The Effects of Autoscaling in Cloud Computing on Product Launch

Amir Fazli          Amin Sayedi          Jeffrey D. Shulman

March 13, 2017

*University of Washington, Foster School of Business. All authors contributed equally to this manuscript. Amir Fazli (fazli@uw.edu) is a Ph.D. Student in Marketing, Amin Sayedi (aminsa@uw.edu) is an Assistant Professor of Marketing and Jeffrey D. Shulman (jshulman@uw.edu) is the Marion B. Ingersoll Associate Professor of Marketing, all at the Michael G. Foster School of Business, University of Washington.
The Effects of Autoscaling in Cloud Computing on Product Launch

Abstract
Web-based firms often rely on computational resources to serve their customers, though rarely is the number of customers they will serve known at the time of product launch. Today, many of these computations are run using cloud computing. A recent innovation in cloud computing known as autoscaling allows companies to automatically scale their computational load up or down as needed. We build a game theory model to examine how autoscaling will affect firms’ decisions to enter a new market and the resulting equilibrium prices, profitability, and consumer surplus. Prior to autoscaling, firms planning to launch a new product needed to set their computational capacity before realizing their computational demands. With autoscaling, a company can be assured of meeting demand and pay only for the demand that is realized. Though autoscaling decreases expenditures on unneeded computational resources and therefore should make market entry more attractive, we find this is not always the case. We highlight strategic forces that determine the equilibrium outcome. Our model identifies the likelihood of a firm’s success in a new market and differentiation among potential entrants in the market as key drivers of whether autoscaling increases or decreases market entry, prices, and consumer surplus.

1 Introduction
Consider an entrepreneur or firm deciding whether to launch a web service in a new market. Getting started requires an investment of time and money into research, development, legal, and other starting expenses prior to knowing whether the new product will ever succeed. Furthermore, web and mobile-based firms rely on computational capacity to serve customers. Every interaction a customer has with an application such as a page load, data transfer, and object viewing requires computational resources. This is particularly relevant to companies launching a Software as a Service (SaaS), an industry estimated at $49 billion and expected to grow to $67 billion by 2018 (Columbus 2015). For any company relying on computational resources, cloud computing allows the company to outsource the server set-up and maintenance to a cloud provider.
More and more companies are adopting the cloud to handle their computational needs. Examples of companies using the cloud include big firms such as Netflix, Airbnb, Pinterest, Samsung, Expedia, and Spotify as well as many small businesses and startups (Gaudin 2015; Bort 2015). In fact, end user spending on cloud services in 2015 was estimated to be above $100 billion (Flood 2013) with an expected annual growth of 44% in workloads (Ray 2013).

While cloud computing technology allows for the outsourcing of computational costs, firms who ran their tasks in the cloud often needed to decide, and pre-commit to, their computational capacity at the time of purchasing the cloud service. As such, due to the unpredictable nature of demand for firms entering a new market, the pre-purchased capacity may be excessive or insufficient for the traffic. This issue is specially important in the case of entrepreneurs launching a web startup. For example, BeFunky, an online photo editing startup, was featured on a popular social media site three weeks after launch and saw 30,000 visitors in three hours, crashing their servers (Nickelsburg 2016). Such events happen regularly enough that there is a term for it: the Slashdot effect, which Klems, Nimis, and Tai (2008) describe as occurring when a startup is featured on a popular network, resulting in a significant increase in traffic load and causing the firm’s servers to slow down or crash. Such a problem can be quite costly as an Aberdeen study found that “a 1-second delay in page load time can result in a 7% loss in conversion and a 16% decrease in customer satisfaction” (Poepsel 2008). Kissmetrics, an analytics company, reports that 1 in 4 people abandon a page if it takes longer than 4 seconds to load (Work 2011). Though capacity can later be increased, the missed demand can be costly. As Amazon CEO Jeff Bezos describes, startups face a serious challenge when choosing computational capacity:

> “And you do face this issue (demand uncertainty) whenever you have a startup company. You want to be prepared for lightning to strike because if you’re not, that generates a big regret. If lightning strikes and you weren’t ready for it, that’s kind of hard to live with. At the same time, you don’t want to prepare your physical infrastructure to hubris levels either in the case that lightning doesn’t strike.”

To address this challenge, cloud providers such as Amazon, Microsoft, and Google have begun to offer autoscaling, a feature that allows firms to scale their computational capacity up or down.

---

automatically in real time. Using autoscaling, firms launching a new web-based service can maintain application availability and scale their computational capacity for serving consumers without having to make capacity pre-commitments.

For companies, autoscaling means having just the right number of servers required for meeting the demand at any point in time, which can provide an attractive solution to handling uncertain demand in new markets. Autoscaling is offered with no additional fees and has been celebrated as one of the most beneficial features of cloud computing. Users of autoscaling such as AdRoll and Netflix find that when a new customer comes on board, they can handle the additional traffic instantly. As Mikko Peltola, the Operations Lead at Rovio, noted regarding the benefits of autoscaling, “We can scale up as the number of players go up ... so we can automatically increase the processing power for our servers.” The Chief Technology Officer for Cloud at General Electric has mentioned this feature as one of the main reasons for the company’s move to the cloud, stating “Running inside a public cloud environment, you’re able to consume unlimited capacity as needed” (Weaver 2015). Anecdotally, some entrepreneurs such as Animoto CEO, Brad Jefferson, who used Amazon Web Services (AWS), see the scaling it offers as a game changer for their product launch: “We simply could not have launched Animoto.com and our professional video rendering platform at our current scale without massive CapEx and a lot of VC funding. The viral spike in Animoto video creations we experienced this week would have been disastrous without AWS.”

In this paper, we develop an analytical model to examine how the emergence of autoscaling in cloud computing will affect web-based firms’ decisions regarding market entry and prices. In particular, despite popular belief, we identify conditions for when autoscaling negatively affects market entry. In other words, fewer firms will enter in certain markets due to the advent of autoscaling. Autoscaling has several properties that make it unique from some previously explored areas in marketing and operations. In particular, autoscaling:

- removes a capacity decision that otherwise has to be made prior to pricing;
- converts computational capacity costs from fixed costs to variable costs at the time of pricing;
- allows capacity to be set after the uncertainties regarding consumers’ level of interest and

---

2 See https://aws.amazon.com/solutions/case-studies/adroll/
3 See https://aws.amazon.com/solutions/case-studies/rovio/
competitors’ strategies are resolved; however, autoscaling still

- preserves the uncertainty that exists at the time of making the entry decision.

Given these properties, the effects of autoscaling on company strategies cannot be addressed by prior research on capacity choices and demand uncertainty. In fact, *our model and predictions diverge from prior literature substantively.*

We uniquely incorporate these properties of autoscaling into a game theory model in which two horizontally differentiated firms have the option to enter a market upon incurring an entry cost. We compare a model in which firms choose computational capacity to a model in which firms can choose to adopt autoscaling in cloud computing. The model is constructed to address the following research questions:

1. How does autoscaling affect firms’ profits in the new market?
2. How does autoscaling affect pricing strategies at launch?
3. How does autoscaling affect market entry decisions?
4. How does autoscaling affect consumer surplus?

We explore the roles of several important market factors in determining the answer to each of the research questions. First, we model the ex ante likelihood of a successful product launch. As Griffith (2014) suggests, the value that a firm brings to a new market is unknown before market entry. In some markets, consumer needs are well known and established, therefore yielding a higher likelihood of successfully creating a product that matches consumer needs. For other markets, the consumer needs are less understood and there is a lower likelihood of a successful venture. We show that the likelihood of a successful venture plays a critical role in determining how autoscaling affects equilibrium strategies and profits.

Secondly, we model competition among market entrants. As Burke and Hussels (2013) suggest, competition in new markets is a key factor in determining the performance of new products. In fact, without accounting for competition, autoscaling has a strictly positive impact on the profit of the firm introducing a new product. However, a conventional study of autoscaling for a monopoly does not capture the strategic interactions inherent to autoscaling. By analyzing a competitive
game theory model, we capture these strategic effects and find autoscaling can actually decrease the profits of competing firms.

We identify three general effects of autoscaling: First, the *downside risk reducing* effect of autoscaling prevents firms from investing in excess capacity in case their product is not successful. Second, the *demand satisfaction* effect of autoscaling allows firms to fully serve the demand without facing insufficient capacity in case their product is successful. Finally, autoscaling also has a *competition intensifying* effect, which can result in lower prices and profitability if multiple firms enter the market.

In answering the first research question, we find that autoscaling can increase or decrease firms’ expected profits in the new market. We identify strategic consequences of autoscaling and the conditions that lead to each possibility. In particular, when the probability of success is sufficiently low, autoscaling increases the firms’ expected profits. However, autoscaling may also create a prisoner’s dilemma situation where firms choose autoscaling, but autoscaling lowers their equilibrium profits. In particular, when entry costs are sufficiently small and the probability of success is moderately high, firms adopt autoscaling in equilibrium; however, their equilibrium profits would be higher if autoscaling was not available. This counterintuitive result is driven by the *competition intensifying* effect outweighing the *demand satisfaction* and *downside risk reducing* effects of autoscaling.

In addressing the second research question, we find that autoscaling may increase or decrease average prices charged by competing firms. Existing economic theory would suggest that removing the capacity decision prior to pricing would result in a shift from a Cournot game to a Bertrand game, thereby decreasing prices (e.g., Kreps and Scheinkman 1983). However, our model shows that when both firms enter the market, autoscaling increases the average prices set by each firm if the probability of a successful venture is not too high. On the other hand, if the probability of a successful venture is sufficiently high, then autoscaling decreases the average prices set by each firm. Our model shows that the probability of success plays a critical role in the firms’ capacity choice without autoscaling and therefore influences the magnitude of the *demand satisfaction* and *downside risk reducing* effects of autoscaling relative to the *competition intensifying* effect.

With regard to the third research question, we find that autoscaling can actually decrease market entry. Though we confirm common intuition that the likelihood of a market being served by at least one firm is improved with autoscaling, we find that, under certain conditions, entry
by multiple firms will not occur because of autoscaling. The counter-intuitive result occurs due to competition in the new market, when entry costs are moderately high and there is a high probability that entrants will have a successful venture. In this region, the two firms anticipate autoscaling will heighten price competition after entry, and one firm, therefore, avoids entering the market altogether.

Finally, in addressing the fourth research question, we show that autoscaling may increase or decrease expected consumer surplus depending on the likelihood of a successful venture and entry costs. Given the fact that autoscaling guarantees companies have the capacity to serve consumers in case of high demand, thereby resolving issues such as the Slashdot effect, one might expect that consumers benefit from autoscaling. However, our model shows that when entry costs are low enough such that both firms enter the market, autoscaling decreases expected consumer surplus if and only if the probability of a successful venture is moderate. Intuitively, autoscaling decreases consumer surplus in the region where firms would not have set constraining capacities without autoscaling. In this region, the competition intensifying effect of autoscaling is weak and autoscaling increases firms’ prices, resulting in lower surplus for consumers.

The findings of this study provide implications for various players in new markets, including startups, cloud providers and consumers. Our analysis informs firms entering web-based markets on how autoscaling affects competitive dynamics in pricing and entry. The results suggest that a firm should consider not only the positive direct effect of autoscaling in reducing costs, but also the negative strategic effect in altering the nature of competition. By evaluating the probability of success, the cost of computational capacity, and entry costs, managers can use the findings from this study to determine whether autoscaling increases or decreases the likelihood of monopoly power over the new market. Our findings also inform cloud providers about how autoscaling affects not only the number of firms using the cloud, but also the number of servers each of those firms purchases. Our model provides insights for consumers on how autoscaling affects the prices charged in the market, showing that for high probabilities of success, average prices decrease with autoscaling and for lower probabilities of success they increase with autoscaling. We also find conditions for which autoscaling will decrease or increase consumer surplus, which can be used for consumer surplus maximizing policy design. To the best of our knowledge, this paper is the first to study the marketing aspects of cloud computing, and how it can affect prices and market entry. With
the growing trend of adopting the cloud by firms, cloud computing is expected to become a major part of any business and this provides the field of marketing with a variety of related new topics to explore.

In addition to contributions to practice, our work contributes to economic theory regarding capacity commitments. Our benchmark model uniquely solves a capacity choice game with demand uncertainty and horizontal differentiation between sellers. Contrary to the previous literature, where capacity commitments lead to higher prices, we show that under demand uncertainty, capacity commitments (relative to autoscaling) can intensify the competition and cause lower equilibrium prices. We also uniquely study the effects of removing capacity commitments made under demand uncertainty (via autoscaling) on firms’ market entry decisions.

The rest of this paper is organized in the following order. In Section 2, we review the literature related to our research problem. In Section 3, we introduce the model. In Section 4, we present the analysis of the model and derive the results. We conduct a series of robustness checks and extensions to our model in Section 5. Finally, the discussion of our findings is presented in Section 6.

2 Literature Review

Academic research on cloud computing is still relatively new and most of the work done on this topic focuses on technological issues of the cloud (e.g., Yang and Tate 2012). The few existing business and economics studies of cloud computing have mainly offered conceptual theories and evidence from surveys and specific cases (e.g., Leavitt 2009; Walker 2009; Gupta, Seetharaman, and Raj 2013). Sultan (2011) suggests that cloud computing can benefit small companies due to its flexible cost structure and scalability. Regarding the ability to autoscale, Armbrust et al. (2009) suggest that elasticity in the cloud shifts the risk of misestimating the workload from the user to the cloud provider. Regarding market entry, Marston, Li, Bandyopadhyay, Zhang, and Ghalsasi (2011) conceptually argue that cloud computing can reduce costs of entry and decrease time to market by eliminating upfront investments. Our paper is the first to model autoscaling in the cloud and show how its effects on market entry are not always positive, and depend on market characteristics.

In addition to cloud computing, our research is related to a number of topics in the literature. In particular, previous research shows uncertainty in demand plays an important role in capacity
and production decisions. Desai, Koenigsberg, and Purohit (2007) find the optimal inventory with
demand uncertainty as a function of a product’s durability. Ferguson and Koenigsberg (2007)
examine how a firm should sell its deteriorating perishable inventory and compare this option to
discarding the previously unsold stock. Desai, Koenigsberg, and Purohit (2010) find a strategic
reason for retailers to carry inventory larger than the expected sales in both high and low demand
states. Biyalogorsky and Koenigsberg (2014) consider product introductions by a monopolist facing
uncertainty about consumer valuations and find whether the firm offers multiple products simulta-
neously or sequentially.

Our research is particularly related to the literature considering optimal timing of production
under demand uncertainty. Van Mieghem and Dada (1999) consider a monopoly firm making three
decisions: capacity investment, production quantity, and price. They study the effect of postponing
the two latter decisions. Anupindi and Jiang (2008) study the strategic effects of competition in a
model with flexible production timing. Anand and Girotra (2007) allow firms to postpone product
differentiation by customizing products after more demand information is revealed. Goyal and
Netessine (2007) allow firms to choose between product-flexible or product-dedicated technologies
and invest in capacity before demand uncertainty is resolved, while postponing production decisions
until demand is revealed. These previous studies on postponing production separate the capacity
decision and the production decision: The capacity decision is assumed to occur before demand is
revealed and it is the production decision that can be postponed. With cloud computing, there
is zero production cost after the firm chooses its computational capacity and autoscaling allows
both capacity and production to simultaneously match with demand. In the mentioned papers
on timing of production, capacity constraints are set before demand realization and they influence
how many customers are served, regardless of the timing of production. However, autoscaling
uniquely eliminates the effect of capacity constraints under demand uncertainty, resulting in findings
different from the production postponement literature. For instance, Anupindi and Jiang (2008)
find production postponement increases capacity investment and profitability, whereas we find
when autoscaling may decrease equilibrium computational expenditures and when it may decrease
profitability.

Che, Narasimhan, and Padmanabhan (2010) allow for eliminating the capacity decision under
demand uncertainty by considering a firm’s decision between adopting a make-to-stock system, a
backorder system, or a combination of both. However, a firm’s decision of using autoscaling is conceptually different from the decision between make-to-stock and backorder production. Backorder production delays the time at which customers are served compared to a make-to-stock production, resulting in factors such as time-sensitivity of customers and firms determining the outcome of the model. With autoscaling, on the other hand, firms avoid capacity decision under demand uncertainty while satisfying the demand at the exact same time as they would have with pre-purchased capacity. Che, Narasimhan, and Padmanabhan (2010) find it is optimal for firms to use a combination of make-to-stock and backorder production. However, our model shows using autoscaling and purchasing fixed capacity simultaneously is not an optimal strategy for firms using the cloud.

Our examination of how cloud computing with autoscaling affects a firm’s entry decision also relates to the literature on market entry. A body of literature looks at the timing of entry and how an incumbent can deter entry (e.g., Spence 1977; Joshi, Reibstein, and Zhang 2009; Milgrom and Roberts 1982; and Ofek and Turut 2013). Narasimhan and Zhang (2000) consider firms’ decisions on order of entry into markets with demand uncertainty. They study how market entry from the first mover can unfold market information and resolve demand uncertainty for the second entrant. In contrast, our model examines simultaneous entry decisions by firms, such that both firms face equal demand uncertainty when making entry decisions. Our model adds to entry literature by jointly considering both entry and capacity decisions, such that each firm’s decision to enter depends on the expected future capacity of both firms and whether this capacity will be chosen ex ante or autoscaled to demand.

We compare autoscaling with cases where firms commit to their computational capacity before pricing and realizing demand. This relates to other papers examining capacity commitments. Kreps and Scheinkman (1983) find that a Bertrand pricing game becomes a Cournot game when capacity is chosen prior to pricing. Reynolds and Wilson (2000) extend this model to include uncertainty about market size and find there is no symmetric, pure-strategy equilibrium capacity choice when there is significant demand variation. Nasser and Turcic (2015) find symmetric horizontally differentiated firms use asymmetric strategies on whether to commit to capacity or not. This is consistent with our findings. However, since they do not allow for demand uncertainty, capacity commitments always alleviate competition in their model, whereas, in ours, capacity commitments sometimes intensify competition. Furthermore, at least one firm commits to capacity in any equilibrium
in their model, whereas, in our model, both firms may use autoscaling. Swinney, Cachon, and Netessine (2011) examine the optimal timing of capacity investment in a model in which market price is given by a demand curve and firms can choose to set capacity early at one marginal cost of capacity or after demand is realized at a different cost of capacity. Van Mieghem and Dada (1999) allow firms to choose the time of their pricing decisions and find that postponing pricing until after demand is realized makes the capacity decision less sensitive to demand uncertainty. Our research expands this literature by considering demand uncertainty and market entry decisions in a horizontally differentiated market with and without capacity commitments. In contrast to the previous literature where capacity commitments always alleviate competition, we show that, depending on the level of demand uncertainty, capacity commitments may indeed intensify the competition. Furthermore, in our model, firms decide whether to adopt autoscaling or to pre-commit to capacity; we find that firms do not always follow symmetric strategies in regards to adoption of autoscaling. Finally, we uniquely explore how entry decisions are affected by existence of capacity commitments under demand uncertainty.

Autoscaling in cloud computing also has the effect of converting up-front capacity costs to variable costs that change with demand. Prior research examining the effect of converting fixed to variable costs via outsourcing (e.g., Shy and Stenbacka 2003; Buehler and Haucap 2006; Chen and Wu 2013) have found that prices and profitability rise with this conversion. However, these models do not allow for demand uncertainty. In our model, the decision between up-front investments in capacity and opting for variable cost of capacity through autoscaling is dependent on the level of demand uncertainty in the market. Also, relative to these outsourcing models, up-front capacity cost is endogenous in our model since firms can choose their capacity. Accounting for demand uncertainty and endogenous upfront costs provides novel insights on fixed versus variable capacity costs. In contrast to prior outsourcing literature, we show average prices can fall when cloud computing with autoscaling is used even in conditions for which entry is unaffected.

