \bar{\theta}$ and the derivative in (A.12) is positive and for p_E high enough the derivative in (A.12) is negative. Thus there always exists such a p_E^*. Setting (A.12) equal to zero and re-arranging gives (19).

Appendix B A More General Formulation

B.1 Generalized Model and Results

This appendix provides a more general formulation of the core ideas where, instead of having only two groups of consumers, we allow for the possibility that any consumer is willing to switch to the entrant. We also allow for more than two dimensions of quality. The main aim of the section is to show that the core insights from the model in the main text carry over to this more general setting.

Suppose then that consumers differ in their benefit $b \in [\underline{b}, \bar{b}]$ from switching to the entrant, with distribution function $F(b)$ that admits a density and is log-concave.[23] We make no assumption about the signs of $\underline{b}$ and $\bar{b}$, so consumers may prefer to stay with the incumbent, or to switch to the entrant, when offered the same quality levels by both.

The continuous benefit from switching generalizes the idea of rigid and flexible consumers. This benefit can arise for a variety of reasons that are relevant for schools and hospitals, reflecting, for example, the geographical location of the incumbent or entrant which makes use of one of the providers more convenient for some consumers. It might also proxy for other intrinsic attributes.

Our general formulation also allows for vectors of both types of quality. Specifically, let q^1 be an M-element vector of observable qualities, with generic element q^1_m, that a provider can commit to before consumers choose their provider, q^2 be an N-element vector of unobservable qualities, with generic element q^2_n, to which commitment is infeasible before consumers choose their provider, and q be the overall vector of qualities (q^1, q^2). For notational convenience let π denote the vector of parameters in the model.[24] All consumers have the same utility $U(q, \pi)$ from provision by the incumbent, which is everywhere strictly increasing in each element of q. A consumer with switching benefit b has utility $U(q, \pi) + b$ from being served by an entrant that provides quality vector q. As before, consumers choose provision if and only if they attain utility of at least zero and those indifferent between providers choose the entrant.

Providers have constant returns to scale and serve all consumers who come to them.[25] They enter the market if and only if they achieve a positive payoff from doing so and the order of moves is the same as in the main text.

As before, for $j \in \{FP, NP\}$ and $i \in \{I, E\}$, let $u^j(\pi)$ be the utility to consumers delivered by a type j provider if not constrained by competition and $\tilde{U}^j_i(\pi) > u^j(\pi)$ be the highest consumer payoff type j is willing to provide, but now both net of switching

[23] This is weaker than the more widely used assumption that the density F' is log-concave; see Jewitt (1987) for discussion.

[24] For the model in the main text, $\pi = (\theta_I, \theta_E, p_I, p_E, \alpha, \lambda, \beta)$. However, the parameterization in the generalized model can be richer than that.

[25] It is straightforward to introduce an entry cost.

benefit. Also, let $v_i^j(U,\pi)$ be the highest payoff available to provider type j if delivering consumer utility U conditional on having entered the market. This is assumed to be continuously differentiable and strictly concave in U for all U in excess of what the provider would offer in the absence of competition.[26]

Conditional on utility offers U_I and U_E from the incumbent and entrant respectively, both net of switching benefit, consumers with switching benefit b choose I if $U_I > U_E + b$. Let

$$b^*(U_I, U_E) = \begin{cases} \underline{b}\,, & \text{if } U_I < U_E + \underline{b} \\ U_I - U_E, & \text{if } U_E + \underline{b} \le U_I \le U_E + \bar{b}; \\ \bar{b}\,, & \text{if } U_I > U_E + \bar{b}; \end{cases} \tag{B.1}$$

be the value of b that determines consumer choices given U_I and U_E and let $U_I^j(U_E,\pi)$ denote the best response utility offer for a type $j \in \{FP, NP\}$ incumbent if the entrant offers U_E. We assume $\underline{b}$ sufficiently low that the incumbent always chooses to retain some consumers. For this generalized formulation, the following result corresponds to Proposition 2.