In summary, our paper uniquely compares market entry, pricing, and profitability between computing resources requiring capacity pre-commitments and cloud computing with autoscaling. The advent of cloud computing with autoscaling has several properties: 1. Autoscaling allows for capacity decisions to be made after the demand uncertainty is resolved. 2. Autoscaling converts the fixed cost of computational resources that is sunk prior to the firms’ pricing decision into
variable costs that are directly affected by the pricing decision. 3. Autoscaling makes it such that capacity and pricing decisions are made simultaneously rather than sequentially. However, autoscaling does not affect the level of uncertainty at the time a firm makes its entry decision. Though previous research has separately examined demand uncertainty, the timing of capacity and pricing decisions, or the conversion of fixed costs to variable costs, our paper is unique in its comprehensive examination of the effect of autoscaling. In particular, our paper is the first to solve for entry, capacity and pricing decisions in a model of horizontally differentiated firms with demand uncertainty and to compare this equilibrium to the entry and pricing decisions of horizontally differentiated firms who make entry decisions with demand uncertainty but whose capacity can autoscale to the demand realized upon setting prices.

3 Model

We consider two symmetric firms who could potentially enter a particular web or mobile application market. To enter the market, the firms would incur a fixed entry cost, \( F \). This cost includes starting expenses such as legal, research and development, and human capital investments. To model post-entry competition, we adopt a discrete horizontal differentiation model (e.g., Narasimhan 1988; Iyer, Soberman, and Villas-Boas 2005; Zhang and Katona 2012; Zhou, Mela, and Amaldoss 2015) with three consumer segments, each consumer demanding at most one unit of the product.\(^5\) Upon entry, each Firm \( i \) will find a segment of consumers, Segment \( i \) with \( i \in \{1, 2\} \), who will buy from Firm \( i \) if and only if the price \( p_i \) is below their reservation value \( v_i \) and who will derive zero value from the competitor’s product. This captures the reality that consumers vary in their taste preferences regardless of firm entry and that firms have idiosyncratic differences that will allow them to serve these tastes differently from each other upon successful entry. The size of each Segment \( i \) is given by \( \alpha < 1/2 \) for \( i \in \{1, 2\} \). The remaining \( 1 - 2\alpha \) consumers are in Segment 3 and are indifferent between firms, preferring to buy from the firm with the lowest price. The parameter \( \alpha \) can be interpreted as the extent to which consumers vary in their taste preferences. Note that \( \alpha \) also represents the level of competition in the market; for \( \alpha = 1/2 \), Segment 3 disappears, each firm

\(^5\)We should note that the Hotelling model also leads to mixed strategy equilibrium in the pricing subgame when there are capacity constraints. The reason that we use the discrete model in Narasimhan (1988) is that, unlike the Hotelling model, it gives us ordinary differential equations when we add capacity decisions to that model.
gets a local monopoly, and there is no competition between the firms. As $\alpha$ becomes smaller, the
segment of consumers for which both firms compete grows and competition intensifies.

In the absence of autoscaling, the timing of the game is as follows:

**Stage 1:** Firms simultaneously decide whether or not to enter the market and thereby incur the
cost $F$. We allow for uncertainty in whether a firm will find the venture successful in terms of
whether $v_i$ is high or low. We assume the ex ante probability of a firm finding success in this
market is $\gamma$, which is common knowledge. In other words, if Firm $i$ enters the market, $v_i$ is an i.i.d.
draw from a binary distribution in which $v_i = 1$ with probability $\gamma$ and $v_i = 0$ with probability
$1 - \gamma$. This assumption reflects the idea that the value provided to customers is unclear for potential
entrants. As Lilien and Yoon (1990) argue, the fit between market requirements and the offering of
the new entrant is highly unpredictable and is critical to the success of the entrant. In a survey of
101 startups, it was reported that the number one reason for the failure of a startup is the lack of
market need for the offered product (Griffith 2014), suggesting that the value created for customers
in a new market is unknown to many firms before entry. In an extension, we allow the low value
condition to be $v_i = v_L > 0$ and verify our results are robust to the assumption.

To remark on the structure of demand and uncertainty, notice that our model set up has several
desirable properties. In particular, it allows a firm to be uncertain about the size of the potential
market and the effect of its price on realized demand: the firm may find itself a monopolist, the firm
may find itself with very low demand (normalized to zero), or the firm may find itself competing
head-to-head. Moreover, a firm’s price relative to its competitor’s is not the only source of de-
mand uncertainty. Though one can explore alternative model specifications to capture these same
properties, the current specification allows for tractability while uncovering a novel mechanism.

**Stage 2:** Firms that enter simultaneously choose computational capacity $k_i$ and incur a computa-
tional capacity cost $ck_i$.

**Stage 3:** The reservation value for each Firm $i$, $v_i$, becomes common knowledge and each firm in
the market simultaneously chooses $p_i$ to maximize profit.

**Stage 4:** Demand is realized. In the case a firm experiences demand greater than its computational
capacity, we assume an efficient rationing rule (see Tirole 1988, p. 213) in which demand from
Segments 1 and 2 is satisfied prior to demand from Segment 3. In an extension, we show that our
results are robust to an alternative proportional rationing rule. Residual demand from Segment 3
Stage 1
Firms simultaneously decide whether to enter.

Stage 2
Firms simultaneously choose capacity, $k_i$.

Stage 3
Firms learn $v_i$ and simultaneously choose price, $p_i$.

Stage 4
Demand is realized and allocated according to an efficient rationing rule.

Figure 1: Sequence of events with no autoscaling

is allocated to the competing firm, provided it has available capacity.

When autoscaling is available, firms simultaneously decide whether to use autoscaling or to choose a computational capacity in Stage 2.\(^6\) If a firm chooses autoscaling, it incurs the computational cost $c$ only on each unit of realized demand. In practice, changing capacity decisions in the absence of autoscaling takes at least a few hours, and in some cases days, before coming into effect on the cloud servers. The Befunky example, the Slashdot effect, and the “lightning strike” analogy by Amazon’s CEO, discussed in the introduction, highlight the fact that demand often changes faster than what firms can respond to in terms of computational capacity. Our assumption that capacity decision is made before the demand is realized captures this reality. However, our main results are robust to this assumption. In particular, even if firms can choose to adopt autoscaling after the demand is realized, our results in Propositions 2, 3, 4, and 5 still hold.

The timing of the game is depicted in Figures 1 and 2. A summary of notation is in Table 1.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Size of each of Segments 1 and 2</td>
</tr>
<tr>
<td>$k_i$</td>
<td>Capacity chosen by Firm $i$</td>
</tr>
<tr>
<td>$v_i$</td>
<td>Reservation value consumers have for Firm $i$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Probability that $v_i = 1$</td>
</tr>
<tr>
<td>$c$</td>
<td>Cost per unit of computational capacity</td>
</tr>
<tr>
<td>$p_i$</td>
<td>Price chosen by Firm $i$</td>
</tr>
<tr>
<td>$F$</td>
<td>Cost of entry</td>
</tr>
</tbody>
</table>

Table 1: Summary of notation

\(^6\)In practice, many companies such as Facebook and Netflix have revealed that they use autoscaling, and startups are highly recommended to do so. We should also note that assuming that firms can observe competitors’ choices of autoscaling when setting their prices allows us to characterize the equilibria in the whole parameter space; however, we do not need this assumption for our main findings. Our main results come from regions in which using autoscaling is a \textit{weakly dominant strategy}, and therefore, firms can rationally infer that their competitors are using autoscaling even if they cannot observe their decisions.
Firms simultaneously decide whether to enter.

Stage 2
Firms decide whether to adopt autoscaling or set capacity $k_i$.

Stage 3
Firms learn $v_i$ and simultaneously choose price, $p_i$.

Stage 4
Demand is realized and allocated according to an efficient rationing rule.

Figure 2: Sequence of events with autoscaling

4 Analysis

Our research objective is to identify how the advent of autoscaling affects equilibrium prices, profits, and market entry. To this end, we first examine equilibrium capacity and prices in the situation in which computational capacity must be determined prior to demand realization. We will subsequently characterize the equilibrium when autoscaling is available. We will conclude with a comparison across these possibilities.

4.1 Choice of Computational Capacity

In this section, we find the equilibrium choices of price and capacity and evaluate the effect of autoscaling on these choices. Throughout this section, we assume entry costs are such that both firms will have entered the market. The analysis of firms’ choice of market entry is left for Section 4.3.

Equilibrium Choices without Autoscaling

We start with solving the model in which there is no autoscaling via backward induction, beginning with the pricing subgame equilibrium. First suppose that $k_1 + k_2 > 1$. We denote this condition as overlapping capacities. We want to calculate equilibrium prices of this game. Without loss of generality, assume that $k_2 \geq k_1$. Also, it is easy to see that firms never set their capacity to $k_i > 1 - \alpha$ or $k_i < \alpha$; therefore, it is sufficient to consider the case where $k_i \in [\alpha, 1 - \alpha]$ for $i \in \{1, 2\}$.

We start by showing that this game does not have a pure strategy equilibrium. Assume for sake of contradiction that the firms use prices $p_1$ and $p_2$ in a pure strategy equilibrium. If $p_1 \neq p_2$, then the firm with a lower price can benefit from deviating by increasing its price to $\frac{p_1 + p_2}{2}$. If $p_1 = p_2$, then
then Firm 2 can benefit from deviating by decreasing its price to \( p_2 - \varepsilon \), for sufficiently small \( \varepsilon \), to acquire more consumers from Segment 3. Therefore, a pure strategy equilibrium cannot exist.

Next, we find a mixed strategy equilibrium for this game. Mixed strategies can be interpreted as sales or promotions and are common in the marketing literature (e.g., Chen and Iyer 2002, Iyer, Soberman, and Villas-Boas 2005, Zhang and Katona 2012). Provided \( k_1 \leq 1 - \alpha \), Firm 2 can choose to *attack* with a price that clears its capacity or *retreat* with a price equal to 1 that harvests the value from the \( 1 - k_1 \) consumers that Firm 1 cannot serve due to its capacity constraint. Let \( z \) be the price at which Firm 2 is indifferent between *attacking* to sell to \( k_2 \) consumers at price \( z \) and *retreating* to sell to \( 1 - k_1 \) consumers at price 1. We have \( z = \frac{1 - k_1}{k_2} \). Figures 3 and 4 demonstrate the different appeals of these two pricing strategies. The choice between *retreating* and *attacking* for each firm depends on the choice of the other firm. If Firm 1’s price is high, it becomes easier for Firm 2 to attract the consumer segment that is in both firms’ reach resulting in Firm 2 choosing to *attack*. On the other hand, if Firm 1’s price is low, Firm 2 would prefer to *retreat* than to compete with Firm 1 over the overlapping consumers. In the equilibrium that we find, both firms use a mixed strategy with prices ranging from \( z \) to 1. Suppose that \( F_i(\cdot) \) is the cumulative distribution function of price set by Firm \( i \) and \( F_j(\cdot) \) is the cumulative distribution function of price set by competing firm \( j \).\(^7\) The profit of Firm \( i \) earned by setting price \( x \), excluding the sunk cost of capacity, is

\[
\pi_i(x) = F_j(x)(1 - k_j)x + (1 - F_j(x))k_i x.
\]

Using equilibrium conditions, we know that the derivative of this function must be zero for \( x \in (z, 1) \). Therefore, we have

\[-x(k_i + k_j - 1)F'_j(x) - F_j(x)(k_i + k_j - 1) + k_i = 0.\]

The solution to this differential equation is

\[
F_j(x) = \frac{k_i}{k_i + k_j - 1} + \frac{C_j}{x}
\]

\(^7\)We are implicitly assuming that \( F_i \) and \( F_j \) are piecewise differentiable. The game could have other mixed strategy equilibria where cumulative distribution functions of prices are not differentiable. We cannot find those equilibria using this method.
where constant $C_j$ is determined by the boundary conditions. As for the boundary conditions, we use $F_1(1) = 1$. Therefore, we get

$$
F_1(x) = \begin{cases}
0 & \text{if } x < z \\
\frac{(k_1-1)+k_2x}{x(k_1+k_2-1)} & \text{if } z \leq x < 1 \\
1 & \text{if } x \geq 1
\end{cases}
$$

This implies that Firm 1 mixes on prices between $z$ and 1 such that Firm 2 is indifferent between using any two prices in this range. Furthermore, given $F_1(.)$, Firm 2 strictly prefers any price in $[z,1]$ to any price outside this interval. To have an equilibrium, the strategy of Firm 2 should be such that Firm 1’s strategy is not suboptimal. In other words, Firm 1 should be indifferent between any two prices in $[z,1]$, and should weakly prefer any price in $[z,1]$ to any price outside this interval. Therefore, we have to use the boundary condition $F_2(z) = 0$ to make sure that (1) Firm 1’s indifference condition is satisfied in $[z,1]$, and (2) Firm 2 does not set the price to lower than $z$, as we already know from $F_1(.)$ that such prices are suboptimal for Firm 2. As such, we get

$$
F_2(x) = \begin{cases}
0 & \text{if } x < z \\
\frac{k_1((k_1-1)+k_2x)}{k_2x(k_1+k_2-1)} & \text{if } z \leq x < 1 \\
1 & \text{if } x \geq 1
\end{cases}
$$

Note that $F_2(x)$ is discontinuous at $x = 1$, and jumps from $\frac{k_1}{k_2}$ to 1. This implies that Firm 2 uses price 1 with probability $1 - \frac{k_1}{k_2}$. In other words, $f_2(1) = (1 - \frac{k_1}{k_2})\delta(0)$, where $f_2(.)$ is the probability density function for price of Firm 2 and $\delta(.)$ is Dirac delta function.\(^8\)\(^9\)

Given $F_i(.)$, we can calculate the expected profit of each firm in this mixed strategy equilibrium. Excluding the sunk cost of capacity, we have

$$
\pi_1 = \frac{(1-k_1)k_1}{k_2} \quad \text{and} \quad \pi_2 = (1-k_1).
$$

---

8 See Hassani (2009), pp 139-170.
9 One might wonder if the probability density $1 - \frac{k_1}{k_2}$ allocated to price 1 by Firm 2 could be instead allocated to price $z$. The answer is that it cannot. While such strategy would still keep Firm 1 indifferent between any two prices in $[z,1]$, it would make price $z - \varepsilon$ (for sufficiently small $\varepsilon$) a strictly better strategy for Firm 1, which violates equilibrium conditions.
Figure 3: How $k_1$ and $k_2$ affect firm incentives to adopt a *retreating* price

![Figure 3](image)

Figure 4: How $k_1$ and $k_2$ affect firm incentives to adopt an *attacking* price

![Figure 4](image)

Note that the higher capacity firm, Firm 2, earns a profit equal to the profit it would have made if it had chosen a *retreating* strategy while pricing at 1; as such, the expected profit of Firm 2 is independent of its capacity $k_2$. On the other hand, the lower capacity firm, Firm 1, earns more than what it would have earned if *retreating* was chosen, since $1 - k_2 < k_1 < k_2$ requires $(1 - k_1)k_2 > (1 - k_2)$. As expected, after excluding sunk costs, the higher capacity firm makes a higher profit than the lower capacity firm.

Note that the mixed strategy pricing equilibrium bears some resemblance to Chen and Iyer (2002) who find in a model of customized pricing the ratio of profits is equal to the ratio of consumer addressability. In our model, the profit ratio is equal to the ratio of capacities. However, the model in Chen and Iyer (2002) is conceptually very different from ours. In particular, the overlap between the customers of the two firms is always non-zero in Chen and Iyer (2002), whereas in our model the overlap is non-zero only if the sum of the capacities is larger than the market size, i.e., $k_1 + k_2 > 1$. Furthermore, even though the ratio of profits is the same in both papers, the actual profit functions are very different. For example, as mentioned above, and in contrast to Chen and Iyer (2002), the profit of the firm with larger capacity does not depend on its own capacity in our model.

Now suppose $k_1 + k_2 \leq 1$. We denote this condition as *separated capacities*. This implies that
each firm that enters the market can sell to its capacity without directly competing with the other firm for consumers in Segment 3. As such, each firm that successfully enters the market can charge \( p_i = 1 \) and sell \( k_i \) units for profit \((1 - c)k_i\). Increasing the price will result in zero sales and profit, decreasing the price will still sell \( k_i \) units but at lower revenue.

Next consider the capacity subgame equilibrium. The capacity decision is made in anticipation of the possible combinations of values for \( v_1 \) and \( v_2 \). If both firms find success (i.e., \( v_1 = v_2 = 1 \)), then the profit depends on how \( k_i \) and \( k_j \) relate to each other and relate to \( \alpha \). The expected profit for Firm \( i \) depends on its capacity relative to the capacity of competing Firm \( j \) and can be written as follows for \( k_i \in [\alpha, 1 - \alpha] \):

\[
E(\pi_i) = \begin{cases} 
\gamma k_i - ck_i & \text{if } k_i + k_j \leq 1 \\
\gamma(1 - \gamma)k_i + \gamma^2(1 - k_i) - ck_i & \text{if } 1 - k_j < k_i < k_j \\
\gamma(1 - \gamma)k_i + \gamma^2(1 - k_j) - ck_i & \text{if } 1 - k_i < k_j \leq k_i
\end{cases}
\]

where index \( j \) indicates the other firm. The equilibrium capacity choices are summarized in the following proposition.

**Proposition 1** Suppose both firms enter the market initially. The equilibrium capacity choices depend on \( \gamma \) as follows:

- If there is a low probability of a successful venture (i.e., \( \gamma < c \)), then both firms choose \( k_i = 0 \) and earn zero profit.

- If there is a moderate probability of a successful venture (i.e., \( \gamma(1 - \gamma) > c \)), then capacities overlap such that one firm sets \( k = 1 - \alpha \) and the other firm sets \( k = k^* \), where \( \alpha < k^* \leq 1 - \alpha \) is defined in the appendix.

- If there is a high probability of a successful venture (i.e., \( \gamma > c \) and \( \gamma(1 - \gamma) < c \)), then capacities do not overlap and the unique symmetric equilibrium is \( k_1 = k_2 = 1/2 \).

The results of Proposition 1 are depicted in Figure 5. Region 2 represents overlapping capacities such that \( k_1 + k_2 > 1 \). Region 3 represents separated capacities such that \( k_1 + k_2 = 1 \) and in the
Proposition 1 highlights a non-monotonic effect of $\gamma$ on the equilibrium capacity choice. Intuitively, if there is a low probability of success then neither firm wishes to invest in computational capacity because there is a high probability of it going unused. Interestingly, when there is a high probability of a successful venture, firms dampen competition by choosing a capacity that just covers the market. To understand this, consider the extreme case in which $\gamma = 1$. If firms choose separated capacities such that $k_1 + k_2 = 1$, both firms can charge their monopoly price for all of their consumers. The moment capacities overlap such that firms compete even for a single consumer, the firms are unable to avoid intense price competition for that consumer, thereby affecting revenues from all of their customers. Though an additional unit of capacity can result in an additional sale, the subsequent effect on price competition is severe enough such that firms refrain from competing directly.\footnote{10}

Another interesting facet of Proposition 1 is that a moderate probability of a successful venture leads to excessive capacity choices. Therefore, a reduction in the probability of success can actually cause an increase in capacity. To understand this result, consider two competing effects of decreasing $\gamma$. On the one hand, lower $\gamma$ implies greater downside risk that the chosen capacity will go completely unused due to $v_i = 0$ and the resulting failure in the market. This effect would suggest that capacity should decrease as $\gamma$ decreases. On the other hand, a firm’s chance at having monopoly power over all of Segment 3 is maximized at moderate levels of $\gamma$. The latter monopoly harvesting effect dominates the former downside risk effect at moderate levels of $\gamma$ resulting in overlapping capacities.

We show in the appendix that when cost of capacity is low enough for overlapping capacities, $c < \gamma(1 + \alpha(\gamma - 1))/(1 - \alpha) - 2\gamma^2$, then both firms set maximum capacity (i.e., $k^* = 1 - \alpha$). Intuitively, when cost of capacity is negligible, even if capacity goes unused, firms do not incur a big loss. Thus both firms focus on fully benefiting from the high probability of being a monopolist, $\gamma(1 - \gamma)$, by setting maximum capacity, without being concerned about the downside risk effect.

\footnote{In Figures 5–9, we use parameters $\alpha = 1/4$, $c = 1/2$, and $F = 0$, unless that parameter is being used as a variable in the figure.}

\footnote{Note that $\alpha$ does not affect the decision between separated and overlapping capacities. This is because regardless of the size of Segment 3, the mere existence of this segment is what drives price competition when capacities overlap. Thus, as long as overlapping capacities fall in the region $\alpha \leq k_i \leq 1 - \alpha$, the mixed strategy pricing chosen by each firm and therefore firms’ profits do not depend on $\alpha$. However, $\alpha$ does determine the equilibrium capacities chosen in the overlapping capacity region as detailed in the appendix.}
To understand the role the assumptions play in the result, it is worth mentioning that firms can have excessive or insufficient capacity due to either demand uncertainty or randomized price competition or both. In other words, the mixed strategy in pricing decisions is not the only source of mismatch between capacity and demand. For instance, in the region for separated capacities, Firm $i$ realizes high value with probability $\gamma$ and low value with probability $1 - \gamma$. Even if it does not face a competitor in the pricing game (i.e., Firm $j$ draws $v_j = 0$), for high $\gamma$ Firm $i$ will have insufficient capacity if $v_i = 1$ and will have excessive capacity if $v_i = 0$. On the other hand, consider when both firms have high value. With probability $\gamma^2$, excessive capacity can exist due to firms setting overlapping capacities and competing over price. Thus, for a moderate probability of success (i.e., $\gamma(1 - \gamma) > c$), two high-value firms compete with mixed pricing strategies, resulting in excessive capacity for the firm with the higher price. Introducing autoscaling, by definition, eliminates the mismatch between capacity and demand. Next, we look at equilibrium strategies in a model of autoscaling.

**Equilibrium Choices with Autoscaling**

We now turn our attention to the equilibrium when autoscaling is available. Major cloud providers offer autoscaling with no additional fees.\(^{12}\) We solve for the equilibrium pricing by entrants supposing both firms enter and choose autoscaling.\(^{13}\)

---


\(^{13}\)Theoretically, a firm could buy a fixed capacity $k$ and also use autoscaling. We do not allow that in our model to simplify exposition, however, it is easy to see that doing so is always dominated by only using autoscaling (and not buying any fixed capacity). A formal proof is available upon request.
With probability $\gamma^2$, we have $v_1 = v_2 = 1$, and it is straightforward to show there is no pure strategy pricing equilibrium; instead the pricing subgame leads to a mixed strategy equilibrium where the prices of both firms range between $z'$ and 1. Similar to our analysis of mixed strategy equilibrium without autoscaling, $z'$ is the price for which each firm is indifferent between *attacking*, resulting in a profit of $(z' - c)(1 - \alpha)$, and *retreating*, resulting in a profit of $\alpha(1 - c)$. This results in $z' = \frac{\alpha(1-c)}{1-\alpha} + c$.

Supposing that $G_i(.)$ is the cumulative distribution function for the price of Firm $i$, the profit of Firm $i$ when setting price $x$, is

$$\pi_i(x) = \alpha G_j(x)(x - c) + (1 - G_j(x))(1 - \alpha)(x - c)$$

Setting the derivative of this function equal to zero for $x \in (z', 1)$ and using the boundary conditions $G(z') = 0$ or $G(1) = 1$, we find

$$G_j(x) = \begin{cases} 
0 & \text{if } x < z' \\
\frac{(1-\alpha)(x-c) - \alpha(1-c)}{(1-2\alpha)(x-c)} & \text{if } z' \leq x \leq 1 \\
1 & \text{if } x > 1 
\end{cases}$$

which results in the profit $\alpha(1 - c)$ for each firm.

With probability $\gamma(1 - \gamma)$, $v_1 = 1$ and $v_2 = 0$, giving Firm 1 monopoly power over all of Segment 3 and profit of $(1 - c)(1 - \alpha)$. Thus, the expected profit of each firm when both use autoscaling is $(1 - c)(\gamma^2 \alpha + \gamma(1 - \gamma)(1 - \alpha))$. Though we allow for firms to choose to set capacity (rather than adopt autoscaling) even when autoscaling is available, we show in the Appendix that there exists a $\hat{c}$ such that both firms will choose autoscaling in equilibrium if $c \geq \hat{c}$. In the paper, we focus on both firms choosing autoscaling (i.e., $c \geq \hat{c}$), but show in the Appendix that our results also hold for $c < \hat{c}$, which results in only one firm choosing autoscaling.

Next, we study the effect of autoscaling on firms using the cloud and find how average prices change with the introduction of autoscaling.
4.2 Effect of Autoscaling on Firms’ Prices

Given the firms’ equilibrium strategies, we can examine how autoscaling will affect equilibrium prices in the event that entry costs are low enough such that both firms enter.

**Proposition 2** Suppose entry costs are such that both firms enter the market with or without autoscaling. The effect of autoscaling on average prices depends on $\gamma$ as follows:

- In the region for separated capacities (i.e., $\gamma > c$ and $\gamma(1 - \gamma) < c$), autoscaling decreases the average price set by each firm.

- In the region for overlapping capacities (i.e., $\gamma(1 - \gamma) > c$), autoscaling increases the average price set by each firm for low enough cost of capacity (i.e., $c < \frac{\gamma(1+\alpha(\gamma-1))}{1-\alpha} - 2\gamma^2$).

Proposition 2 shows demand uncertainty creates an important distinction from previous literature on capacity choice (e.g., Kreps and Scheinkman 1983), as it results in a capacity game that decreases average prices relative to the pricing game that arises with autoscaling. It may be expected that autoscaling should strictly reduce prices by allowing both firms to freely compete over all consumers without capacity constraints. However, Proposition 2 shows firms may actually increase their prices when they no longer have to commit to a fixed capacity under demand uncertainty. Therefore, the probability of a successful venture is critical in determining whether autoscaling increases or decreases prices; a result that is new to the literature.

In general, autoscaling has two opposing effects on price: First, autoscaling turns the cost of each server from a sunk cost to a cost that depends on the number of consumers served by the firm. Without autoscaling, firms do not consider the cost of servers in their pricing decision, as this cost is sunk. Thus, they receive no negative utility from serving a larger portion of the market and are more flexible to do so by decreasing price. However, with autoscaling, each additional customer adds an additional cost, resulting in firms having less incentive to decrease their price to get more customers compared to when costs were sunk. This is the positive effect of autoscaling on average prices.

Second, autoscaling can intensify competition between two firms, in cases where capacity constraints without autoscaling stopped firms from competing head to head. This is the negative effect of autoscaling on average prices.
Proposition 2 shows the effect of autoscaling on the average prices charged by each firm for different regions of Figure 5. In the region for separated capacities, the second effect of autoscaling is the dominant effect. In this region firms choose not to attack without autoscaling, since they restrict their capacity to dampen competition. The equilibrium choice of capacity results in both firms charging the monopoly price. Autoscaling removes this separation of targeted consumers and increases competition between the two firms, resulting in decreased average prices.

In the region for overlapping capacities, a low enough cost of capacity (i.e., \( c < \frac{\gamma(1+\alpha(\gamma-1))}{1-\alpha} - 2\gamma^2 \)) reduces the downside risk of excessive capacity and thus results in both firms setting maximum capacity; \( k_1 = k_2 = 1 - \alpha \). This means there are no capacity constraints preventing the firms from competing over price without autoscaling. Thus, the second and negative effect of autoscaling on average prices diminishes, since autoscaling does not intensify competition in this region. Therefore, the first effect is dominant and autoscaling increases average prices.\(^{14}\)

So far, we solved the pricing and capacity subgames conditional on both firms having had entered the market. Next, we consider firms’ choice of entering the market and study how autoscaling affects this decision. Autoscaling allows the firms to avoid over- or under-spending on computational capacity. However, there is also a strategic effect of autoscaling that can lead to dampened or intensified competition. In the following section, we analyze how these two effects of autoscaling combine to influence entry decisions.

### 4.3 Effect of Autoscaling on Entry Decisions

We now turn our attention to the entry decision. We start by deriving the expected profits for each equilibrium strategy given the price and capacity choices described in Section 4.1.

First, suppose both firms enter the market. In the absence of autoscaling, the profits depend on equilibrium capacities. With overlapping capacities, the higher capacity firm earns expected profit equal to \((1 - \alpha)((\gamma(1 - \gamma)) - c) + \gamma^2(1 - k^*)\) and the lower capacity firm earns expected profit equal to \( \gamma(1 - \gamma)k^* + \gamma^2((1-k^*)k^* - ck^*) \), where \( k^* \) is the capacity chosen by the lower capacity firm and is defined in the appendix. With separated capacities, each firm earns an expected

\(^{14}\)Note that for \( c > \frac{\gamma(1+\alpha(\gamma-1))}{1-\alpha} - 2\gamma^2 \) in the region for overlapping capacities, the two firms have different average prices without autoscaling. Since with both firms using autoscaling they both set the same price, evaluating the effect of autoscaling on average charged prices is not as straightforward as for \( c < \frac{\gamma(1+\alpha(\gamma-1))}{1-\alpha} - 2\gamma^2 \). Later in this section, we use consumer surplus as a proxy to average prices to study the effects of autoscaling in this region.
profit of $(\gamma - c)/2$. As stated previously, when both firms use autoscaling, each firm’s profit is 
$(1 - c)(\gamma^2 \alpha + \gamma(1 - \gamma)(1 - \alpha))$.

Now consider the case when only one firm enters the market. Without autoscaling, the firm will be a monopolist, optimally choosing $k = 1 - \alpha$ and earning expected profit $(\gamma - c)(1 - \alpha)$, if $\gamma > c$, and optimally choosing $k = 0$ to earn zero profit otherwise. In the presence of autoscaling, a single entrant earns a profit of $\gamma(1 - c)(1 - \alpha)$.

Given the firms’ equilibrium strategies and their expected profits, we can derive their entry decisions. We summarize the entry decisions with and without autoscaling in the following lemma.

**Lemma 1** Firms’ entry decisions depending on the presence of autoscaling are as follows.

- **Without Autoscaling:** 
  - If $F > (\gamma - c)(1 - \alpha)$, then there is no entry. Otherwise:
    - In the region for overlapping capacities (i.e., $\gamma(1 - \gamma) - c > 0$), both firms enter if $F < \gamma(1 - \gamma)k^* + \gamma^2 \frac{(1-k^*)k^*}{1-\alpha} - ck^*$. Otherwise only one firm enters.
    - In the region for separated capacities (i.e., $\gamma > c$ and $\gamma(1 - \gamma) - c < 0$), both firms enter if $F < (\gamma - c)/2$. Otherwise only one firm enters.

- **With Autoscaling:** 
  - If $F > \gamma(1 - c)(1 - \alpha)$, then there is no entry. Both firms enter the market if $F < \text{Max}[(1 - c)(\gamma^2 \alpha + \gamma(1 - \gamma)(1 - \alpha)), \frac{c(\alpha(\gamma^2-1) + (\alpha(-\gamma)+\alpha-1))^2}{4(\alpha-1)^3(\gamma^2(c-1))}]$. Otherwise, only one firm enters.

Lemma 1 is graphically depicted in Figures 6a and 6b. Note that in Figure 6a, there is a jump in the size of the region with double entry as $\gamma$ grows. This is because when $\gamma$ becomes sufficiently large, firms change their strategies from overlapping capacities to separated capacities.

By comparing the results presented in Lemma 1, we can determine the effect of autoscaling on firm entry. We next examine how autoscaling affects whether multiple firms enter the market.

**Proposition 3** When probability of success, $\gamma$, is sufficiently large and cost of entry, $F$, is moderate (region A in Figure 6c), fewer firms enter the market in equilibrium when autoscaling is available than when it is not.

Proposition 3 finds the counter-intuitive result that autoscaling can decrease market entry. Though autoscaling has a downside risk reducing effect, it also has a competition intensifying
In other words, autoscaling makes it less costly for a firm to find out if it has a successful venture on its hands, but also allows a firm to fight aggressively for consumers in Segment 3. To further explain these effects and when each is dominant, we consider the three potential outcomes if both firms enter.

When both firms enter, there is a $\gamma(1-\gamma)$ probability that a firm finds itself a monopolist, a $\gamma^2$ probability that a firm finds itself competing head-to-head, and a $1-\gamma$ probability that a firm finds $v_i = 0$. In the former case, autoscaling weakly benefits firms because they are assured of having the computational capacity to satisfy the demand of all consumers in Segment 3. Without autoscaling, firms acknowledging the downside risk choose $k_i \leq 1 - \alpha$ and thus cannot satisfy all demand when given monopoly power over all consumers in Segment 3. This is the problem that startup Befunky experienced without autoscaling in the earlier example and represents the positive demand satisfaction effect of autoscaling. In the latter case, autoscaling weakly benefits firms because it prevents them from over-purchasing capacity. Without autoscaling, $k_i \geq 0$ and thus firms have excess computational capacity when $v_i = 0$. This is the positive downside risk reducing effect of
autoscaling. However, autoscaling weakly disadvantages firms if they compete head-to-head. With autoscaling, firms are assured of having computational capacity to satisfy demand of all consumers in Segment 3. As such, autoscaling intensifies competition. Without autoscaling, the fact that \( k_i \leq 1 - \alpha \) allows the firms to include a more profitable *retreating* price in the equilibrium mixed strategy. In this case, the demand satisfaction effect actually leads to the negative *competition intensifying* effect of autoscaling.

The *downside risk reducing* effect is most dominant when \( \gamma \) is low. The *demand satisfaction* effect is most dominant when \( \gamma(1 - \gamma) \) is high (i.e., moderate \( \gamma \)) and the *competition intensifying* effect is most dominant when \( \gamma \) is high. A high \( \gamma \) increases the probability of competition and also makes it such that firms without autoscaling choose capacities such that this competition is avoided. Therefore, autoscaling decreases market entry when the probability of a successful venture is sufficiently high.

We next consider the effect of autoscaling on participation in the market by any of the firms, i.e., how autoscaling affects whether at least one of the two firms enters.

**Proposition 4** *Autoscaling increases the range of entry costs, \( F \), such that at least one firm enters the market.*

Proposition 4 confirms common intuition that autoscaling can make entry more attractive for at least one firm. The reason is that it allows firms with uncertain likelihood of success to incur the cost of computational needs after demand is realized. This highlights the *downside risk reducing* effect of autoscaling. Without autoscaling, firms have to invest in the cost of entry \( F \) and the cost of computational capacity \( c_k \) prior to realizing whether the venture will be successful. Autoscaling increases the range of entry costs for which at least one firm enters by \( c(1 - \gamma)(1 - \alpha) \). Thus, the higher the cost of capacity and the lower the probability of success, the more effective autoscaling will be in guaranteeing that the market will be served by at least one firm. Propositions 3 and 4 suggest that to find the effect of autoscaling on the number of new entrants, we must consider the probability of success as well as the cost of entry, which goes against the intuition that autoscaling always facilitates market entry.

The combined results of Propositions 3 and 4 are graphically depicted in Figure 6c. The solid (dotted) lines in the figure are the thresholds on \( F \) without (with) autoscaling, described in
Lemma 1. As shown in this figure, there are four regions of interest. In region A, autoscaling decreases entry due to the competition intensifying effect. Autoscaling allows one firm to be a monopolist because the other firm cannot profitably enter given the anticipated level of competitive intensity. In regions B and D, the market will not be served by either firm unless there is autoscaling. In region C, a firm would have monopoly power because the downside risk of capacity pre-purchase makes it unprofitable for a second entrant, but autoscaling alleviates this effect and results in competing firms entering the market.\(^{15}\)

Next, we consider the effect of autoscaling on the expected profit of the two firms entering the market.

**Corollary 1** Suppose entry costs are such that both firms enter the market with or without autoscaling. Autoscaling can create a prisoner’s dilemma, such that both firms use autoscaling even though they earn greater expected profit in the absence of autoscaling.

As noted previously, the competition intensifying effect can outweigh the demand satisfaction effect and the downside risk reducing effect for sufficiently high $\gamma$. If $F$ is sufficiently low, both firms will choose to enter with or without autoscaling. Furthermore, as shown in Corollary 1, they both choose autoscaling in equilibrium. Interestingly, this leads to a prisoner’s dilemma situation where the firms’ adoption of autoscaling results in diminished expected profitability of both firms. This result is depicted in Figure 7. The dashed lines in Figure 7 correspond to regions when autoscaling is not available (from Figure 5), and show how autoscaling affects firms’ equilibrium profits in different regions. When the probability of success, $\gamma$, is very high, only one firm uses autoscaling while the competing firm can strategically limit its computational capacity to soften competition. Also, when $\gamma$ is sufficiently low, both firms use autoscaling, but due to the downside risk reducing effect of autoscaling, both firms get higher profits with autoscaling. However, a moderately high $\gamma$ creates a prisoner’s dilemma situation where the competition intensifying effect of autoscaling dominates the downside risk reducing effect, but the firms still use autoscaling. Therefore, both firms would be better off if autoscaling was not available in this region.

We now turn our attention to the impact of autoscaling on consumers.

\(^{15}\)Note that $\alpha < \frac{1}{2}$ is a necessary condition in the proof of Proposition 3; when $\alpha = \frac{1}{2}$, region A in Figure 6c disappears. In other words, autoscaling decreases market entry only in the existence of competition. Without competition (i.e., when $\alpha = \frac{1}{2}$ and size of Segment 3 equals zero), there are no downsides to using autoscaling, and thus autoscaling always increases market entry.
4.4 Effect of Autoscaling on Consumer Surplus

So far, we studied the effects of autoscaling on firms’ strategies and their profit. Autoscaling can increase price competition between firms. It can also increase market entry. Both of these effects, intuitively, should lead to higher surplus for consumers. However, autoscaling also changes the capacity cost from sunk cost at the time of pricing to variable cost. Therefore, as shown in Proposition 2, autoscaling can lead to higher average prices, and thus lower surplus, for consumers. Furthermore, as shown in Proposition 3, autoscaling can also increase the likelihood of a monopoly market. In this section, we study the effect of these opposing forces on consumer surplus.

Expected consumer surplus can be derived from calculating the difference between expected social welfare and combined expected firm profit. Social welfare is equal to the combined value consumers get (i.e., the number of purchases times a value of 1) minus the cost to deliver that value (i.e., $c$ times the computational capacity). Therefore, the expected consumer surplus can be written as

$$E[CS] = E[#\text{purchases}] - c \times E[\text{computational capacity}] - E[\pi_1] - E[\pi_2]$$  \hspace{1cm} (3)$$

where #purchases and computational capacity indicate the total number of consumers who purchase the product and the total computational capacity reserved by the firms, respectively. If there is only one firm in the market, in both cases with and without autoscaling, that firm sets the price to 1, resulting in zero consumer surplus. If both firms are in the market, consumer surplus depends on whether firms use autoscaling or set capacity. We present the values of consumer surplus derived
from Equation (3) in the Technical Appendix. Comparing across conditions, we have the following result.

**Proposition 5** For sufficiently small $F$, sufficiently small $c$ and a moderate value of $\gamma$, autoscaling decreases expected consumer surplus. If $F$ is sufficiently large, autoscaling has no effect on the expected consumer surplus. Otherwise, autoscaling increases the expected consumer surplus.

Proposition 5 shows the effect of autoscaling on consumer surplus. The result is also depicted in Figures 8 and 9. Figure 8 is analogous to Figure 5 with the contours from Figure 5 marked with dashed gray lines. As we can see, the region where autoscaling decreases consumer surplus is located in the region for overlapping capacities. Intuitively, since autoscaling increases average prices in this region, it decreases consumer surplus.

Figure 9 is analogous of Figure 6 with the contours from Figure 6 marked as dashed gray lines. The figure shows that autoscaling can only affect consumer surplus if the cost of entry is sufficiently low. In particular, unless both firms enter the market in at least one of the two cases (i.e., autoscaling or no autoscaling available), consumer surplus will be zero; therefore, if $F$ is not low enough, autoscaling can have no effect on consumer surplus. Figure 9 also shows that when autoscaling increases market entry to two firms, it also increases expected consumer surplus.
5 Extensions

So far, we studied the effect of autoscaling on prices, profits, market entry, and consumer surplus using a model of market entry for horizontally differentiated firms. In this section, we examine three extensions of the model to establish the robustness of our results and obtain new insights. First, in Section 5.1, we consider the case in which when firms do not succeed, consumers still have positive reservation value for their offerings \( V_l > 0 \). In the main model, we assumed an efficient rationing rule when allocating firms’ capacity to consumer demand. In Section 5.2, we relax that assumption by considering a proportional rationing rule. Finally, in Section 5.3, we look at the incentives of the service provider and endogenize the price of capacity \( c \).