Proposition 8 *For $j,k \in \{FP, NP\}$, a sufficient condition for entry by a type k potential entrant facing a type j incumbent is that $\tilde{U}_I^j(\pi) < \tilde{U}_E^k(\pi) + \bar{b}$. For $\tilde{U}_I^j(\pi) \ge \tilde{U}_E^k(\pi) + \bar{b}$, a necessary and sufficient condition for entry by a type k potential entrant facing a type j incumbent is*

$$\frac{\partial}{\partial U_I} v_I^j(U_I,\pi) + v_I^j(U_I,\pi)\, F'(\bar{b}) \le 0, \text{ for } \quad U_I = \tilde{U}_E^k(\pi) + \bar{b}, \;\; j,k \in \{FP, NP\}\,. \tag{B.2}$$

If the incumbent would set $u^j(\pi) \in \left(0, \tilde{U}_I^i(\pi)\right)$ in the absence of entry and entry occurs, all consumers strictly gain from entry.

Entry occurs as long as the entrant has a non-negative payoff from servicing the consumers with the highest benefit from switching, those with $b = \bar{b}$. Thus there is entry for sure if the highest utility the entrant is willing to offer attracts some consumers even when the incumbent also offers the highest utility it is willing to offer (that is, if $\tilde{U}_I^j(\pi) < \tilde{U}_E^k(\pi) + \bar{b}$). Otherwise, there is entry if and only if the incumbent prefers to cede part of the market at the highest utility the entrant is willing to offer, a condition captured by (B.2), which generalizes (14).

The main economic difference from this more general formulation is that even consumers who do not switch to an entrant can strictly gain from entry,[27] which strength-

[26] These properties are satisfied by the specific functional forms in the main text.

[27] That is certainly the case if the incumbent is a not-for-profit that offers strictly positive utility in the absence of entry (that is, $u^{NP}(\pi) > 0$) because then, with all consumers potentially flexible, it is always worth the incumbent offering at least marginally higher utility to retain some additional consumers. It may not be the case with a for-profit incumbent who, as in the previous model, sets $u^{FP}(\pi) = 0$.

ens the welfare results. This is because competition may lead the incumbent to offer higher utility to retain additional consumers.

For the model of the main text, the probability of entry by a for-profit provider is lower with a not-for-profit incumbent than with a for-profit incumbent. The next result gives a general condition for any parameter change to reduce the probability of entry in the generalized model.

Proposition 9 *For $j, k \in \{FP, NP\}$, consider an equilibrium that, conditional on entry, has U_E such that $U_I^j(U_E, \pi) \in (0, \tilde{U}_I^j(\pi))$. A change in any parameter in π that increases $\frac{\partial v_I^j\left(\tilde{U}_E^k(\pi)+\bar{b},\pi\right)/\partial U}{v_I^j\left(\tilde{U}_E^k(\pi)+\bar{b},\pi\right)}$ but does not affect $v_E^k(U_E, \pi)$ reduces the probability of entry.*

This proposition shows that the finding that entry is less likely with a not-for-profit incumbent than with a for-profit one extends beyond the particular formulation in the main text. There are two potential channels at work here. The first is a cost channel; a not-for-profit incumbent that provides a positive (instead of a zero) level of some unobserved quality can deliver given utility at lower cost even with the same (strictly convex) cost function. That results in an increase in optimal U_I for given U_E. With $v_E^k(U_E, \pi)$ unaffected, this increases the critical value of θ_E at which entry becomes worthwhile and hence, for a given distribution of θ_E, reduces the probability of entry. The second is a payoff channel which depends on how a change in parameter that affects preferences changes the incentive of an incumbent to offer a particular level of U.[28]

It is also instructive to see how the result on encouraging or discouraging entry in Proposition 7 is changed in the more general formulation of this section. To generalize the welfare criterion in (18), it is helpful to define the parameter vector $\hat{\pi}$ as the parameter vector π excluding the efficiency parameter of the potential entrant θ_E and the

Then the incumbent may prefer to offer $U_I^{FP}(U_E, \pi) = 0$ for some U_E even with entry and serve only those consumers with highly negative switching benefits if the distribution F is such that there are sufficient of these. Formally, the difference between $u^j(\pi) > 0$ and $u^j(\pi) = 0$ is that the former is an interior solution at which a marginally higher utility always attracts more consumers when entry occurs, whereas the latter is a corner solution.