5.1 Positive Low-State Value

In the main model, we assumed that consumers’ reservation value for a firm that does not succeed is \( v_i = 0 \). This assumption could seem strong as it gives monopoly power to the competitor. In this section, we relax this assumption to show the robustness of our results. We assume that the reservation value of consumers for each firm is \( V_h \) with probability \( \gamma \) and \( V_l \) with probability \( 1 - \gamma \), where \( V_h > V_l > 0 \). We show that our counter-intuitive result in Proposition 3 that autoscaling could lower entry becomes even stronger when \( V_l > 0 \).

Equilibrium Choices without Autoscaling

We use the same techniques as before to solve the pricing subgame, and to calculate the expected profit of Firm \( i \) for given capacities. The details of how we solve the pricing subgame are provided in the appendix. The expected profit of Firm \( i \), where \( i, j \in \{1, 2\} \), depends on the capacities of
the two firms as shown below, assuming $\alpha \leq k_i, k_j \leq 1 - \alpha$.

$$\begin{align*}
E(\pi_i) = \begin{cases}
  k_i(-c + \gamma V_h - \gamma V_l + V_l) & \text{if } k_i + k_j \leq 1 \\
k_i((k_i-1)((\gamma-1)V_l-\gamma V_h)-ck_j)) & \text{if } k_i < \frac{V_l}{V_h} k_j \text{ and } k_i + k_j \geq 1 \\
k_i(-ck_j+\gamma^2(-k_j)V_h+\gamma k_j V_l+\gamma k_j V_l+\gamma(-k_j-1)V_l+\gamma^2 V_h+V_l) & + \frac{k^2(\gamma^2(-V_h)+\gamma V_l-V_l))}{k_j} \gamma^2 V_l-\gamma^2 V_l-\gamma^2 k_j V_l+\gamma k_j V_l) & \text{if } \frac{V_l}{V_h} k_j < k_i < k_i \text{ and } k_i + k_j \geq 1 \\
-c k_i + (1 - \gamma) \gamma k_i \left(V_h - \frac{V_l(k_i+k_j-1)}{k_j}\right) & + (1 - k_j) (\gamma^2 V_h + (\gamma - 1)^2 V_l) + (\gamma - 1) \gamma (k_j - 1) V_l & \text{if } k_j < \frac{V_l}{V_h} k_i \text{ and } k_i + k_j \geq 1 \\
(k_j - 1)((\gamma - 1)V_l - \gamma V_h) - c k_i & & \text{if } k_j < \frac{V_l}{V_h} k_i \text{ and } k_i + k_j \geq 1 
\end{cases}
\end{align*}$$

Given the firms’ expected profits in the pricing subgame, we can calculate the equilibrium capacities by comparing the expected profits for each set of capacities. Assuming both firms enter the market, equilibrium capacity choices depend on the probability of success, the low state and high state values, and the cost of capacity as follows:

- If there is a low probability of a successful venture (i.e., $\gamma(V_h - V_l) + V_l < c$), then both firms choose $k_i = 0$.

- If there is a moderate probability of a successful venture (i.e., $(1 - \gamma) \gamma (V_h - V_l) > c$), then the firms choose overlapping capacities such that $k_i + k_j > 1$.

- If there is a high probability of a successful venture (i.e., $(1 - \gamma) \gamma (V_h - V_l) < c < \gamma(V_h - V_l) + V_l$), then the firms choose separating capacities with the unique symmetric equilibrium being $k_i = k_j = 1/2$.

It is interesting to note that as $V_l$ increases, the region in which firms use separated capacities grows. The region for separated capacities is given by $(1 - \gamma) \gamma (V_h - V_l) < c < \gamma(V_h - V_l) + V_l$. Since we have $\frac{\partial((1-\gamma)\gamma(V_h-V_l))}{\partial V_l} < 0$ and $\frac{\partial(\gamma(V_h-V_l)+V_l)}{\partial V_l} > 0$, this region becomes larger as $V_l$ increases. Intuitively, this is because increasing $V_l$ increases direct competition between a high-value firm and a low-value firm when their capacities overlap. To avoid this competition, firms are more likely to choose separated capacities and gain monopoly pricing power for higher $V_l$. Since autoscaling breaks the firms’ ability to dampen competition through limited capacity, as we see in the next
section, the competition intensifying effect of autoscaling becomes stronger as $V_l$ increases, and, therefore, autoscaling lowers market entry in a larger region.

**Equilibrium Choices with Autoscaling**

When autoscaling is available, we show that one or two firms use autoscaling. The analysis when $V_l > 0$ is similar to the analysis in the main model, and, hence, is relegated to the appendix. The expected profit of Firm $i$ when both firms use autoscaling is

$$E(\pi_{AA}) = -\alpha c + (2\alpha - 1)\gamma^2(V_h - V_l) + (\alpha - 1)\gamma(V_l - V_h) + \alpha V_l$$

**Effect of Low-State Value on Entry**

By comparing the cost of entry, $F$, to firms' profits, we determine how many firms, if any, would choose to enter the market for any given $F$. Figure 10 shows the effect of increasing $V_l$ on entry. As in Figure 6c, the shaded region A is where autoscaling decreases entry from two firms to one firm. In regions B and D neither firm enters the market unless autoscaling is available. Finally, region C is where autoscaling increases entry from one firm to two firms.

As $V_l$ increases, the shaded region in Figure 10 where autoscaling decreases entry from two firms to one firm expands. This is because increasing $V_l$ results in the expansion of the region for separated capacities without autoscaling. Thus, the region where the *competition intensifying* effect of autoscaling is dominant expands as the low-state value increases, decreasing market entry with autoscaling.

Note that regions B and D in Figure 10 disappear for $V_l > c$. A single entrant using autoscaling always sells to all possible $(1 - \alpha)$ customers when the low-state value is greater than the unit cost of capacity. The reason for this is that marginal cost of selling to each customer becomes less than the charged price, regardless of whether the firm realizes low-state or high-state value. Similarly, without autoscaling, a single entrant would set its capacity to $1 - \alpha$ when $V_l > c$. Therefore, for a single entrant, the profits and entry conditions without autoscaling become the same as those with autoscaling. As such, for $V_l > c$, autoscaling does not affect the condition for which at least one firm enters the market. In other words, Proposition 4, which states autoscaling increases the range of entry costs for which at least one firm enters, holds only for $0 \leq V_l < c$.  

32
Finally, Figure 10 shows $V_l$ must be low enough for autoscaling to increase entry, since region C only exists for low-state values that are not too high. Intuitively, a high $V_l$ means less difference between the good and bad outcome when realizing value, which reduces the level of uncertainty faced by firms. Therefore, as $V_l$ increases, the effective uncertainty of firms decreases, thereby diminishing the downside risk reducing effect of autoscaling. This in turn stops autoscaling from causing more entry into the market.

5.2 Proportional Rationing

In Section 4.1, we assumed an efficient rationing rule such that, when demand exceeds capacity, Segments 1 and 2 are served before Segment 3. In this extension, we check the robustness of our findings with respect to the rationing rule and solve the model using proportional rationing. We show that our results continue to hold when consumers across all segments arrive uniformly.

Suppose that after prices are set, consumers in Segment 3 prefer Firm $i$ to Firm $j$. Firm $i$’s capacity is allocated simultaneously to Segment 3 and Segment $i$, where $i \in \{1, 2\}$. With proportional rationing, the ratio of capacity allocated to each of these segments is relative to the size of that segment:

$$\text{Capacity Allocated to Segment } i = \frac{\text{Size of Segment } i}{\text{Sum of Sizes of Segments } i \text{ and } 3} \times k_i = \frac{\alpha}{1 - \alpha} k_i$$

Thus, of the $k_i$ available capacity for Firm $i$, $\frac{\alpha}{1 - \alpha} k_i$ is used by Segment $i$ and $\frac{1 - 2\alpha}{1 - \alpha} k_i$ is used by Segment 3. Note that unlike what we found with efficient rationing, Segment $i$ is not fully satisfied and there are $\alpha(1 - \frac{k_i}{1 - \alpha})$ consumers in this segment that are not served.
Once Firm $i$’s capacity is full, the residual demand of Segment 3 (i.e., $(1 - 2\alpha) - \frac{1 - 2\alpha}{1 - \alpha} k_i$) and the demand from Segment $j$ is satisfied by Firm $j$, provided it has available capacity. Thus, as long as $k_j < (1 - 2\alpha) - \frac{1 - 2\alpha}{1 - \alpha} k_i + \alpha$, each of the two firms can sell up to its full capacity without overlapping with the competitor. Otherwise, for $\frac{1 - 2\alpha}{1 - \alpha} k_i + k_j > 1 - \alpha$, it is not possible for both firms to sell their maximum capacity. In this case, a pure strategy equilibrium for the pricing subgame does not exist and both firms use mixed strategy pricing.

Note that the condition for overlapping capacities, $\frac{1 - 2\alpha}{1 - \alpha} k_i + k_j > 1 - \alpha$, occurs for a greater range of $k_i$ and $k_j$ compared to the overlapping capacities condition with the efficient rationing rule, $k_i + k_j > 1$. In other words, with proportional rationing, there are certain capacities for which $k_i + k_j < 1$ and firms use mixed strategy pricing.

Equilibrium Choices without Autoscaling

Using the same methods as before, we can solve the pricing subgame for given capacities with the new rationing rule. Details of the analysis are provided in the appendix. The expected profit of Firm $i$ when $\alpha < k_i, k_j < 1 - \alpha$ is as follows.

$$E(\pi_i) = \begin{cases} 
  k_i(\gamma - c) & \text{if } k_i < -\alpha - \frac{(1 - 2\alpha)k_j}{1 - \alpha} + 1 \\
  \frac{\gamma^2(-\alpha(\alpha + 2k_j - 2) + k_j - 1)}{\alpha - 1} - ck_i - (\gamma - 1)\gamma k_i & \text{if } k_i > -\alpha - \frac{(1 - 2\alpha)k_j}{1 - \alpha} + 1 \text{ and } k_i < k_j < -\frac{(\alpha - 1)^2}{2\alpha - 1} - k_i \\
  k_j \left( \frac{\gamma^2(-\alpha(\alpha + 2k_i - 2) + k_i - 1)}{\alpha - 1} - c - (\gamma - 1)\gamma \right) & \text{if } k_j > -\alpha - \frac{(1 - 2\alpha)k_j}{1 - \alpha} + 1 \text{ and } k_j < k_i < -\frac{(\alpha - 1)^2}{2\alpha - 1} - k_j \\
  k_i \left( \frac{\gamma^2(-\alpha(\alpha + 2k_j - 2) + k_j - 1)}{\alpha - 1} - c - (\gamma - 1)\gamma \right) & \text{if } k_j > k_i > -\frac{(\alpha - 1)^2}{2\alpha - 1} - k_j \\
  \frac{\gamma^2(-\alpha(\alpha + 2k_j - 2) + k_j - 1)}{\alpha - 1} - ck_i - (\gamma - 1)\gamma k_i & \text{if } k_i > k_j > -\frac{(\alpha - 1)^2}{2\alpha - 1} - k_i \\
\end{cases}$$

Comparing the expected profits for each set of capacities, we find the equilibrium capacities:

- If $\gamma < c$, then firms set $k_i = k_j = 0$.
- If $\gamma(1 - \gamma) > c$, then the firms choose overlapping capacities such that $\frac{1 - 2\alpha}{1 - \alpha} k_i + k_j > 1 - \alpha$.
- If $\gamma(1 - \gamma) < c$ and $\gamma > c$, then the firms choose separated capacities with the unique
symmetric equilibrium being \( k_i = k_j = \frac{(\alpha - 1)^2}{2 - 3\alpha} \).

Thus, the regions for separated and overlapping capacities have the same boundaries as in Proposition 1, when the efficient rationing rule was used. The difference from using a proportional rationing rule only appears in the capacities chosen within each region, not the size of each region.

**Equilibrium Choices with Autoscaling**

When both firms use autoscaling, the rationing rule does not affect the outcome, since neither firm has a capacity constraint. Therefore, each firm’s profit is \( \pi_{aa} = (1 - c) (\alpha \gamma^2 + (1 - \alpha)(1 - \gamma)\gamma) \).

In the appendix, we analyze the case when one firm uses autoscaling and the other chooses a fixed capacity. This case is affected by the rationing rule, which determines how much of the capacity of the firm not using autoscaling is allocated to consumers in Segment 3. We find that the profits are \( \pi_{an} = \frac{1}{2}(1 - \alpha)(-2\gamma c + c + \gamma) \) for the firm using autoscaling and \( \pi_{na} = \frac{(\alpha - 1)^2(\alpha - \gamma)^2}{4(1 - 2\alpha)\gamma(1 - c)} \) for the firm choosing fixed capacity.

**Effect of Autoscaling on Entry with Proportional Rationing**

We compare the cost of entry, \( F \), to firms’ profits, finding the number of firms that choose to enter the market for any given \( F \). We prove in the appendix that there exists a cutoff \( \tilde{\gamma} \), such that for \( \gamma > \tilde{\gamma} \) autoscaling decreases the range of entry costs for which both firms enter the market. This result is similar to what we found about the effect of autoscaling on entry when the efficient rationing rule was applied. Also, the rationing rule does not affect profits and entry conditions when only one firm enters the market. Thus, both Propositions 3 and 4 hold with proportional rationing.

Figure 11 shows the effect of autoscaling on market entry with proportional rationing. In region A, autoscaling decreases entry from 2 to 1 firms. In region C, autoscaling increases entry from 1 to 2 firms. In regions B and D, autoscaling increases entry from 0 to 1 and 2 firms respectively. Comparing Figure 11 with Figure 6c, where the efficient rationing rule was used, we see that the insights of our model for market entry remain the same with proportional rationing.
5.3 Service Provider Incentives

To focus on firms’ competition, in the main model, we assumed that the cloud provider’s decisions were exogenous. In practice, there are many parameters, still exogenous to our model, that affect a cloud provider’s decision on how to price capacity ($c$), and whether to offer autoscaling. For example, a cloud provider may decide to offer autoscaling, or change price $c$, because other cloud providers are doing so. Furthermore, cloud providers have clients from a wider range of industries with different $F$’s, $\alpha$’s and $\gamma$’s, but for practical purposes (e.g., cloud capacities are often sold as an off-the-shelf product), they may have to offer the same, or similar, prices and functionalities across many industries. This would again limit the cloud provider’s ability to optimize $c$ and the choice of offering autoscaling for a given $F$, $\alpha$ and $\gamma$. For these reasons, when studying firms’ competition, an exogenous model for the cloud provider might be a better approximation of the real world. However, from a theoretical point of view, it is interesting to see if and how endogenizing the cloud provider’s decisions affects our results. We show that our counter-intuitive result in Proposition 3, that autoscaling could lower market entry, continues to hold when we endogenize the cloud provider’s decisions. The analysis and proofs from this section are relegated to the appendix.

Provider’s Choice of Capacity Price

In this section, we endogenize the price of capacity, by allowing the cloud provider to choose $c$ as a function of $F$, $\alpha$ and $\gamma$. The cloud provider’s profit is

$$\pi_{CP} = c \times \text{(Purchased Capacity)}.$$
Figure 12: Effect of autoscaling on entry with endogenous \( c \). In region A, autoscaling decreases entry from 2 to 1 firms. In region C, autoscaling increases entry from 1 to 2 firms.

When autoscaling is available, the provider faces a price-volume tradeoff between \( c \) and the number of entrants: increasing \( c \) increases the provider’s profit per unit of capacity but could also decrease market entry depending on \( F \). We show that the provider is indifferent between one firm or two firms entering the market when \( F \) is

\[
\hat{F} = \frac{\gamma(-2+\gamma+\alpha+\gamma-1)^2}{1-\alpha}.
\]

When the cost of entry is low (i.e., \( F < \hat{F} \)), the provider sets \( c = 1 - \frac{\gamma(2\gamma-1)F}{\gamma(\alpha(2\gamma-1)-\gamma+1)} \) so that both firms enter the market; otherwise, for \( \hat{F} \leq F \leq (1-\alpha)\gamma \), the price of capacity is optimally increased to \( c = 1 - \frac{F}{(1-\alpha)\gamma} \), resulting in only one firm entering the market. Finally, for \( F > (1-\alpha)\gamma \), neither firm would enter the market for any \( c \geq 0 \).

When autoscaling is not offered, similar to the case with autoscaling, the provider sets a low \( c \) for low \( F \) to allow both firms to enter, and increases \( c \) for higher \( F \) resulting in single entry. The analysis for the optimal price of capacity when autoscaling is not an option is presented in the appendix. Figure 12 shows the effect of autoscaling on market entry when \( c \) is endogenously chosen by the provider. This figure is a replication of Figure 6c, but with endogenous \( c \). Comparing the two figures, we see that regions B and D from Figure 6c, in which autoscaling increased entry from 0 to 1 or 2 firms, disappear when \( c \) becomes endogenous. As stated in Proposition 4, when \( c \) was given exogenously, autoscaling increased the range of \( F \) for which at least one firm entered the market. The reason for this finding was the downside risk reducing effect of autoscaling, allowing a
single entrant to only pay for capacity when its demand is high. However, when \( c \) is endogenous, the provider sets the price of capacity sufficiently low (when autoscaling is not available) so that still one firm enters. In other words, when autoscaling is not available and \( c \) is endogenous, the provider absorbs the firms’ downside risk of failure by lowering the price of capacity to encourage entry. Therefore, regions B and D disappear, and our result in Proposition 4, where autoscaling increases entry from 0 to 1 or 2 firms, does not hold for endogenous \( c \). Finally, similar to what we had in Figure 6c, Figure 12 also shows that for high enough \( \gamma \), autoscaling decreases the range of \( F \) for which both firms enter the market (region A). Thus, our counter-intuitive finding in Proposition 3 which stated that autoscaling could lead to fewer firms entering the market still holds when \( c \) is endogenous.

**Effect of Autoscaling on Provider’s Profit**

In this section, we find how autoscaling affects the profit of the cloud provider when the provider chooses the price of capacity. Since we do not find closed form solutions for the optimal price under all conditions, we present the results of a numerical comparison of provider profits with and without autoscaling in Figure 13. In Region 1 of Figure 13, autoscaling has no effect on the provider’s profit. This is because with or without autoscaling, only one firm enters the market in Region 1 and the provider makes a profit of \((1 - \alpha)\gamma - F\). Thus, for autoscaling to change provider’s profit, the cost of entry must be low enough so that two firms enter the market.

In Region 2, for medium \( \gamma \) and low enough \( F \), autoscaling increases the provider’s profit. In Region 3, for higher \( \gamma \), autoscaling decreases the provider’s profit. Intuitively, medium probabilities of success result in high demand uncertainty. Thus, for medium \( \gamma \), the *downside risk reducing* and *demand satisfaction* effects of autoscaling are dominant, allowing autoscaling to facilitate market entry and benefit the provider. For high probabilities of success, autoscaling results in a high probability of direct competition between the two firms, which means the provider would in turn have to decrease the price of capacity substantially to create incentive for the firms to enter the market. However, without autoscaling, firms choose capacities that do not overlap for high \( \gamma \),

---

16 We should note, however, that there is a caveat in this argument. Here, we are assuming that the provider optimizes \( c \) as a function of \( \gamma \), \( \alpha \) and \( F \). In other words, we allow the provider to set different prices for firms with different \( \gamma \)'s, \( \alpha \)'s and \( F \)'s in the market. If the provider has to keep the price of capacity the same for all (or most) firms, because of other practical concerns (e.g., because it cannot perfectly price discriminate), then the result of Proposition 4 would hold again.
Figure 13: Effect of autoscaling on cloud provider’s profit. Region 1 denotes autoscaling does not change providers’ profit. Region 2 denotes autoscaling increases provider’s profit. Region 3 denotes autoscaling decreases provider’s profit.

allowing the provider to choose a higher $c$. Thus, when $\gamma$ is high, autoscaling prevents the provider from increasing the price of capacity and earning more profit.

Figure 13 also shows the effect of changing $\alpha$ on the provider’s profits. As $\alpha$ increases, the expected purchased capacity of a single entrant, $(1 - \alpha)$, decreases, making single entry a less attractive option for the provider and shrinking Region 1.

Figure 13 confirms the intuition that for markets with considerable demand uncertainty, represented by medium $\gamma$ in Region 2, autoscaling helps increase the provider’s profit and impacts the market. Note that while autoscaling may not benefit the provider in certain industries with little uncertainty, the provider makes the choice of offering autoscaling based on the aggregate outcome of multiple industries with varying levels of $\gamma$ and $F$. Also, other factors such as competition among providers can be expected to contribute to strengthening the benefit of autoscaling for cloud providers. Observations from the cloud computing industry indicate that autoscaling is offered by all major cloud providers and is not an exclusive offer for only certain markets.\footnote{www.knowthecloud.com/Providers/auto-scaling-providers.html, accessed February 2017.} A thorough examination of such contributing factors affecting providers’ decisions is a potentially rich topic for future research on cloud computing, but outside of the scope of this paper.

In the appendix, we show that even when the cloud provider is allowed to choose two separate capacity prices, $c_k$ for fixed pre-purchased capacity and $c_a$ for autoscaling capacity, the insights are similar to those from Figure 13.
6 Discussion

The emergence of cloud computing and its feature of autoscaling has the potential to influence the entry of web-based firms into new markets. Before autoscaling was available, firms needed to invest heavily in computational capacity before entering a new market so that they could serve highly unpredictable demand levels. The uncertainty in demand meant that the capacity chosen by firms could be more or less than the actual needs, resulting in extra costs or unfulfilled demand. However, autoscaling allows the new entrants a flexible capacity such that the cloud provider will designate the specific computational resources as the actual demand and required resources are realized. This feature can help new entrants to the market pay only for the capacity that is required to meet their demand, reducing the disadvantages of uncertainty in demand. In this paper, we find the effects of autoscaling on entry decisions of web-based firms, their prices, their capacity decisions, and their profits.

Our model shows how the demand satisfaction effect of autoscaling, in which firms can be assured of having capacity to satisfy demand, turns out to be a double-edged sword in that it frees competing firms to aggressively pursue customers. Our research identifies this competition intensifying effect of autoscaling and establishes the conditions under which this effect will outweigh the positive effects of autoscaling. This has several important implications for web-based firms and their product launch strategies.

The results can help guide pricing decisions by firms launching a new product with autoscaling. We find that autoscaling will increase the average launch prices if the probability of a successful venture is not too high. In this case, autoscaling decreases a firm’s costs, but these savings are not passed on to consumers because autoscaling converts a sunk fixed cost into a variable cost that a firm incurs if and only if it attracts more customers. As a consequence, price competition is dampened by autoscaling in the case that entrants would otherwise choose excessive capacities. However, if the probability of a successful venture is sufficiently high, firms would optimally limit their capacities to dampen competition in the absence of autoscaling. In this case, autoscaling gives the competing firms freedom to aggressively pursue customers with price and thus can decrease average prices.

The trade-off between the intensified competition and the reduced downside risk of failure affects
firms’ decisions on whether to enter the market. In particular, when the probability of success is sufficiently high, the increased competition effect of autoscaling dominates the decreased downside risk of failure. As a result, while autoscaling facilitates the entry for one firm, it deters a second firm from simultaneously entering the market. Our findings expand the market entry literature by looking at the topic from the new angle of capacity choice before versus after demand is realized.

Our research also guides managers of web-based firms with their response to the introduction of autoscaling and cloud computing. The research shows if the probability of success is moderately high, autoscaling leads to a prisoner’s dilemma situation where both entrants adopt autoscaling, but they would have been better off if both had avoided it. Our findings suggest that in industries where cost of entry is relatively low, the introduction of the autoscaling feature in cloud computing can reduce the profit of new entrants to the market even though it decreases their downside risk of failure.

Autoscaling could also have positive or negative effects on consumer surplus. On the one hand, it can increase price competition between the firms and facilitate their entry, both of which lead to higher consumer surplus. On the other hand, autoscaling can lower consumer surplus by dampening price competition or by reducing the number of firms that enter the market. Our results show that when cost of entry is low and probability of success is moderate, autoscaling decreases expected consumer surplus.

Overall, our results highlight how web-based firms should use information on cost of entry, probability of success, cost of capacity and level of differentiation to evaluate their entry decisions, capacity commitments and pricing strategies in the presence of autoscaling. Our research is one of the first to consider the marketing aspects of cloud computing and autoscaling. With the rapid adoption of cloud computing by firms across different industries, marketing and economics research on the cloud can be a rich and important topic of study for future research.

References


A Technical Appendix

A.1 Proof of Proposition 1

If $\gamma < c$, then expected profit is strictly decreasing in capacity for any capacity chosen by the competition. Therefore, each firm optimally chooses zero capacity.

If $\gamma > c$, then for any $k_2 < 1 - \alpha$, expected profit is increasing in $k_1$ for any $k_1 < 1 - k_2$. Therefore, we can rule out any $k_1 + k_2 < 1$. First consider $\gamma(1 - \gamma) < c$. Suppose there were a symmetric equilibrium in which $k_1 + k_2 > 1$. By definition, this implies that $k_1 > 1/2$ which
means Firm 1 earns greater expected profit by deviating downward. Therefore, the only potential symmetric equilibrium requires \( k_1 + k_2 = 1 \). If Firm 1 deviates upward, its expected change in profit is \( \gamma (1 - \gamma) - c < 0 \).

If \( \gamma (1 - \gamma) > c \), then \( k_1 + k_2 = 1 \) is no longer an equilibrium because either firm can profitably deviate to harvest the potential of monopoly power. We can therefore focus our attention on \( k_1 + k_2 > 1 \). If \( k_2 > k_1 \), then Firm 2’s expected profit is strictly increasing in \( k_2 \) until \( k_2 = 1 - \alpha \) and is strictly decreasing for any \( k_2 > 1 - \alpha \). If \( k_2 > k_1 \), Firm 1’s expected profit for any \( k_1 < 1 - \alpha \) is given by \( \gamma (1 - \gamma) + \gamma^2 \frac{(1-k_1)k_1}{k_2} - k_1 c \), which is concave in \( k_1 \) and maximized at \( k_1 = \frac{\gamma (1-a(1-\gamma)) - c(1-\alpha)}{2\gamma} \). Note this value of \( k_1 \) is in fact less than \( 1 - \alpha \) if and only if \( c > \frac{\gamma (1+a(\gamma-1))}{1-a} - 2\gamma^2 \). We verify that Firm 2 cannot benefit from a global deviation undercutting \( k_1 \) at this level. Supposing that Firm 2 deviates to a lower capacity than Firm 1, its profit would become
\[
\tilde{k}_2 = c (1 + \alpha) + c\gamma (2 + 2\alpha(-1 + \gamma) - \gamma) + \gamma^2 (-1 + \alpha + \gamma - 2\alpha\gamma + (-2 + \alpha)\gamma^2).
\]

It is easily shown that any values of \( c \) that allow for \( \tilde{k}_2 < k_1 \) will result in Firm 2’s profit at \( \tilde{k}_2 \)
\[
\left(\frac{c^2(-1+\alpha)+c\gamma(2-\gamma+2(-1+\gamma)\alpha)+\gamma^2(-1+\gamma+\gamma^2(-2+\alpha)+\alpha-2\gamma\alpha)}{8\gamma^4(c(-1+\alpha)+\gamma(1+(-1+\gamma)\alpha))}\right)^2
\]
to be less than Firm 2’s profit at our equilibrium.

If \( c < \gamma (1+a(\gamma-1)) - 2\gamma^2 \), then neither firm benefits from deviating from \( k_j = 1 - \alpha \). Thus, \( k^* \) from Proposition 1 is defined as \( k^* = Min\left[\frac{\gamma (1-a(1-\gamma)) - c(1-\alpha)}{2\gamma^2}, 1 - \alpha\right] \).

### A.2 Proof for Firms’ Decision to Use Autoscaling

In this section, we consider the firms’ choice of using autoscaling when autoscaling is available to see how many firms, if any, use autoscaling.

We begin with the case when only one firm uses autoscaling. Without loss of generality, we assume that Firm 2 adopts autoscaling and Firm 1 chooses capacity \( k_1 \). We start by showing that this game does not have a pure strategy pricing equilibrium. Assume for sake of contradiction that the firms use prices \( p_1 \) and \( p_2 \) in a pure strategy equilibrium. If \( p_1 \neq p_2 \), then the firm with a lower price can benefit from deviating by increasing its price to \( \frac{p_1 + p_2}{2} \). If \( p_1 = p_2 \), then Firm 2 can benefit from deviating by decreasing its price to \( p_2 - \varepsilon \), for sufficiently small \( \varepsilon \), to acquire all consumers in
Segment 3. Therefore, a pure strategy equilibrium cannot exist.

Next, we find a mixed strategy equilibrium for this game. Provided $k_1 \leq 1 - \alpha$, Firm 2 can choose to attack with a price that clears its capacity or retreat with a price 1 that harvests the value from the $1 - k_1$ consumers that Firm 1 cannot serve due to its capacity constraint. Let $z''$ be the price at which Firm 2 is indifferent between attacking to sell to $1 - \alpha$ consumers at price $z''$ and retreating to sell to $1 - k_1$ consumers at price 1. We have $(1 - k_1)(1 - c) = (1 - \alpha)(z'' - c)$ which gives us $z'' = \frac{\alpha c - ck_1 + (k_1 - 1)}{\alpha - 1}$. In equilibrium, both firms use a mixed strategy with prices ranging from $z''$ to 1. Suppose that $H_i(.)$ is the cumulative distribution function of prices used by Firm $i$. The profit of Firm 1 earned by setting price $x$ is

$$\pi_1(x) = H_2(x)\alpha x + (1 - H_2(x))k_1 x - k_1 c$$

Using equilibrium conditions, we know that the derivative of this function with respect to $x$ must be zero for $x \in (z'', 1)$. By solving the differential equation we get

$$H_2(x) = \begin{cases} 0 & \text{if } x < z'' \\ \frac{k_1(-ck_1+(-1+k_1)+\alpha(c-x)+x)}{(-1+\alpha)(\alpha-k_1)x} & \text{if } z'' \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Similarly, the profit of Firm 2 earned by setting price $x$ is

$$\pi_2(x) = H_1(x)(1 - k_1)(x - c) + (1 - H_1(x))(1 - \alpha)(x - c)$$

By setting the derivative with respect to $x$ to zero for $x \in (z'', 1)$ and solving the differential equation, we get

$$H_1(x) = \begin{cases} 0 & \text{if } x < z'' \\ \frac{-ck_1+(-1+k_1)+\alpha(\alpha-x)+x}{(k_1-\alpha)(x-c)} & \text{if } z'' \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Prior to the pricing game, the optimal capacity $k_1$ that maximizes expected profit for Firm 1
is given by:

\[ k_{na}^* = \frac{\gamma(1 - \alpha + \alpha\gamma) - c(1 + \alpha(-1 + \gamma^2))}{2(1 - c)\gamma^2} \]

We may now examine the equilibrium adoption of autoscaling. The payoffs from each possible firm choice of autoscaling or capacity \( k \) are summarized in Table 2.

<table>
<thead>
<tr>
<th>Firm 2 uses Autoscaling</th>
<th>Firm 2 uses capacity ( k )</th>
</tr>
</thead>
</table>
| **Firm 1 uses Autoscaling** | \( \pi_1 = \pi_{AA} = (1 - c)(\gamma^2\alpha + \gamma(1 - \gamma)(1 - \alpha)) \)  
|                           | \( \pi_2 = \pi_{AA} = (1 - c)(\gamma^2\alpha + \gamma(1 - \gamma)(1 - \alpha)) \)  
| **Firm 1 uses capacity \( k \)** | \( \pi_1 = \pi_{AN} = \gamma^2(1 - k)(1 - c) + \gamma(1 - \gamma)(1 - \alpha)(1 - c) \)  
|                           | \( \pi_2 = \pi_{NA} = \gamma^2(1 - k)(1 - c) + \gamma(1 - \gamma)(1 - \alpha)(1 - c) \)  
|                           | Profits are the same as in Section 4.1  

Table 2: Payoffs from autoscaling adoption strategies assuming both firms enter the market

**Proof That At Least One Firm Uses Autoscaling**

We show that both firms not using autoscaling cannot be an equilibrium. As seen in Table 2, the expected profit of using autoscaling, when the opponent uses capacity \( k \), is

\[ \bar{\pi}_{AN}(k) = \gamma^2(1 - k)(1 - c) + \gamma(1 - \gamma)(1 - \alpha)(1 - c). \]

We compare the profit of each firm when neither use autoscaling to the profit from deviating to autoscaling (i.e., \( \bar{\pi}_{AN}(k) \)).

First, consider when neither firm uses autoscaling and firms set separated capacities. Each firm’s profit is \( \frac{1}{2}(\gamma - c) \). A firm’s profit from unilaterally using autoscaling is \( \bar{\pi}_{AN}(\frac{1}{2}) = \frac{1}{2}\gamma(2\alpha(\gamma - 1) - \gamma + 2)(1 - c) \). For all \( \alpha \leq \frac{1}{2} \), we have \( \bar{\pi}_{AN}(\frac{1}{2}) > \frac{1}{2}(\gamma - c) \). Thus, each firm has incentive to use autoscaling instead of remaining in the separated capacity equilibrium.

Next, consider the case when firms set overlapping capacities instead of using autoscaling. Based on the proof of Proposition 1, for \( c < \frac{\gamma(1 + \alpha(\gamma - 1))}{1 - \alpha} - 2\gamma^2 \) each firm sets capacity to \( 1 - \alpha \), earning a profit of \((1 - \alpha)((\gamma(1 - \gamma)) - c) + \gamma^2\alpha \), which is less than \( \bar{\pi}_{AN}(1 - \alpha) = \gamma(\alpha(2\gamma - 1) - \gamma + 1)(1 - c) \).

For \( c > \frac{\gamma(1 + \alpha(\gamma - 1))}{1 - \alpha} - 2\gamma^2 \), capacities are \( k_i = 1 - \alpha \) and \( k_j = k^* \). The resulting profits without autoscaling are \( \pi_{NNi} = \frac{1}{2}((\alpha - 1)c + \gamma(\alpha(\gamma - 1) + 1)) \) and \( \pi_{NNj} = \frac{(\alpha - 1)c + \gamma(\alpha(\gamma - 1) + 1))^2}{4(1 - \alpha)^2\gamma^2} \). For all \( \frac{\gamma(1 + \alpha(\gamma - 1))}{1 - \alpha} - 2\gamma^2 < c < \gamma(1 - \gamma) \), the condition for the asymmetric overlapping capacities, we have
\( \pi_{NNi} < \tilde{\pi}_{AN}(k^*) \) and \( \pi_{NNj} < \tilde{\pi}_{AN}(1-\alpha) \). Therefore, deviating to autoscaling is strictly profitable for each firm and both firms using fixed capacities cannot be an equilibrium.

Intuitively, when both firms set capacity constraints, competition is restricted and firms set high prices. Thus, it is always beneficial to react to a fixed capacity by using autoscaling and undercutting the price of the opposition to win over Segment 3.

**Proof for the Choice of Using Autoscaling**

Given the fact that at least one firm uses autoscaling in equilibrium, we examine whether both firms use autoscaling or only one, by comparing the profits from Table 2. If Firm 2 adopts autoscaling, Firm 1’s best response is to adopt autoscaling if and only if

\[
(1 - c)(\gamma^2\alpha + \gamma(1 - \gamma)(1 - \alpha)) \geq \frac{c(\alpha(\gamma^2 - 1) + 1) + \gamma(\alpha(-\gamma) + \alpha - 1))}{4(1 - \alpha)\gamma^2(1 - c)},
\]

where the right hand side of the inequality is Firm 1’s profit if it chooses the best possible capacity \( k \), given by \( k_{na}^* = \frac{c(\alpha(\gamma^2 - 1) + 1) + \gamma(\alpha(-\gamma) + \alpha - 1))}{2\gamma^2(1 - c)} \), to Firm 2’s autoscaling decision.

Thus, only one firm uses autoscaling when \( \gamma > (1 - \alpha)/(2 - 3\alpha) \) and \( c < \hat{c} \), where \( \hat{c} \) is defined as follows:

\[
\hat{c} = \frac{2(1 - \gamma)^2\gamma^2(2\alpha\gamma - \alpha - \gamma + 1) + \gamma((2 - 3\alpha)^2\gamma^3 + ((9 - 5\alpha)\alpha - 4)\gamma^2 - (\alpha - 1)\alpha\gamma + (\alpha - 1)^2)}{2(1 - \gamma)^2(1 - \alpha)^2 + 1}.
\]

For \( c > \hat{c} \) or \( \gamma < (1 - \alpha)/(2 - 3\alpha) \), both firms will choose autoscaling.

**A.3 Proof of Proposition 2**

We start by comparing average prices with and without autoscaling in the region for overlapping capacities. Based on the proof of Proposition 1, when \( c < \frac{(1 - \alpha)(\gamma - 1)}{1 - \alpha} - 2\gamma^2 \), both firms set their capacities equal to 1 - \( \alpha \) without autoscaling. The cumulative distribution function of prices set by each firm is

\[
F(x) = \frac{1}{\pi_{1-\alpha}}
\]

Thus, the probability density function equals \( f(x) = \frac{\alpha}{x(1-2\alpha)} \) and the average price of each firm is \( \bar{p}_{NN} = \int_{\gamma}^{1} f(x) x \, dx = \frac{\alpha}{x(1-2\alpha)} \). Now suppose autoscaling is offered in this region. For \( c < \frac{(1 - \alpha)(\gamma - 1)}{1 - \alpha} - 2\gamma^2 \), we have \( \gamma < (1 - \alpha)/(2 - 3\alpha) \). As shown in the proof of the choice of using autoscaling, when autoscaling is offered and \( \gamma < (1 - \alpha)/(2 - 3\alpha) \), both firms use autoscaling. With both firms using autoscaling, the probability density function of the price of each firm equals \( g(x) = \frac{\alpha}{x(1-2\alpha)(x-c)} \) and the average price of each firm becomes \( \bar{p}_{AA} = \int_{\gamma}^{1} g(x) x \, dx = \frac{(1 - 2\alpha)\alpha(1-c) + \alpha \log(\frac{1-c}{1-2\alpha}(1-c))}{1-2\alpha} \). Thus we have \( \bar{p}_{AA} - \bar{p}_{NN} = c(1 - \frac{\alpha \log(\frac{1-c}{1-2\alpha})}{1-2\alpha}) \).
We know that \( \log(s) < s - 1 \), for any \( s > 1 \). Allowing \( s = \frac{1-\alpha}{\alpha} \), we show that \( \log(\frac{1-\alpha}{\alpha}) < \frac{1-2\alpha}{\alpha} \), thus \( \bar{p}_{AA} > \bar{p}_{NN} \).

Next consider the region for separated capacities. Without autoscaling, \( k_1 + k_2 = 1 \) and both firms set their price equal to 1. When both firms use autoscaling, the average price equals \( \bar{p}_{AA} \). As we showed, \( \log(\frac{1-\alpha}{\alpha}) < \frac{1-2\alpha}{\alpha} \). Thus, \( \bar{p}_{AA} = \frac{(1-2\alpha)c+\alpha \log(\frac{1-\alpha}{\alpha})(1-c)}{1-2\alpha} < 1 \).

When only one firm (Firm 2) uses autoscaling and the other (Firm 1) chooses capacity \( k_1 \), the probability density function of Firm 1’s price equals \( h_1(x) = \frac{(k_1-1)(c-1)}{(a-k_1)(c-x)^2} \) and its average price is \( \bar{p}_{NA} = \int_1^\infty h_1(x)dx = c + \frac{(1-k_1)(1-c)\log(\frac{c}{k_1-\alpha})}{k_1-\alpha} \). For all \( s > 0 \), we know that \( s \log(s) > s - 1 \). Allowing \( s = \frac{1-\alpha}{1-k_1} \), we have \( \frac{1-k_1}{k_1-\alpha} < (1-\alpha)\log((\frac{1-\alpha}{1-k_1})) \) and \( \bar{p}_{NA} < 1 \).

For Firm 2, the probability density function is \( h_2(x) = \frac{k_1(-\alpha+c+k_1-k_1+1)}{(a-1)x^2(a-k_1)} + \left( 1 - \frac{k_1(c-1)}{(a-1)} \right) \delta(x-1) \) and the average price equals \( \bar{p}_{AN} = \int_1^\infty h_2(x)dx = c \frac{(1-\alpha)}{c(k_1-\alpha)+1(k_1)} \). We know that \( \log(s) < s - 1 \), for any \( s > 1 \). Allowing \( s = \frac{1-\alpha}{c(k_1-\alpha)+1(k_1)} \), we find \( \log\left(\frac{1-\alpha}{-\alpha+c+k_1-k_1+1}\right) < \frac{1}{-\alpha+c+k_1-k_1+1} \) and therefore, \( \bar{p}_{AN} < 1 \). Thus, autoscaling decreases average prices in the region for separated capacities.