[28]The following example illustrates the payoff channel at work. Suppose the not-for-profit incumbent has payoff function $\alpha U(q, \pi) + \Pi(q, \pi)$, where $\Pi(q, \pi)$ is its profit and $\alpha > 0$, and let $q(U, \pi)$ denote the incumbent's optimal choice of quality vector to deliver utility U given the constraints it faces. Then

$$v_I^j(U, \pi) = \alpha U(q(U, \pi)) + \Pi(q(U, \pi), \pi) \ .$$

This reduces to the for-profit payoff function $v_I^{FP}(U, \pi)$ for $\alpha = 0$. Since α affects profit only through the choice of q, it follows from the envelope theorem that $\partial v_I^j(U, \pi)/\partial\alpha = U(q(U, \pi)) > 0$, and hence $\partial^2 v_I^j(U, \pi)/\partial U \partial\alpha = 1$, even if there were no change in unobservable qualities. Moreover, for any best response, $v_I^j(U, \pi)$ is non-increasing in U. Straightforward differentiation then establishes that $\frac{\partial v_I^j(U,\pi)/\partial U}{v_I^j(U,\pi)}$ is increasing in α as long as $U < \tilde{U}_I^j(\pi)$. That also results in an increase in optimal U_I for given U_E. With $v_E^k(U_E, \pi)$ unaffected, this increases the critical efficiency level at which entry becomes worthwhile. Thus an increase in α from zero (which corresponds to moving from a for-profit incumbent to a not-for-profit incumbent) reduces the probability of entry.

payment to the entrant p_E. (That is, $\hat{\pi} = \pi \backslash (\theta_E, p_E)$.) For consistency with the earlier model, the entrant's cost of supplying quality is decreasing in θ_E. Then, with $\hat{}$ used to specify equilibrium values conditional on the parameters, the welfare criterion given incumbent type j and potential entrant type k is

$$G\left(\hat{\theta}_E\left(p_E,\hat{\pi}\right)\right)u^j\left(\hat{\pi}\right)-\xi p_I+\int_{\hat{\theta}_E(p_E,\hat{\pi})}^{\bar{\theta}}\left\{F\left(\hat{b}\left(\theta_E,p_E,\hat{\pi}\right)\right)\hat{U}_I^j\left(\theta_E,p_E,\hat{\pi}\right)\right.$$
$$\left.+\int_{\hat{b}(\theta_E,p_E,\hat{\pi})}^{\bar{b}}\left[\hat{U}_E^k\left(\theta_E,p_E,\hat{\pi}\right)+b-\xi\left(p_E-p_I\right)\right]dF\left(b\right)\right\}dG\left(\theta_E\right), \quad \text{(B.3)}$$

where $\hat{\theta}_E(p_E,\hat{\pi})$ denotes the entrant efficiency at which entry becomes just worthwhile and $\hat{b}\left(\theta_E,p_E,\hat{\pi}\right)=\hat{U}_I^j\left(\left(\theta_E,p_E,\hat{\pi}\right)\right)-\hat{U}_E^k\left(\theta_E,p_E,\hat{\pi}\right)$. The following result is the counterpart to Proposition 7.

Proposition 10 *Suppose, for $j,k\in\{FP,NP\}$, a type j incumbent competes with a type k potential entrant and $\hat{\theta}_E(p_E,\hat{\pi})<\bar{\theta}$ at $p_E=p_I$. Then a policy-maker increases welfare by encouraging entry by increasing p_E above p_I if*

$$-\left[\hat{U}_E^k\left(\hat{\theta}_E\left(p_E,\hat{\pi}\right),p_E,\hat{\pi}\right)+F\left(\hat{b}\left(\hat{\theta}_E\left(p_E,\hat{\pi}\right),p_E,\hat{\pi}\right)\right)\hat{b}\left(\hat{\theta}_E\left(p_E,\hat{\pi}\right),p_E,\hat{\pi}\right)\right.$$
$$\left.+\int_{\hat{b}\left(\hat{\theta}_E(p_E,\hat{\pi}),p_E,\hat{\pi}\right)}^{\bar{b}}bF'\left(b\right)db-u^j\left(\hat{\pi}\right)\right]g\left(\hat{\theta}_E\left(p_E,\hat{\pi}\right)\right)\frac{\partial\hat{\theta}_E\left(p_E,\hat{\pi}\right)}{\partial p_E}$$
$$+\int_{\hat{\theta}_E(p_E,\hat{\pi})}^{\bar{\theta}}\left\{\frac{\partial\hat{U}_E^k\left(\theta_E,p_E,\hat{\pi}\right)}{\partial p_E}+F\left(\hat{b}\left(\theta_E,p_E,\hat{\pi}\right)\right)\frac{\partial\hat{b}\left(\theta_E,p_E,\hat{\pi}\right)}{\partial p_E}\right.$$
$$\left.-\left[1-F\left(\hat{b}\left(\theta_E,p_E,\hat{\pi}\right)\right)\right]\xi\right\}dG\left(\theta_E\right)>0 \quad \text{(B.4)}$$

and discouraging it if the strict inequality is reversed.