### A.4 Proof of Proposition 3

We compare the cut-offs on \( F \) such that both firms enter. Let \( F_{AA} \) denote the cut-off for two firms to enter with autoscaling, \( F_{NA} \) denote the cut-off for two firms to enter when only one firm uses autoscaling, and \( F_{NN} \) denote the cut-off for two firms to enter when autoscaling is not available.

Comparing the profits from Table 2 in the presence of autoscaling, it is straightforward to show \( \pi_{NA} < \pi_{AN} \). Thus, if the subgame equilibrium involves only one firm choosing autoscaling, both firms will enter if \( F < \pi_{NA} \). Both firms will choose autoscaling if \( \pi_{NA} < \pi_{AA} \). As such, both firms will enter if \( F < \text{Max}[^{\pi_{NA}, \pi_{AA}}] \).

First consider the two cases in which \( \gamma(1-\gamma) > c \). If \( c < \frac{2(1+\alpha(\gamma-1))}{1-\alpha} - 2\gamma^2 \), then \( 0 < c < 1 \) requires \( \gamma < (1-\alpha)/(2-3\alpha) \). Therefore, both firms will use autoscaling if autoscaling is available. Thus, \( F_{AA} - F_{NN} = c(1-\alpha + (\gamma - \alpha + 2\alpha\gamma)) \), which is decreasing in \( \alpha \) and positive at \( \alpha = 1/2 \) and therefore positive for all \( \alpha < 1/2 \). If \( c > \frac{2((1+\alpha(\gamma-1))}{1-\alpha} - 2\gamma^2 \), \( F_{AA} - F_{NN} \) is positive for \( c > c \), where both firms use autoscaling.

Also, \( F_{NA} > F_{NN} \) for \( \gamma(1-\gamma) > c \). Thus, in the region for overlapping capacities, the cutoff \( F \) for two firms entering is bigger with autoscaling.

Now consider when \( \gamma > c \) and \( \gamma(1-\gamma) < c \). In this case, \( F_{AA} - F_{NN} = \gamma(1-c)(1-\alpha - \gamma(1-2\alpha)) - (\gamma-c)/2 \) which is convex in \( \gamma \), equal to \(-c(1-c)(1-2\alpha)/2 < 0 \) when evaluated at \( \gamma = 1 \), decreasing in \( \gamma \) through \( \gamma = 1 \), and equal to zero at \( \gamma_{AA} = \frac{(1-2\alpha)-c(1-\alpha)+\sqrt{c(1-c)(1-2\alpha)(1-2\alpha)-2c(1-\alpha)(2-2\alpha)^2}}{4(1-c)(1-2\alpha)} \). Therefore, \( F_{AA} - F_{NN} \) is negative for any \( \gamma > \gamma_{AA} \).
Also for \( \gamma > c \) and \( \gamma (1 - \gamma) < c \), we have \( F_{NA} - F_{NN} = \frac{(c(\alpha(\gamma^2-1)+1)+\gamma(\alpha(-\gamma)+\alpha-1))^2}{4(\alpha-1)\gamma^2(c-1)} - \frac{1}{2}(\gamma - c) \), which is equal to \( ((-1 + 2\alpha)(c - 1))/(4(-1 + \alpha)) < 0 \) at \( \gamma = 1 \). Also, \( \frac{\partial(F_{NA} - F_{NN})}{\partial \gamma} < 0 \) at \( \gamma = 1 \). We find \( F_{NA} = F_{NN} \) has one root between \( \gamma = 0 \) and \( \gamma = 1 \), which is \( \hat{\gamma}_{NA} = (1-\alpha)(1-\gamma+\sqrt{1-2\alpha}) + \sqrt{(\alpha-1)\alpha^2(\gamma^2-1)^2+\alpha^2(\sqrt{1-2\alpha}-1)^2+2(\sqrt{1-2\alpha}-1)(\gamma^2-1)^2+2(\sqrt{1-2\alpha}-1)\gamma^2+2(\sqrt{1-2\alpha}-1))}\). Therefore, \( F_{NA} - F_{NN} \) is negative for any \( \gamma > \hat{\gamma}_{NA} \).

Thus, autoscaling decreases the range of \( F \) such that both firms enter for \( \gamma > \hat{\gamma} = Max[\hat{\gamma}_{AA}, \hat{\gamma}_{NA}] \).

### A.5 Proof of Proposition 4

Suppose autoscaling is not available. A firm will enter if its expected profit is greater than its entry cost. A firm’s best response to its competitor not entering the market is to enter the market, if and only if \( F < (\gamma - c)(1 - \alpha) \). Supposing one firm enters the market, the remaining firm’s best response to its competitor’s entry is to also enter the market provided the expected profit earned when competing is greater than the cost of entry. The anticipated payoffs associated with being one of two firms entering are reported in Section 4.3 and the conditions on \( F \) are presented in Lemma 1.

If only one firm enters, this firm will enjoy monopoly power over \((1 - \alpha)\) consumers with probability \( \gamma \). Using autoscaling, the firm’s profit equals to \( \gamma (1 - c)(1 - \alpha) \), which is strictly larger than \( (\gamma - c)(1 - \alpha) \), the profit earned without autoscaling, for any \( \gamma > 0 \). With autoscaling, at least one firm enters the market if and only if \( F < F_A \equiv \gamma (1 - c)(1 - \alpha) \) whereas without autoscaling at least one firm enters if and only if \( F < F_N \equiv (\gamma - c)(1 - \alpha) \).

### A.6 Proof of Corollary 1

We show that for sufficiently small \( F \) (such that both firms enter the market), when \( \gamma > c \), \( \gamma (1 - \gamma) < c \), and \( \pi_{NA} < \pi_{AA} < \frac{2-c}{2} \), we have a prisoner’s dilemma situation where both firms use autoscaling, even though their profits would be higher if autoscaling was not available.

When \( \gamma > c \) and \( \gamma (1 - \gamma) < c \), firms set separated capacities, and each earn expected profit \( \frac{2-c}{2} \), if autoscaling is not available. However, when autoscaling is available, since \( \pi_{NA} < \pi_{AA} \), both firms use autoscaling and each earn \( \pi_{AA} < \frac{2-c}{2} \) in equilibrium, which creates the prisoner’s dilemma. Now, we have to prove that all these conditions can be satisfied at the same time to show that the described region, in which prisoner’s dilemma happens, actually exists.

Let \( \gamma = \frac{1-\alpha}{2-3\alpha} \). After algebraic simplifications, we have both firms using autoscaling in equilibrium (i.e., \( \pi_{NA} < \pi_{AA} \)) if and only if \( c < \frac{4(\alpha-1)^2}{10\alpha^2-15\alpha+6} \). Furthermore, using algebraic simplifications, the expected profit of autoscaling equilibrium for each firm is lower than that when firms do not use autoscaling (i.e., \( \pi_{AA} < \frac{2-c}{2} \)) if and only if \( c > \frac{(\alpha-1)\alpha}{\alpha^2+2\alpha-2} \). It is easy to see that \( \frac{(\alpha-1)\alpha}{\alpha^2+2\alpha-2} < \frac{4(\alpha-1)^2}{10\alpha^2-15\alpha+6} \) for any \( \alpha < 1/2 \).
Therefore, when \( c \in \left( \frac{(\alpha-1)\alpha}{\alpha^2+2\alpha-2}, \frac{4(\alpha-1)^2}{100\alpha^2-150\alpha+51} \right) \), the prisoner’s dilemma situation holds.

### A.7 Proof of Proposition 5

If \( c > \hat{c} \), then both firms would choose autoscaling upon entry and consumer surplus is \( E[CS_{AA}] = \gamma^2(1-c)(1-2\alpha) \), if \( F < (1-c)((1-\alpha)(1-\gamma)\gamma + c\alpha^2) \).

If \( c < \hat{c} \), then only one firm would choose autoscaling upon entry and the other chooses \( k_{na} = \frac{c(\alpha(\gamma^2-1)\gamma+\gamma(\alpha-\gamma)+\alpha-1)}{2(\alpha-1)^2} \). The expected capacity of the firm using autoscaling is

\[
E[k_{na}] = \frac{(k_{na}c-1)c-1}{(c-1)(c_{na}-c)} + \frac{(c-1)(c_{na}-c)}{(c-1)c^2(k_{na}-c)}
\]

Thus, the expected consumer surplus is \( E[CS_{NA}] = -c((1-\alpha)(1-\gamma)\gamma + \gamma^2E[k_{na}] + k_{na}) + (\gamma^2(1-\gamma)\gamma((1-\alpha)+k_{na})-\frac{1}{2}\gamma(\alpha(\gamma-1)+1)-c(\alpha(\gamma-1)^2+2\gamma-1)) = \frac{c(\alpha(\gamma^2-1)\gamma+\gamma(\alpha-\gamma)+\alpha-1)^2}{4(\alpha-1)^2(\gamma^2+1)} \), when \( F < \frac{(c(\alpha(\gamma^2-1)\gamma+\gamma(\alpha-\gamma)+\alpha-1)^2}{4(\alpha-1)^2(\gamma^2+1)} \).

Similarly, when autoscaling is not available, the expected consumer surplus is

\[
E[CS_{NN}] = \begin{cases} \gamma^2(1-2\alpha) & \text{if } c < \frac{\gamma(1-\alpha)(\gamma-1)}{1-\alpha} - 2\gamma^2 \text{ and } F < (1-\alpha)(\gamma(1-\gamma)-c) + \gamma^2 \\
\frac{(1-\alpha)c^2}{4\gamma^2} + \frac{(\alpha-1)c(2\gamma+1)}{2\gamma} + \frac{(\alpha(\gamma-1)+1)(\gamma+\gamma+1)}{4(\alpha-1)} & \text{if } \frac{\gamma(1-\alpha)(\gamma-1)}{1-\alpha} - 2\gamma^2 < c < \gamma(1-\gamma) \text{ and } F < \frac{(\gamma(1-\alpha)(\gamma-1)-c(1-\alpha))^2}{4\gamma^2(1-\alpha)} \\
0 & \text{otherwise.}
\end{cases}
\]

We start the comparison of consumer surplus with and without autoscaling in the region for overlapping capacities. For \( c < \frac{\gamma(1-\alpha)(\gamma-1)}{1-\alpha} - 2\gamma^2 \), both firms set their capacity to \( 1-\alpha \) without autoscaling. Note that \( c < \frac{\gamma(1-\alpha)(\gamma-1)}{1-\alpha} - 2\gamma^2 \) and \( 0 < c < 1 \) require \( \gamma < (1-\alpha)/(2-3\alpha) \), therefore it is not possible that only one firm uses autoscaling in this region. Thus, we only compare cases when both firms use autoscaling with cases when autoscaling is not available. If \( F < (1-\alpha)(\gamma(1-\gamma)-c) + \gamma^2 \), then both firms enter the market with or without autoscaling and \( E[CS_{NN}] > E[CS_{AA}] \). For \( (1-\alpha)(\gamma(1-\gamma)-c) + \gamma^2 < F < (1-c)((1-\alpha)(1-\gamma)\gamma + \gamma^2) \), \( E[CS_{AA}] > E[CS_{NN}] = 0 \). Finally, for \( F > (1-c)((1-\alpha)(1-\gamma)\gamma + \gamma^2) \), we have \( E[CS_{AA}] = E[CS_{NN}] = 0 \).

Next, consider when \( \frac{\gamma(1-\alpha)(\gamma-1)}{1-\alpha} - 2\gamma^2 < c < \gamma(1-\gamma) \), where one firm sets its capacity to \( k^* < 1-\alpha \) and the other chooses \( k = 1-\alpha \), when autoscaling is not available. First, we analyze the cases when both firms use autoscaling. We find that \( E[CS_{NN}] > E[CS_{AA}] \) when \( F < \frac{\gamma(1-\alpha)(\gamma-1)-c(1-\alpha))^2}{4\gamma^2(1-\alpha)} \) and \( c < c_M = \frac{\gamma(4\alpha\gamma^3 - \sqrt{7} \sqrt{4(1-2\alpha)^2\gamma^2 - 8(\alpha-1)(2\alpha-1)^2\gamma - 4(\alpha-1)(2\alpha-1)\gamma^2 + (\alpha(13\alpha-20)+8)\gamma+4(\alpha-1)^2-2\alpha\gamma - 2\gamma^2 + 2\gamma+1}}{1-\alpha} \). Note that these conditions only hold when both firms use autoscaling, since \( c_M \) is positive if and only if \( \gamma <
\[(1 - \alpha)/(2 - 3\alpha)\). Also when both firms use autoscaling, \(E[CS_{AA}] > E[CS_{NN}] = 0\) if \(\gamma(1-\alpha(1-\gamma)) - (c - (1 - \alpha))^2 < F(1 - c)((1 - \alpha)(1 - \gamma) + 1)\), and \(E[CS_{AA}] = E[CS_{NN}] = 0\) if \(F(1 - c)((1 - \alpha)(1 - \gamma) + 1)\).

Next, we compare the consumer surplus without autoscaling for \(\gamma(1 + \alpha(\gamma - 1)) - 2\gamma^2 < c < \gamma(1 - \gamma)\), with consumer surplus when only one firm uses autoscaling, which occurs for \(\gamma > (1 - \alpha)/(2 - 3\alpha)\) and \(c < \hat{c}\).

For all \(s > 0\), we know that \(s \log(s) > s - 1\). Assuming \(s = \frac{-\alpha c + (\alpha - 2)\gamma c(e - 1) + (\alpha - 1)\gamma}{c(e - 1)(\alpha(\gamma + \alpha - 1) + \gamma((\alpha - 2)\gamma - \alpha + 1))}\), we find a lower bound for \(E[CS_{NA}]\), denoted \(E[CS_{NA}]_L\). Thus, we have

\[
E[CS_{NA}] > E[CS_{NA}]_L = \frac{\gamma^4(c - 1)(2(\alpha - 1)\alpha c^2 + ((8 - 5\alpha)\alpha - 4)c + \alpha^2) + (\alpha - 1)^2 c^2}{4(\alpha - 1)\gamma^2(c - 1)}
\]

\[
= \frac{(\alpha - 1)^2\gamma^2(2e^3 - 4e^2 - 1) + 2(\alpha - 1)^2 c\gamma + 2(\alpha - 1)^2 c\gamma^3(c - 2)}{4(\alpha - 1)\gamma^2(c - 1)}
\]

Comparing \(E[CS_{NA}]_L\) and \(E[CS_{NN}]\), we find that for \(c = 0\), we have \(E[CS_{NA}]_L = E[CS_{NN}]\). Also, for \(c > 0\), we have \(\frac{\partial E[CS_{NA}]_L}{\partial c} > \frac{\partial E[CS_{NN}]}{\partial c}\), when \(\gamma > (1 - \alpha)/(2 - 3\alpha)\) and \(c < \hat{c}\). Therefore, we have \(E[CS_{NA}] > E[CS_{NA}]_L > E[CS_{NN}]\), for any \(c > 0\).

Finally, if \(c > \gamma(1 - \gamma)\), then when autoscaling is not available, both firms either do not enter the market or enter and set their prices equal to 1. Both of these cases result in zero consumer surplus. Therefore autoscaling increases consumer surplus when both firms use autoscaling and \(F < (1 - c)((1 - \alpha)(1 - \gamma) + 1)\), or when one firm uses autoscaling and \(F < (1 + (\alpha(\gamma - 1) + (\alpha - 1)\gamma))^{\frac{1}{2(\alpha - 1)\gamma + 1}}\). Otherwise, autoscaling does not affect consumer surplus.

### A.8 Proofs of Section 5.1

**Equilibrium Strategies without Autoscaling**

In outcomes in which both firms have the same reservation value, the solution to the pricing subgame is similar to that of Section 4.1, replacing \(vi = 1\) with \(vi\) equal to \(V_h\) or \(V_i\). We focus on the analysis of the case when one firm has high value \(V_h\) while the other firm has low value \(V_l\). We denote the capacity chosen by the high-value firm \(k_h\) and the capacity of the low-value firm \(k_l\). Assuming the firms’ capacities overlap such that \(k_l + k_h > 1\), the pricing subgame has no pure strategy equilibrium and both firms choose mixed strategy pricing.

Suppose Firm \(i\)'s range of prices in the mixed strategy equilibrium is \([p_i^{Min}, v_i]\), where \(p_i^{Min}\) is the lowest price in the support of the price distribution chosen by Firm \(i\). Similarly, suppose Firm \(j\)'s range of prices is \([p_j^{Min}, v_j]\). We prove that in equilibrium \(v_i - p_i^{Min} = v_j - p_j^{Min}\). Assume to the contrary that \(v_i - p_i^{Min} > v_j - p_j^{Min}\). In a mixed strategy equilibrium, Firm \(i\) gains the same expected profit from any price
$p_i \in [p_i^{min}, v_i]$. Firm $i$’s expected profit from setting $p_i = p_i^{min}$ is $p_i^{min}(1 - \alpha)$, since $v_i - p_i^{min} > v_j - p_j$ for any $p_j \in [p_j^{min}, v_j]$ and all consumers in Segment 3 prefer Firm $i$ to Firm $j$. Now, suppose Firm $i$ chooses $p_i = p_i^{min} + \varepsilon$, where $\varepsilon$ is an infinitely small positive number. Still, all consumers in Segment 3 prefer Firm $i$ to Firm $j$, since $v_i - (p_i^{min} + \varepsilon) > v_j - p_j$ for any $p_j \in [p_j^{min}, v_j]$. Thus, the expected profit of Firm $i$ becomes $(p_i^{min} + \varepsilon)(1 - \alpha)$. Since $\varepsilon > 0$, Firm $i$’s expected profit from $p_i = p_i^{min} + \varepsilon$ is strictly larger than the profit from $p_i = p_i^{min}$. This contradicts the assumption that both prices $p_i = p_i^{min}$ and $p_i = p_i^{min} + \varepsilon$ are included in Firm $i$’s mixed strategy pricing equilibrium. Thus, both firms must have the same length of price interval, where the length of price interval for Firm $i$ is defined as $v_i - p_i^{Min}$.

Let $z_l$ be the price at which the low-value firm is indifferent between attacking to sell to $k_l$ consumers at price $z_l$ and retreating to sell to $1 - k_l$ consumers at price $V_l$. We have $z_l = \frac{V_l(1-k_l)}{k_l}$. Similarly, $z_h = \frac{V_h(1-k_l)}{k_h}$ is the price at which the high-value firm is indifferent between attacking and retreating. The firms’ price intervals in the mixed strategy equilibrium are such that $v_i - p_i^{Min} = v_j - p_j^{Min}$ and the high-value and low-value firms’ prices are greater than $z_h$ and $z_l$ respectively. Thus, the length of the price intervals of both firms must be equal to $Min[V_l - z_l, V_h - z_h]$. For $k_l > \frac{V_l}{V_h} k_h$, we have $V_l - z_l < V_h - z_h$ and therefore, the length of the price intervals of both firms is $V_l - z_l$. Thus, for $k_l > \frac{V_l}{V_h} k_h$, the low-value firm chooses the price range $p_l \in (z_l, V_l)$ and the high-value firm chooses the price range $p_h \in (V_h - (V_l - z_l), V_h)$. For $k_l < \frac{V_l}{V_h} k_h$, we have $V_l - z_l > V_h - z_h$ and therefore, the length of the price intervals of both firms is $V_h - z_h$. Thus, for $k_l < \frac{V_l}{V_h} k_h$, the high-value firm chooses the price range $p_h \in (z_h, V_h)$ and the low-value firm chooses the price range $p_l \in (V_l - (V_h - z_h), V_l)$.

When $V_l - p_i > V_h - p_h$, the low-value firm sells all of its capacity, $k_l$, leaving the remaining $1 - k_l$ consumers for the high-value firm. When $V_l - p_i < V_h - p_h$, the high-value firm sells all of its capacity, $k_h$, and the low-value firm sells to the remaining $1 - k_l$ consumers. Suppose that $F_l(\cdot)$ is the cumulative distribution function of the price of the low-value firm and $F_h(\cdot)$ is the cumulative distribution function of the price of the high-value firm. Thus, excluding the sunk cost of capacity, the profit of the low-value firm setting price $x$ is $\pi_l(x) = x((1 - k_h)F_h((V_h - V_l) + x) + k_l(1 - F_h((V_h - V_l) + x))).$ Similarly, the profit of the high-value firm is $\pi_h(x) = x(k_h(1 - F_l(x - (V_h - V_l))) + (1 - k_l)F_l(x - (V_h - V_l)))$.