The term in square brackets on the top two lines of (B.4) is the utility gain to those consumers who would have switched to an entrant with cost parameter $\hat{\theta}$ from having entry occur at a marginally lower cost parameter as the result of the marginal increase in the payment to the entrant. Unlike in the simple model, it involves an integral term because those consumers switching to the entrant differ in their benefit from doing so. The lower two lines of (B.4) incorporate the welfare effect of the *change* in the proportion of consumers who switch to the entrant because the payment to the entrant affects the utility the entrant offers those who switch. This second effect does not arise in the simple model because there the proportion of consumers who switch is fixed. This second effect complicates evaluation of having different payments for the entrant and the incumbent. But the essential point, in line with the simpler model above, is that there is no more reason to presume that it is optimal to set the same payment for both incumbent and entrant when all consumers are potentially flexible than when only a fixed proportion are.

Overall, the results in this appendix confirm that a range of insights generated by the simple model are indeed robust to having a continuous benefit from switching and arbitrary dimensions of quality. It should also be clear that we do not need to stick to the specific way that we modeled not-for-profit preferences for the core results to hold as long as they satisfy the key assumptions outlined here.[29]

B.2 Proofs for Generalized Model

Lemma 1 *A type j incumbent's best response to an entrant offering U_E that attracts some consumers is the unique $U_I^j(U_E,\pi)$ that satisfies*

$$F\left(b^*\left(U_I^j(U_E,\pi),U_E\right)\right)\frac{\partial}{\partial U_I}v_I^j\left(U_I^j(U_E,\pi),\pi\right) + v_I^j\left(U_I^j(U_E,\pi),\pi\right)F'\left(b^*\left(U_I^j(U_E,\pi),U_E\right)\right)=0 \quad \text{(B.5)}$$

or, equivalently,

$$-\frac{\partial v_I^j\left(U_I^j(U_E,\pi),\pi\right)/\partial U_I}{v_I^j\left(U_I^j(U_E,\pi),\pi\right)}=\frac{F'\left(b^*\left(U_I^j(U_E,\pi),U_E\right)\right)}{F\left(b^*\left(U_I^j(U_E,\pi),U_E\right)\right)}. \quad \text{(B.6)}$$

Proof. A type j incumbent's best response to an entrant offering U_E is $U_I^j(U_E,\pi)$ that satisfies

$$U_I^j(U_E,\pi)\in\arg\max_{U_I} v_I^j(U_I,\pi)\,F\left(b^*(U_I,U_E)\right). \quad \text{(B.7)}$$

The first-order necessary condition for this best response to be interior (that is, with $U_I \in \left(0,\tilde{U}_I^i(\pi)\right)$ such that $b^*(U_I,U_E)\in\left(\underline{b},\bar{b}\right)$) is (B.5) because $\partial b^*(U_I,U_E)/\partial U_I = 1$ for $b^*(U_I,U_E)\in\left(\underline{b},\bar{b}\right)$ from (B.1). Moreover, (B.5) can be written as (B.6). With $v_I^j(U,\pi)$ non-negative and strictly concave in U in the relevant range, the left-hand side of (B.6) is strictly increasing in U_I. With F log concave, F'/F is non-increasing, so the right-hand side of (B.6) is non-increasing in U_I since $\partial b^*(U_I,U_E)/\partial U_I = 1$ at any interior solution from (B.1). There can, therefore, be at most one solution to (B.5) with $U_I\in\left(0,\tilde{U}_I^i(\pi)\right)$ such that $b^*(U_I,U_E)\in\left(\underline{b},\bar{b}\right)$ and hence, by continuity, at most one $U_I^j(U_E,\pi)$ that satisfies (B.7) with $b^*(U_I,U_E)\in\left(\underline{b},\bar{b}\right)$. By assumption, $\underline{b}$ is sufficiently low that the incumbent always chooses to retain some consumers, so U_I such that $b^*(U_I,U_E)=\underline{b}$ cannot be a best response and $b^*(U_I,U_E)=\bar{b}$ corresponds to no entry. Thus there is at most one $U_I^j(U_E,\pi)$ that satisfies (B.7) for given U_E at which entry can

[29] In Section B.3 of this appendix, we give a specific parameterized example where all of these assumptions are satisfied.

occur and this satisfies (B.5) and (B.6). ■

Proof of Proposition 8 With constant returns to scale, the potential entrant enters if and only if it can attract at least the consumers with the largest benefit from switching $\bar{b}$. The proof considers separately the sufficient conditions for entry, the necessary condition for entry, and consumer utility conditional on entry as specified in the proposition.