We first solve the pricing subgame without autoscaling for $k_l < \frac{V_l}{V_h} k_h$, where price ranges are $p_l \in (V_l - (V_h - z_h), V_l)$ and $p_h \in (z_h, V_h)$. We set the derivative of the profit functions $\pi_l(x)$ and $\pi_h(x)$ to zero and use the boundary conditions $F_l(V_l - (V_h - z_h)) = 0$ and $F_h(z_h) = 0$. We find $F_l(x) = \frac{V_h(k_h + k_l - 1) + k_h(x - V_l)}{(k_h + k_l - 1)(V_h - V_l)}$ and $F_h(x) = \frac{k_l(-k_h x - k_h V_h + V_l)}{k_h(k_h + k_l - 1)(V_h - V_l - x)}$. Note that $F_h(x)$ jumps from $\frac{k_l V_h}{k_h V_l}$ to 1 at $x = V_h$, which means the high-value firm uses price $V_h$ with probability $1 - \frac{k_h V_h}{k_h V_l}$. Thus, when $k_l < \frac{V_l}{V_h} k_h$, the profits excluding sunk costs are

$$\pi_l = k_l \left( V_l - \frac{V_h(k_h + k_l - 1)}{k_h} \right) \quad \text{and} \quad \pi_h = V_h - k_l V_h.$$
Next, we analyze the pricing subgame without autoscaling when \( k_l > \frac{V_l}{k_h} k_h \), where price ranges are \( p_l \in (z_l, V_l) \) and \( p_h \in (V_l - (V_l - z_l), V_h) \). We set the derivative of the profit functions to zero, using the boundary conditions \( F_l(z_l) = 0 \) and \( F_h(V_h - (V_l - z_l)) = 0 \). We find \( F_l(x) = \frac{k_h((k_h - 1)V_l + k_l x)}{k_l(k_h + k_l - 1)(V_h - V_l + x)} \) and \( F_h(x) = \frac{(k_h - 1)V_l + k_l x}{k_h(k_h + k_l - 1)} \). Note that \( F_l(V_l) = \frac{k_h V_l}{k_h V_h} \), implying the low-value firm sets its price to \( V_l \) with probability \( 1 - \frac{k_h V_l}{k_h V_h} \). Thus, when \( k_l > \frac{V_l}{k_h} k_h \), the profits excluding sunk costs are

\[
\pi_l = V_l - k_h V_l \quad \text{and} \quad \pi_h = k_h \left( V_h - \frac{V_l (k_h + k_l - 1)}{k_l} \right).
\]

Assuming each firm realizes high value with a probability \( \gamma \), expected profits are derived as follows.

\[
E(\pi) = \begin{cases} 
  k_i(-c + \gamma V_h - \gamma V_l + V_l) & \text{if } k_i + k_j \leq 1 \\
  \frac{k_j((-1)\gamma V_l - \gamma V_h - c k_j)}{k_j} & \text{if } k_i < \frac{V_l}{k_h} k_j \text{ and } k_i + k_j \geq 1 \\
  k_j(-c k_j + \gamma^2 V_l + \gamma k_h V_l + \gamma^2 V_h - \gamma^2 V_l - \gamma^2 V_h + \gamma k_j V_l) + &\frac{k_j(-c k_j + \gamma^2 V_l + \gamma k_h V_l + \gamma^2 V_h - \gamma^2 V_l - \gamma^2 V_h + \gamma k_j V_l)}{k_j} \\
  -c k_i + (1 - \gamma) \gamma k_i \left( V_h - \frac{V_l (k_i + k_j - 1)}{k_j} \right) + &\frac{-c k_i + (1 - \gamma) \gamma k_i \left( V_h - \frac{V_l (k_i + k_j - 1)}{k_j} \right) +}{k_j} \\
  (1 - k_j) \left( \gamma^2 V_l + (\gamma - 1)^2 V_l \right) + (\gamma - 1) \gamma (k_j - 1) V_l &\frac{(1 - k_j) \left( \gamma^2 V_l + (\gamma - 1)^2 V_l \right) + (\gamma - 1) \gamma (k_j - 1) V_l}{k_j} \\
  (k_j - 1)((\gamma - 1) V_l - \gamma V_h - c k_i) &\frac{(k_j - 1)((\gamma - 1) V_l - \gamma V_h - c k_i)}{k_j} \\
\end{cases}
\]

We find the optimal capacities that maximize the firms’ expected profits. For \( k_i + k_j < 1 \), we have \( \frac{\partial E(\pi)}{\partial k_i} = -c + \gamma V_h - \gamma V_l + V_l \), which is negative for \( \gamma < \frac{\gamma V_h - \gamma V_l}{\gamma V_l} \). Thus, for \( c > \gamma (V_h - V_l) + V_l \), equilibrium capacities are \( k_i = k_j = 0 \).

For \( k_j < \frac{V_l}{k_h} k_i \) and \( k_i + k_j \geq 1 \), we have \( \frac{\partial E(\pi)}{\partial k_i} = -c \). Therefore, there can be no equilibrium where capacities are so far apart such that \( k_i < \frac{V_l}{k_h} k_j \), since the high capacity firm would have incentive to decrease its capacity until \( k_i < \frac{V_l}{k_h} k_j \) no longer holds.

For \( \frac{V_l}{k_h} k_j < k_i < k_j \) and \( k_i + k_j \geq 1 \), the optimal capacities are

\[
k_i = \hat{k}^* = \frac{(\gamma - 1) \gamma V_l (-c + \gamma (\gamma + 1) V_h + (\gamma - 2) (\gamma - 1) V_l)}{c^2 + 2(\gamma - 1) \gamma c (V_h - V_l) + (\gamma - 1) \gamma \left( ((\gamma - 1) \gamma V_h^2 + 2 \gamma (\gamma + 1) V_h V_l + (\gamma - 4) (\gamma - 1) V_l^2) \right)}
\]

\[
k_j = \hat{k}^{**} = \frac{(\gamma^2 V_h - \gamma V_l + V_l) (c + (\gamma - 1) \gamma (V_h + V_l))}{c^2 + 2(\gamma - 1) \gamma c (V_h - V_l) + (\gamma - 1) \gamma \left( ((\gamma - 1) \gamma V_h^2 + 2 \gamma (\gamma + 1) V_h V_l + (\gamma - 4) (\gamma - 1) V_l^2) \right)}
\]

For this equilibrium to satisfy \( k_i + k_j > 1 \), we must have \( c < (1 - \gamma) \gamma (V_h - V_l) \). Otherwise, for \( (1 - \gamma) \gamma (V_h - V_l) < c < \gamma (V_h - V_l) + V_l \), equilibrium capacities are such that \( k_i + k_j = 1 \).

Comparing these capacities with \( 1 - \alpha \), we find the higher capacity reaches its ceiling (i.e., \( \hat{k}^{**} > 1 - \alpha \)) for
\[ c < c^* = \sqrt{\gamma^4 V_h^2 - 2(\gamma - 1)\gamma^2 V_h V_l (4(2\alpha^2 - 3\alpha + 1)\gamma + 1) + (\gamma - 1)^2 V_l^2 (8(2\alpha^2 - 3\alpha + 1)\gamma + 1) - 2\alpha\gamma V_h + 2\alpha\gamma V_h + \gamma^2 V_h - 2\gamma V_h + (\gamma - 1)V_l (2(\alpha - 1)\gamma + 1)}. \]

For \( c < c^* = -\frac{3\alpha\gamma^2 V_h + \alpha\gamma V_h + 2\gamma V_h - \gamma V_l + \alpha\gamma V_l - 2\alpha V_l - \gamma^2 V_l + V_l}{\alpha - 1} \), we have \( k^* > 1 - \alpha \) and thus both firms set their capacity equal to \( 1 - \alpha \).

Thus, the equilibrium capacities and profits are derived as shown below.

\[
\begin{aligned}
    k_i &= k_j = 0 \quad \text{and} \quad \pi_i = \pi_j = 0 \quad \text{if } c > \gamma(V_h - V_l) + V_l \\
    k_i &= k_j = \frac{1}{2} \quad \text{and} \quad \pi_i = \pi_j = \frac{1}{2}(-c + \gamma V_h - \gamma V_l + V_l) \quad \text{if } (1 - \gamma)\gamma(V_h - V_l) < c < \gamma(V_h - V_l) + V_l \\
    k_i &= k^*, k_j = k^{**} \quad \text{and} \quad \pi_i = \pi_j = \frac{\gamma V_l (c - \gamma V_h + (\gamma - 1)V_l)}{c(\gamma - 1)V_l - \gamma^2 V_h} \quad \text{if } c^* < c < (1 - \gamma)\gamma(V_h - V_l) \\
    k_i &= k_j = 1 - \alpha \quad \text{and} \quad \pi_i = \pi_j = \pi^*_{\text{sym}} \quad \text{if } c^{**} < c < c^* \\
    k_i &= k_j = 1 - \alpha \quad \text{and} \quad \pi_i = \pi_j = \pi^*_{\text{asym}} \quad \text{if } 0 < c < c^{**}
\end{aligned}
\]

Where we have:

\[
\pi^*_{\text{asym}} = \frac{(\alpha - 1)^2 c^2 - 2(\alpha - 1)c((\gamma - 1)(-\alpha\gamma V_h + (\alpha - 1)\gamma V_l + V_l) - \gamma V_h) + \gamma^2 V_h^2(\alpha(\gamma - 1) + 1)^2}{2(\gamma - 1)\gamma V_h V_l ((\alpha - 1)\alpha\gamma^2 + (\alpha - 3)\alpha\gamma + \alpha + \gamma - 1) + (\gamma - 1)^2 V_l^2((\alpha - 1)\gamma(\alpha(\gamma - 4) - \gamma + 2) + 1)} + \frac{4(\alpha - 1)((\gamma - 1)V_l - \gamma^2 V_h)}{4(\alpha - 1)((\gamma - 1)V_l - \gamma^2 V_h)}
\]

\[
\pi^{**}_{\text{asym}} = \frac{(\alpha - 1)^2 c^2 - 2(\alpha - 1)c((\gamma - 1)(-\alpha\gamma V_h + (\alpha - 1)\gamma V_l + V_l) - \gamma V_h) + \gamma^2 V_h^2(\alpha(\gamma - 1) + 1)^2}{2((\alpha - 1)c + \gamma V_h - (\gamma - 1)(-\alpha\gamma V_h + (\alpha - 1)\gamma V_l + V_l))} + \frac{2(\gamma - 1)\gamma V_h V_l ((\alpha - 1)\alpha\gamma^2 + (\alpha - 3)\alpha\gamma + \alpha + \gamma - 1) + (\gamma - 1)^2 V_l^2((\alpha - 1)\gamma(\alpha(\gamma - 4) - \gamma + 2) + 1)}{2((\alpha - 1)c + \gamma V_h - (\gamma - 1)(-\alpha\gamma V_h + (\alpha - 1)\gamma V_l + V_l))}
\]

\[
\pi^*_{\text{sym}} = (\alpha - 1)c + (2\alpha - 1)\gamma^2 (V_h - V_l) + (\alpha - 1)\gamma(V_l - V_h) + \alpha V_l
\]

Thus, both firms enter the market without autoscaling, if and only if \( F < \text{Min} [\pi_i, \pi_j] \).

**Equilibrium Strategies when Both Firms Use Autoscaling**

We solve the case where both firms use autoscaling, assuming one firm has high value and the other has low value. For \( V_l < c \), the low-value firm does not benefit from selling to any consumers with autoscaling, and thus the high-value firm sells to \( 1 - \alpha \) consumers at the price of \( V_h \). Next, suppose \( V_l > c \) such that the low-value firm has incentive to sell with autoscaling. There is no pure strategy equilibrium for the pricing subgame and firms use mixed strategy prices. The price at which the firm is indifferent between attacking and retreating is \( z_l = \frac{-2ac + c + aV_l}{1 - a} \) for the low-value firm, and \( z_h = \frac{-2ac + c + aV_h}{1 - a} \) for the high-value firm. It is easy to show \( V_h - z_h > V_l - z_l \). Thus, using a logic similar to our analysis of the case without autoscaling,
we show the length of the price interval chosen by both firms in the mixed strategy equilibrium should be $V_l - z_l$. This means the low-value firm chooses the price range $p_l \in (z_l, V_l)$ and the high-value firm chooses the price range $p_h \in (V_h - (V_l - z_l), V_h)$.

Given the price distributions, the profits are $\pi_l(x) = (x-c)((1-\alpha)(1-F_h((V_h-V_l)+x)) + \alpha F_h((V_h-V_l)+x))$ for the low-value firm and $\pi_h(x) = (x-c)((1-\alpha)(1-F_l(x-(V_h-V_l))) + \alpha F_l(x-(V_h-V_l)))$ for the high-value firm. Using boundary condition $F_h(V_h-(V_l-z_l)) = 0$, we find $F_h(x) = \frac{(2\alpha-1)c+\alpha(V_l+c(x)+V_h+x)}{(2\alpha-1)(c+V_h-V_l-x)}$, resulting in $\pi_l = \alpha(V_l-c)$. Similarly, with boundary condition $F_l[z_l] = 0$, we have $F_l(x) = \frac{(2\alpha-1)c-\alpha(V_l+x) + x}{(2\alpha-1)(c-V_h+V_l-x)}$, where $F_l(x)$ jumps from $\frac{V_l-c}{c-V_h}$ to 1 at $x = V_l$. This distribution results in $\pi_h = -\alpha(c+V_h-2V_l) + V_h - V_l$.

Considering the different possible values for each firm, the expected profit of Firm $i$ when both firms use autoscaling is $E(\pi_{AA}) = -\alpha c + (2\alpha-1)\gamma^2(V_h-V_l) + (\alpha-1)\gamma(V_l-V_h) + \alpha V_l$.

**Equilibrium Strategies when Only One Firm Uses Autoscaling**

Consider the case where only one firm uses autoscaling and the two firms have different values. The pricing strategy of each firm depends on 1) whether it is using autoscaling, and 2) whether its value is $V_l$ or $V_h$. Let $z_{an}(.)$ be the price at which the firm using autoscaling is indifferent between *attacking* to sell to $1-\alpha$ consumers at price $z_{an}(.)$ and *retreating* to sell to $1-k_{na}$ consumers at price $v_{an}$, where $v_{an}$ is the value of the firm using autoscaling and $k_{na}$ is the capacity chosen by the firm not using autoscaling. We have $z_{an}(v_{an}) = \frac{c(k_{na}-\alpha)+(1-k_{na})v_{an}}{1-\alpha}$. Similarly, $z_{na}(v_{na}) = \frac{\alpha v_{na}}{k_{na}}$ is the price at which the firm not using autoscaling is indifferent between *attacking* and *retreating*, where $v_{na}$ is the value of the firm not using autoscaling.

First, suppose the firm using autoscaling has low value $V_l$ and the firm with fixed capacity has high value $V_h$. Assume $V_l > c$ so that the low-value firm has incentive to sell with autoscaling. We have $V_l - z_{an}(V_l) < V_h - z_{na}(V_h)$. Thus, the low-value firm chooses the price range $p_l \in (z_{an}(V_l), V_l)$ and the high-value firm chooses the price range $p_h \in (V_h - (V_l - z_{an}(V_l)), V_h)$.

Next, suppose the firm using autoscaling has high value $V_h$ and the firm with fixed capacity has low-value $V_l$. For $V_h - V_l > c$ and $k_{na} > \frac{(\alpha-1)V_l}{c-V_h}$, we have $V_h - z_{an}(V_h) > V_l - z_{na}(V_l)$; otherwise, $V_h - z_{an}(V_h) < V_l - z_{na}(V_l)$. Thus, for $V_h - V_l > c$ and $k_{na} > \frac{(\alpha-1)V_l}{c-V_h}$, the low-value firm chooses the price range $p_l \in (z_{na}(V_l), V_l)$ and the high-value firm chooses the price range $p_h \in (V_h - (V_l - z_{na}(V_l)), V_h)$. Otherwise, the low-value firm chooses the price range $p_l \in (V_l - (V_h - z_{an}(V_h)), V_l)$ and the high-value firm chooses the price range $p_h \in (z_{an}(V_h), V_h)$.

Using the same steps as before, we derive the pricing distribution for each firm. First, suppose the firm that uses autoscaling has $V_l$, such that $V_l > c$, and the firm with fixed capacity has $V_h$. We find $F_h(x) = \frac{(k_{na}-1)(V_l-x)}{(\alpha-k_{na})(c+V_h-V_l-x)} + 1$ and $F_l(x) = \frac{k_{na}(c(k_{na}-\alpha)-(k_{na}-1)V_l)+V_h+(\alpha-1)x}{(\alpha-1)(k_{na}-\alpha)(V_h-V_l+x)}$, where $F_l(x)$ jumps from $\frac{k_{na}(c-V_l)}{(\alpha-1)V_h}$.
to 1 at the price of \( x = V_l \). Corresponding profits are \( \pi_l = (k_{na} - 1)(c - V_l) \) for the firm using autoscaling and \( \pi_h = \frac{k_{na}(c-k_{na}+(c-V_h-V_l)+k_{na}V_l-V_h)}{\alpha-1} \) for the firm with fixed capacity, excluding sunk costs.

Next, consider the case where the firm using autoscaling realizes \( V_h \) and the firm with fixed capacity realizes \( V_l \). If \( V_h - V_l < c \) or \( k_{na} < \frac{(\alpha-1)V_l}{c-V_h} \), the price ranges are \( p_l \in (V_l - (V_h - z_{an}(V_h)), V_l) \) and \( p_h \in (z_{an}(V_h), V_h) \). The price distributions are \( F_l(x) = \frac{(k_{na} - 1)(V_l - x)}{(k_{na} - \alpha)(c - V_h + V_l - x)} + 1 \) and \( F_h(x) = \frac{k_{na}(c-k_{na}+(c-V_h-V_l)+k_{na}V_h-V_l)}{(\alpha-1)(k_{na} - \alpha)(V_h - V_l - x)} \), where \( F_h(x) \) jumps from \( k_{na} \) to 1 at the price of \( x = V_h \). The profits are \( \pi_l = \frac{k_{na}(c-k_{na}+(c-V_h-V_l)+k_{na}V_h-V_l)}{\alpha-1} \) for the firm that sets capacity and \( \pi_h = (k_{na} - 1)(c - V_l) \) for the firm that uses autoscaling.

Finally, we assume the firm using autoscaling has \( V_h \), the other firm has \( V_l \), and we have \( V_h - V_l > c \) and \( k_{na} > \frac{(\alpha-1)V_l}{c-V_h} \). Price ranges are \( p_l \in (z_{an}(V_l), V_l) \) and \( p_h \in (V_h - (V_l - z_{na}(V_l)), V_h) \). The firms use the price distributions \( F_h(x) = \frac{k_{na}(-V_h + V_l - x) - V_l}{(k_{na} - \alpha)(c(V_h + V_l - x))} \) and \( F_l(x) = \frac{(\alpha-1)(k_{na} - \alpha)V_l}{(k_{na} - \alpha)(c(V_l - V_h + V_l - x))} \), where \( F_l(x) \) jumps from \( k_{na} \) to 1 at the price of \( x = V_l \). The profits are \( \pi_l = \alpha V_l \) for the firm that sets capacity and \( \pi_h = \frac{(k_{na} - 1)(c-V_h - V_l)}{k_{na}} \) for the firm that uses autoscaling.

We denote \( E(\pi_{AN}) \) as the expected profit of the firm using autoscaling and \( E(\pi_{NA}) \) as the expected profit of the firm setting fixed capacity.