Sufficient conditions for entry: If $\tilde{U}_I^j(\pi) < \tilde{U}_E^k(\pi) + \bar{b}$, the potential entrant is prepared to offer a higher payoff to type $\bar{b}$ consumers than the incumbent is prepared to offer them, so entry is worthwhile. For $\tilde{U}_I^j(\pi) \geq \tilde{U}_E^k(\pi) + \bar{b}$, suppose (B.2) holds. From Lemma 1, there is at most one solution to (B.5) so the incumbent would not increase its payoff by offering more than $\tilde{U}_E^k(\pi) + \bar{b}$ to retain the consumers with the greatest benefit from switching to the entrant. The entrant would be prepared to offer $\tilde{U}_E^k(\pi)$ to attract those consumers.

Necessary condition for entry: Suppose (B.2) does not hold. Then, even if the entrant offers the highest consumer payoff it is prepared to offer to attract the consumers, $\tilde{U}_E^k(\pi)$, the incumbent's payoff is increasing in U_I at the value that retains even the consumers with the greatest benefit from switching $\bar{b}$. Moreover, with at most one solution to (B.5), the incumbent would obtain a lower payoff by offering any lower U_I.

Consumer utility conditional on entry: In the absence of entry, a type j incumbent chooses U_I to satisfy (B.7) given $b^*(U_I, U_E) = \bar{b}$ so $F(b^*(U_I, U_E)) = 1$ for all $U_I \geq 0$. By definition, the solution to that is $u^j(\pi)$, the payoff to all consumers in the absence of entry. If $u^j(\pi) > 0$, it must satisfy $\partial v_I^j(u^j(\pi), \pi)/\partial U_I = 0$. Conditional on entry, the part of the left-hand side of (B.5) on the lower line is strictly positive for $U_I = u^j(\pi) < \tilde{U}_I^i(\pi)$, as assumed. With $v_I^j(U, \pi)$ strictly concave in U, that implies $U_I^j(U_E, \pi) > u^j(\pi)$. Thus even consumers who do not switch to the entrant receive strictly higher utility conditional on entry as, *a fortiori*, do those who choose to switch to the entrant.

Proof of Proposition 9 By assumption, $\underline{b}$ is sufficiently low that the incumbent always chooses to retain some consumers, so U_I such that $b^*(U_I, U_E) = \underline{b}$ cannot be a best response and $b^*(U_I, U_E) = \bar{b}$ corresponds to no entry. Thus, conditional on entry, the incumbent's best response $U_I^j(U_E, \pi)$ to U_E such that $U_I^j(U_E, \pi) \in \left(0, \tilde{U}_I^j(\pi)\right)$ is, from Lemma 1, given by the unique solution to (B.6). A change in any parameter in π that increases $\frac{\partial v_I^j(U,\pi)/\partial U}{v_I^j(U,\pi)}$ for all $U \in (0, \tilde{U}_I^j(\pi))$ reduces the left-hand side of (B.6) for all $U \in (0, \tilde{U}_I^j(\pi))$, which implies an increase in optimal U_I for given U_E. But an increase in optimal U_I for given U_E with $v_E^k(U_E, \pi)$ unaffected increases the critical value of θ_E at which entry becomes worthwhile and hence, for a given distribution of θ_E, reduces the probability of entry.

Proof of Proposition 10 Substitution for $\hat{U}_I^j(\theta_E, p_E, \hat{\pi})$ in (B.3) using $\hat{b}(\theta_E, p_E, \hat{\pi}) = \hat{U}_E^k(\theta_E, p_E, \hat{\pi}) - \hat{U}_I^j(\theta_E, p_E, \hat{\pi})$ and differentiation with respect to p_E, with p_E set equal to p_I, yields the left-hand side of (B.3). If this is strictly positive, welfare is increased by raising p_E above p_I. If it is strictly negative, welfare is increased by reducing p_E below p_I, as claimed in the proposition.

B.3 Example with multiple qualities

The following example with multiple qualities exhibits properties of $v_i^j(U, \pi)$ that satisfy the assumptions in Appendix B.1. Suppose that the not-for-profit provider has objective function

$$\alpha U(q, \pi) + \Pi(q, \pi), \tag{B.8}$$

where $\Pi(q, \pi)$ is its profit function and $\alpha > 0$, and the utility and profit functions have the forms

$$U(q, \pi) = \sum_{n=1}^{N} r_n q_n; \ \Pi(q, \pi) = p_i - \frac{1}{2\theta_i} \sum_{n=1}^{N} q_n^2, \ \text{ with } \ \theta_i > 0, \text{ for } \ i \in \{I, E\}, \tag{B.9}$$

with $r_n > 0$ for all $n = 1, \ldots, N$, normalized so that $\sum_{m=1}^{M} r_m = 1$, and the other notation as before. We can think of q_n as the square root of the relative monetary expenditure on quality dimension n and r_n as the linearized marginal utility of additional q_n at the enforceable level of quality 0.

A provider chooses quality dimensions $n = M+1, \ldots, N$ to maximize its payoff subject only to the breakeven constraint and non-negativity of the q_n because consumers have already chosen their provider. Its optimization problem at this stage is

$$\max_{q_n, n=M+1,\ldots,N} \left\{ \alpha \sum_{n=M+1}^{N} r_n q_n - \frac{1}{2\theta_i} \sum_{n=M+1}^{N} q_n^2 \right\} \ \text{ subject to} \tag{B.10}$$

$$p_i - \frac{1}{2\theta_i} \sum_{n=1}^{N} q_n^2 \geq 0 \tag{B.11}$$

$$q_n \geq 0, \ \text{for } n = M+1, \ldots, N, \ \text{and given } q_m, \ \text{for } \ m = 1, \ldots, M. \tag{B.12}$$

The first-order condition for an interior solution to q_n is

$$\alpha r_n - \frac{1 + \lambda_i}{\theta_i} q_n = 0, \ \text{ for } n = M+1, \ldots, N,$$

where $\lambda_i \geq 0$ is a multiplier satisfying a complementary inequality with the breakeven constraint (B.11). This gives the solution

$$q_n = r_n \frac{\alpha \theta_i}{1 + \lambda_i}, \ \text{ for } n = M+1, \ldots, N; i \in \{I, E\}. \tag{B.13}$$

Provider i's optimization problem for quality dimensions $m = 1, \ldots, M$ (chosen before consumers have chosen a provider) to deliver utility U must ensure that the q_n satisfy (B.13) and is thus

$$\max_{q_m, m=1,\ldots,M} \left\{ \alpha \sum_{m=1}^{N} r_m q_m + [p_i - \frac{1}{2\theta_i} \sum_{m=1}^{N} q_m^2] \right\} \quad \text{subject to} \tag{B.14}$$

$$p_i - \frac{1}{2\theta_i} \sum_{m=1}^{N} q_m^2 \geq 0 \tag{B.15}$$

$$\sum_{m=1}^{N} r_m q_m \geq U \tag{B.16}$$

$$q_m \geq 0, \text{ for } m = 1, \ldots, M, \text{ and } q_n, \text{ for } n = M+1, \ldots, N, \text{ satisfies (B.13).} \tag{B.17}$$

If U is sufficiently high that the breakeven constraint is binding, $\Pi(q, \pi) = 0$, so a provider with $\alpha > 0$ must be maximizing $\alpha U(q)$, or equivalently $U(q)$, subject to the breakeven constraint regardless of the specific value of α (as long as it is strictly positive). This corresponds to delivering the highest utility that is feasible given the constraints that, for the purposes of this example, we denote $\tilde{U}_i^{\alpha}(\pi)$. Moreover, a for-profit provider with $\alpha = 0$ delivers $\tilde{U}_i^{0}(\pi)$ if its profits are zero. Thus, if $U < \tilde{U}_i^{\alpha}(\pi)$, the breakeven constraint is not binding. Then λ_i in (B.13) is zero and the first-order condition for an interior solution to q_m is

$$r_m(\alpha + \mu_i) - \frac{1}{\theta_i} q_m = 0, \text{ for } m = 1, \ldots, M,$$

where $\mu_i \geq 0$ is a multiplier satisfying a complementary inequality with the utility constraint (B.16). So

$$q_m = r_m (\alpha + \mu_i) \theta_i, \text{ for } m = 1, \ldots, M. \tag{B.18}$$

Use of (B.13) with $\lambda = 0$ and (B.18) in the utility function in (B.9), along with the normalization $\sum_{m=1}^{M} r_m^2 = 1$ and $R = \sum_{n=M+1}^{N} r_n^2$, gives utility

$$(\alpha + \mu_i) \theta_i + \alpha \theta_i R.$$

If this satisfies the utility constraint with $\mu_i = 0$, q_m is given by (B.18) with $\mu_i = 0$. If not, then μ_i must satisfy

$$(\alpha + \mu_i)\theta_i = U - \alpha \theta_i R.$$

Used in (B.18), these give

$$q_m = r_m \max \{\alpha \theta_i, (U - \alpha \theta_i R)\}, \quad \text{for } U \in \left[0, \tilde{U}_i^{\alpha}(\pi)\right), \; m = 1, \ldots, M. \tag{B.19}$$

This and (B.13) for $\lambda = 0$ can be used in the profit function in (B.9) to give, when the breakeven constraint is not binding,

$$\Pi(q,\pi) = p_i - \frac{1}{2\theta_i}\sum_{m=1}^{M}\left(r_m \max\{\alpha\theta_i, U - \alpha\theta_i R\}\right)^2 - \frac{1}{2\theta_i}\sum_{n=M+1}^{N}\left(r_n\alpha\theta_i\right)^2$$

or, with $\sum_{m=1}^{M} r_m^2 = 1$ and $R = \sum_{n=M+1}^{N} r_n^2$,

$$\Pi(q,\pi) = p_i - \frac{1}{2\theta_i}\left[\left(\max\{\alpha\theta_i, U - \alpha\theta_i R\}\right)^2 + \left(\alpha\theta_i\right)^2 R\right]. \tag{B.20}$$

This can be used to check the conditions under which the breakeven constraint (B.11) is not binding. Specifically, the breakeven constraint is not binding if

$$\left(\max\{\alpha\theta_i, U - \alpha\theta_i R\}\right)^2 \leq 2\theta_i p_i - \left(\alpha\theta_i\right)^2 R. \tag{B.21}$$

$\tilde{U}_i^{\alpha}(\pi)$ satisfies (B.21) with equality.

Use of (B.20) in (B.8) gives the payoff to type α from delivering utility $U \in \left[0, \tilde{U}_i^{\alpha}(\pi)\right)$ as

$$v_i^{\alpha}(U,\pi) = \alpha U + p_i - \frac{1}{2\theta_i}\left[\left(\max\{\alpha\theta_i, U - \alpha\theta_i R\}\right)^2 + \left(\alpha\theta_i\right)^2 R\right], \qquad \text{for} \quad U \in \left[0, \tilde{U}_i^{\alpha}(\pi)\right).$$

Note that this function is identical to the case in which there is just one of each type of quality, with marginal utilities of 1 and R respectively. Moreover,

$$\frac{\partial v_i^{\alpha}(U,\pi)}{\partial U} = \begin{cases} \alpha, & \text{for } U \in [0, \alpha\theta_i(1+R)), \\ \alpha(1+R) - \frac{U}{\theta_i}, & \text{for } U \in \left[\alpha\theta_i(1+R), \tilde{U}_i^{\alpha}(\pi)\right). \end{cases}$$

This is positive for

$$U < \alpha\theta_i(1+R),$$

which implies that the utility $u^{\alpha}(\pi)$ offered by the incumbent in the absence of entry is $u^{\alpha}(\pi) = \max\{0, \alpha\theta_I(1+R)\}$, and it is continuous for $U \in \left[u^{\alpha}(\pi), \tilde{U}_i^{\alpha}(\pi)\right)$. It is always negative for $\alpha = 0$, in which case the utility constraint is always binding. Moreover, $\partial^2 v_i^{\alpha}(U,\pi)/\partial U^2 < 0$ for $U \in \left[\alpha\theta_i(1+R), \tilde{U}_i^{\alpha}(\pi)\right)$, so $v_i^{\alpha}(U,\pi)$ is strictly concave in U for $U \in \left[u^{\alpha}(\pi), \tilde{U}_i^{\alpha}(\pi)\right)$.