For \( 0 < V_l < c \), the firm using autoscaling does not sell to any consumers if it realizes low value and we have

\[
E(\pi_{AN}) = \gamma \left( \gamma(k_{na} - 1)(c - V_h) + \frac{(\alpha-1)(\gamma - 1)(\alpha V_l - k_{na}(c - V_h + V_l))}{k_{na}} \right)
\]

\[
E(\pi_{NA}) = \frac{\gamma^2 k_{na} (c - k_{na}) + (k_{na} - 1)V_h}{\alpha - 1} - ck_{na} - \gamma(\gamma - 1)k_{na}V_h + (\gamma - 1)^2 k_{na}V_l - \alpha \gamma(\gamma - 1)V_l
\]

For \( c < V_l < V_h - c \) and \( k_{na} > \frac{(\alpha-1)V_l}{c-V_h} \), we have

\[
E(\pi_{AN}) = \gamma^2 (k_{na} - 1)(c - V_h) + \frac{((\alpha-1)(\gamma - 1)(\alpha V_l - k_{na}(c - V_h + V_l)))}{k_{na}} +
\]

\[
(\gamma - 1)^2 (k_{na} - 1)(c - V_l) - \gamma(\gamma - 1)(k_{na} - 1)(c - V_l)
\]

\[
E(\pi_{NA}) = \frac{\gamma(-\alpha (k_{na}(c - V_h + V_l) + V_l) + k_{na} (ck_{na} - (k_{na} - 2)V_l - V_h) + \alpha^2 V_l)}{\alpha - 1} +
\]

\[
\frac{\gamma^2 ((-k_{na} - \alpha)(ck_{na} - k_{na}V_h - \alpha V_l + V_l) - (k_{na} - 1)k_{na}(c - V_l))}{\alpha - 1}
\]

Finally, for \( V_l > V_h - c \) or \( k_{na} < \frac{(\alpha-1)V_l}{c-V_h} \), expected profits are

\[
E(\pi_{AN}) = (k_{na} - 1)(c - V_h + (\gamma - 1)V_l)
\]

\[
E(\pi_{NA}) = \frac{-(k_{na} - 1)k_{na}(c - V_h + (\gamma - 1)V_l)}{\alpha - 1}
\]
The firm not using autoscaling sets the optimal capacity, \( k_{na}^* \), such that \( \frac{\partial E(\pi_{NA})}{\partial k_{na}} = 0 \):

\[
k_{na}^* = \begin{cases} 
\frac{c(\alpha(\gamma+\gamma-2)+\gamma+\gamma-2\alpha k_1)}{2\gamma^2(c-V_h)} & \text{if } 0 < V_i < c \\
\frac{c(\alpha(\gamma+\gamma-2)+\gamma+\gamma-2\alpha k_1)}{2\gamma^2(c-V_h)} & \text{if } c < V_i < V_h - c \\
\left\lfloor \frac{c}{2} \right\rfloor & \text{if } V_i > V_h - c \text{ and } V_i > c
\end{cases}
\]

Inserting \( k_{na}^* \) into \( E(\pi_{NA}) \) and \( E(\pi_{AN}) \), we derive the optimal profits when only one firm uses autoscaling.

Note that we have \( E(\pi_{NA}) = \frac{k_{na}}{1-\alpha} E(\pi_{AN}) < E(\pi_{AN}) \), for \( V_i > c \). Therefore, the condition for both firms entering the market, when only one firm uses autoscaling, is \( F < \min[E(\pi_{NA}), E(\pi_{AN})] = E(\pi_{NA}) \). Thus, when autoscaling is offered, both firms enter the market if and only if \( F < \max[E(\pi_{AA}), E(\pi_{NA})] \).

### A.9 Proofs of Section 5.2

#### Equilibrium Strategies without Autoscaling

Consider overlapping capacities without autoscaling. If the price of Firm \( i \) is higher than its competitor’s, Firm \( i \) sells to Segment \( i \) and what is left is of Segment 3 after Firm \( j \) sells all of its capacity. Thus, Firm \( i \) sells to \((\alpha + (1 - 2\alpha) - \frac{(1-2\alpha)k_1}{1-\alpha})\) consumers if \( p_i > p_j \) and to \( k_i \) consumers if \( p_i < p_j \). Suppose \( F_j(\cdot) \) is the cumulative distribution function of the price set by Firm \( j \) when the overlapping condition on capacities holds. The profit of Firm \( i \) is derived as \( \pi_i(x) = x \left( k_i(1 - F_j(x)) + F_j(x) \left( \alpha + (1 - 2\alpha) - \frac{(1-2\alpha)k_j}{1-\alpha} \right) \right) \).

Let \( z_i \) be the price at which Firm \( i \) is indifferent between attacking to sell to \( k_i \) consumers at price \( z_i \) and retreating to sell to \((1 - \alpha) - \frac{(1-2\alpha)k_j}{1-\alpha}\) consumers at price 1. We have \( z_i = \frac{-\alpha - (1-2\alpha)k_j}{k_i} \). As shown in the proof of Section 5.1, the length of the price intervals chosen by both firms must be the same in equilibrium to satisfy the necessary condition of mixed strategy equilibrium. Also Firm \( i \) does not choose any price lower than \( z_i \). Thus, both firms choose prices in the range of \( p \in (\max[z_1, z_2], 1) \). Without loss of generality, suppose \( k_1 < k_2 \). For \( k_1 + k_2 > \frac{(a-1)^2}{2a-1} \), the price range is \((z_2, 1)\); otherwise, the price range is \((z_1, 1)\).

We first solve the pricing subgame for \( k_1 > \frac{(a-1)^2}{2a-1} - k_2 \), for which both firms’ price range in the mixed strategy equilibrium is \((z_2, 1)\). Solving the differential equation \( \frac{\partial \pi_2}{\partial x} = 0 \) and using the boundary condition \( F_1(z_2) = 0 \), we have \( F_1(x) = \frac{((a-1)^2 +(2a-1)k_1 + (a-1)k_2 x)}{x(2a-1)k_1 + (a-1)(2a-1)k_2} \). Also, solving \( \frac{\partial \pi_2}{\partial x} = 0 \) and using the boundary condition \( F_2(z_2) = 0 \) results in \( F_2(x) = \frac{k_1((a-1)^2 +(2a-1)k_1 + (a-1)k_2 x)}{k_2((a-1)^2 +(2a-1)k_1 + (2a-1)k_2)} \). Note that \( F_2(1) \) equals \( \frac{k_1((2a-1)k_1 + (a-1)(2a-1)k_2)}{k_2((a-1)^2 +(2a-1)k_1 + (2a-1)k_2)} \), which is less than 1 for \( k_1 + k_2 > \frac{(a-1)^2}{2a-1} \). This means Firm 2 is setting its price to 1 with a probability of \( 1 - \frac{k_1((2a-1)k_1 + (a-1)(2a-1)k_2)}{k_2((a-1)^2 +(a-1)(2a-1)k_1 + (2a-1)k_2)} \). Firms’ profits, excluding costs of capacity,
when \( k_1 > -\frac{(\alpha-1)^2}{2\alpha-1} - k_2 \), are

\[
\pi_1 = \frac{k_1 (-\alpha - 1)^2 - 2\alpha k_1 + k_1)}{(\alpha-1)k_2} \quad \text{and} \quad \pi_2 = \frac{(-\alpha - 1)^2 - 2\alpha k_1 + k_1)}{\alpha - 1}.
\]

Next, we find the solution for \(-\alpha \frac{(1-2\alpha)k_2}{1-\alpha} + 1 < k_1 < -\frac{(\alpha-1)^2}{2\alpha-1} - k_2 \), where both firms’ price range in the mixed strategy equilibrium is \((z_1,1)\). This time, we use the boundary conditions \( F_1(z_1) = 0 \) and \( F_2(z_1) = 0 \) to solve \( \frac{\partial \pi_2}{\partial k_2} = 0 \) and \( \frac{\partial \pi_1}{\partial k_1} = 0 \). We find

\[
F_1(x) = k_2((\alpha - 1)k_2 x + ((\alpha - 1)^2 + (2\alpha - 1)k_2))
\]

and

\[
F_2(x) = \frac{\partial \pi_2}{\partial k_2} = 0
\]

where \( F_1(x) \) jumps from \( k_2((\alpha - 1)^2 + (2\alpha - 1)k_2) \) to 1 at \( x = 1 \). Excluding sunk costs, firms’ profits when \(-\alpha \frac{(1-2\alpha)k_2}{1-\alpha} + 1 < k_1 < -\frac{(\alpha-1)^2}{2\alpha-1} - k_2 \) are

\[
\pi_1 = \frac{((\alpha - 1)^2 + 2\alpha k_2 - k_2)}{1 - \alpha} \quad \text{and} \quad \pi_2 = \frac{k_2((\alpha - 1)^2 + 2\alpha k_2 - k_2)}{(1 - \alpha)k_1}
\]

Finally, when \( k_1 < -\alpha \frac{(1-2\alpha)k_2}{1-\alpha} + 1 \), each firm has a local monopoly and sets its price at \( p_1 = 1 \), earning a profit of \( k_1 \). Note that for \( -\frac{(\alpha-1)^2}{2\alpha-1} - k_1 < -\alpha \frac{(1-2\alpha)k_2}{1-\alpha} + 1 \), there is no capacity overlap for \( p_2 < p_1 \), but there is capacity overlap for \( p_1 < p_2 \). In both these cases, regardless of what \( p_2 \) is, Firm 1 can sell to \( k_1 \) consumers. Thus Firm 1 chooses \( p_1 = 1 \). In order to sell all of its capacity, Firm 2 needs to choose \( p_2 < p_1 \), and does so by setting \( p_2 = 1 - \epsilon \), where \( \epsilon \) is an infinitely small positive number. Therefore, the equilibrium outcome for \( -\frac{(\alpha-1)^2}{2\alpha-1} - k_1 < -\alpha \frac{(1-2\alpha)k_2}{1-\alpha} + 1 \) is \( \pi_1 = k_1 \) and \( \pi_2 = k_2 \), the same as when \( k_1 < -\frac{(\alpha-1)^2}{2\alpha-1} - k_2 \).

Assuming Firm \( i \) realizes \( v_i = 1 \) with probability \( \gamma \), expected profits for \( \alpha < k_i , k_j < 1 - \alpha \) are

\[
E(\pi_i) = \begin{cases} 
 k_i(\gamma - c) & \text{if } k_i < -\alpha \frac{(1-2\alpha)k_j}{1-\alpha} + 1 \\
\frac{\gamma^2 (-\alpha (\alpha + 2k_j - 2) + k_j - 1)}{(\alpha - 1)k_j} - ck_i - (\gamma - 1)\gamma k_i & \text{if } k_j > -\alpha \frac{(1-2\alpha)k_i}{1-\alpha} + 1 \text{ and } k_i < k_j < -\frac{(\alpha-1)^2}{2\alpha-1} - k_i \\
k_i \left( \frac{\gamma^2 (-\alpha (\alpha + 2k_i - 2) + k_i - 1)}{(\alpha - 1)k_i} - c - (\gamma - 1)\gamma \right) & \text{if } k_i > -\alpha \frac{(1-2\alpha)k_j}{1-\alpha} + 1 \text{ and } k_j < k_i < -\frac{(\alpha-1)^2}{2\alpha-1} - k_j \\
k_i \left( \frac{\gamma^2 (-\alpha (\alpha + 2k_i - 2) + k_i - 1)}{(\alpha - 1)k_i} - c - (\gamma - 1)\gamma \right) & \text{if } k_i > k_j > -\frac{(\alpha-1)^2}{2\alpha-1} - k_j \\
\frac{\gamma^2 (-\alpha (\alpha + 2k_i - 2) + k_i - 1)}{(\alpha - 1)k_i} - ck_i - (\gamma - 1)\gamma k_i & \text{if } k_i > k_j > -\frac{(\alpha-1)^2}{2\alpha-1} - k_i
\end{cases}
\]

Given the expected profits for each set of capacities, we solve for equilibrium capacity choices. For \( k_i < -\alpha \frac{(1-2\alpha)k_j}{1-\alpha} + 1 \), we have \( \frac{\partial E(\pi_i)}{\partial k_i} = \gamma - c \). Thus for \( \gamma < c \), equilibrium capacities are \( k_i = k_j = 0 \).

For \( k_i > -\alpha \frac{(1-2\alpha)k_j}{1-\alpha} + 1 \) and \( k_i < k_j < -\frac{(\alpha-1)^2}{2\alpha-1} - k_i \), we have \( \frac{\partial E(\pi_i)}{\partial k_i} = \gamma(1 - \gamma) - c \). Thus, for \( \gamma(1 - \gamma) < c \), the firm with the lower capacity in this region, decreases its capacity until

\[
k_i + \frac{(1-2\alpha)k_j}{1-\alpha} = 1 - \alpha.
\]
The symmetric equilibrium requires \( k_i = k_j = \frac{(\alpha - 1)^2}{2 - 3\alpha} (\gamma - c) \), resulting in each firm earning a monopoly profit of \( \pi^*_{nn} = \frac{(\alpha - 1)^2}{2 - 3\alpha}(\gamma - c) \).

For \( k_i > k_j > \frac{(\alpha - 1)^2}{2 - 3\alpha} - k_i \), we have \( \frac{\partial E(\pi_i)}{\partial k_i} = \gamma(1 - \gamma) - c \). Thus, when \( \gamma(1 - \gamma) > c \), for \( k_i + k_j > \frac{(\alpha - 1)^2}{4\alpha - 1} \), the firm with the higher capacity increases its capacity to \( 1 - \alpha \). When \( k_i = 1 - \alpha \), Firm \( j \) maximizes its profit by choosing \( k_j = k^* = \frac{(\alpha - 1)^2}{4\alpha - 1}(c - \gamma) \). Thus, profits become \( \pi_i = \frac{\alpha}{4}(\alpha - 1)(c - \gamma) \) and \( \pi_j = -\frac{(\alpha - 1)^2}{4\alpha - 1}(c - \gamma)^2 \).

For \( c < \frac{2(\alpha - 1)^2}{\alpha - 1}(\gamma + 2\gamma - 1) \), we have \( k^* > 1 - \alpha \). Therefore, if \( c < \frac{2(\alpha - 1)^2}{\alpha - 1}(\gamma + 2\gamma - 1) \), both firms set their capacity at \( 1 - \alpha \), and each earn a profit of \( (\alpha - 1)c + \gamma(\alpha(2\gamma - 1) - \gamma + 1) \).

**Equilibrium Strategies when Only One Firm Uses Autoscaling**

Suppose one firm uses autoscaling, setting the price \( p_{an} \), while the other firm chooses a fixed capacity of \( k_{na} \) and sets the price \( p_{na} \). When \( p_{an} < p_{na} \), the firm using autoscaling sells to \((1 - \alpha)\) consumers and the firm with fixed capacity sells to \(\alpha\) consumers. When \( p_{an} > p_{na} \), the firm with fixed capacity sells to \( k_{na} \) consumers, of whom \( \frac{1 - 2\alpha}{\alpha - 1} k_{na} \) come from Segment 3. Thus, the firm using autoscaling can only sell to \((1 - 2\alpha) - \frac{1 - 2\alpha}{\alpha - 1} k_{na} + \alpha \) consumers. Suppose \( F_{an}(\cdot) \) is the cumulative distribution function of price for the firm using autoscaling and \( F_{na}(\cdot) \) is the cumulative distribution function of price for the firm with fixed capacity.

The profits of the two firms, excluding sunk costs, are \( \pi_{na}(x) = x(k_{na}(1 - F_{an}(x)) + \alpha F_{an}(x)) \) for the firm with fixed capacity and \( \pi_{an}(x) = (x - c)(F_{na}(x)(\alpha + (1 - 2\alpha) - \frac{1 - 2\alpha}{\alpha - 1} k_{na} + (1 - \alpha)(1 - F_{an}(x))) \) for the firm using autoscaling.

The firm setting capacity is indifferent between attacking and retreating at the price of \( z_{na} = \frac{\alpha}{k_{na}} \). The firm using autoscaling is indifferent between attacking and retreating at the price of \( z_{an} = \frac{c k_{na} - 2\alpha k_{na}}{(\alpha - 1)^2} + \frac{(\alpha - 1)^2 (2\alpha - 1) k_{na}}{(\alpha - 1)^2} \). As shown in the proof of Section 5.1, the condition for mixed strategy equilibrium is that the length of the price interval for both firms should be equal. If \( k_{na} > \frac{(\alpha - 1)^2 - \sqrt{(\alpha - 1)^2(4\alpha(1 - 2\alpha) + (1 - 3\alpha)^2)}}{2(2\alpha - 1)(c - 1)} \), we have \( z_{an} > z_{na} \) and both firms choose prices in the range \((z_{an}, 1)\); otherwise, prices are set in the range \([z_{na}, 1)\).

First, suppose \( k_{na} < \frac{(\alpha - 1)^2 - \sqrt{(\alpha - 1)^2(4\alpha(1 - 2\alpha) + (1 - 3\alpha)^2)}}{2(2\alpha - 1)(c - 1)} \). Solving \( \frac{\partial \pi_{na}}{\partial x} = 0 \) and using the boundary condition \( F_{an}(z_{na}) = 0 \) returns \( F_{an}(x) = \frac{k_{na} x - \alpha}{k_{na} x - \alpha} \) and \( \pi_{na} = \alpha \). Thus, the expected profit of the firm choosing capacity is \( E(\pi_{na}) = -ck_{na} + \gamma(\alpha\gamma - \gamma k_{na} + k_{na}) \). We have \( \frac{\partial E(\pi_{na})}{\partial k_{na}} > 0 \) for \( c < \gamma(1 - \gamma) \). Thus, the firm choosing capacity increases its capacity until \( k_{na} = \frac{(\alpha - 1)^2 - \sqrt{(\alpha - 1)^2(4\alpha(1 - 2\alpha) + (1 - 3\alpha)^2)}}{2(2\alpha - 1)(c - 1)} \) and there is no equilibrium for \( k_{na} < \frac{(\alpha - 1)^2 - \sqrt{(\alpha - 1)^2(4\alpha(1 - 2\alpha) + (1 - 3\alpha)^2)}}{2(2\alpha - 1)(c - 1)} \).

For \( k_{na} \geq \frac{(\alpha - 1)^2 - \sqrt{(\alpha - 1)^2(4\alpha(1 - 2\alpha) + (1 - 3\alpha)^2)}}{2(2\alpha - 1)(c - 1)} \), we solve differential equations \( \frac{\partial \pi_{na}}{\partial x} = 0 \) and \( \frac{\partial \pi_{na}}{\partial x} = 0 \) using the boundary conditions \( F_{an}(z_{an}) = 0 \) and \( F_{na}(z_{an}) = 0 \). We find \( F_{na}(x) = \frac{(2\alpha - 1) k_{na} x + (\alpha - 1)^2 - 2\alpha k_{na} + k_{na} + (\alpha - 1)^2 x}{(2\alpha - 1) k_{na} x - c - x} \) and \( F_{an}(x) = \frac{k_{na} (x - c - x) + (\alpha - 1)^2 x (\frac{1}{\alpha - 1} - \frac{c}{\alpha - 1})}{(\alpha - 1)^2 x (k_{na} - \alpha)} \), where \( F_{an}(x) \) jumps from \( \frac{(2\alpha - 1) k_{na}^2}{(\alpha - 1) k_{na} x - c} \) to \( 1 \) at \( x = 1 \). The profit of the firm using autoscaling is \( \frac{(c - 1)(\alpha - 1)^2 + (2\alpha - 1) k_{na}}{\alpha - 1} \) and the profit of the
other firm is $k_{na}(c(k_{na}-2\alpha k_{na})+(\alpha-1)^2(2\alpha-1)k_{na})$. The expected profit of the firm using autoscaling is
$$E(\pi_{an}) = \frac{\gamma(c-1)((\alpha-1)^2+(2\alpha-1)\gamma k_{na})}{(\alpha-1)^2}$$
and the expected profit of the firm setting capacity is
$$E(\pi_{na}) = k_{na}\left(-\frac{(2\alpha-1)\gamma k_{na}(c-1)}{(\alpha-1)^2} - c + \gamma\right),$$
which is maximized at $k_{na} = k^*_na = -\frac{(\alpha-1)^2(c-\gamma)}{2(2\alpha-1)\gamma(c-1)}$. Inserting $k_{na} = k^*_na$ into $E(\pi_{an})$ and $E(\pi_{na})$, we find
$$\pi^*_an = \frac{1}{2}(1-\alpha)(-2\gamma c + c + \gamma)$$
and
$$\pi^*_na = \frac{(\alpha-1)^2(c-\gamma)^2}{4(1-2\alpha)\gamma^2(1-c)}.$$

**Effect of Autoscaling on Entry**

Next, we find the region where autoscaling decreases entry. We have $\pi^*_nn > \pi^*_na$ for
$$\gamma > \gamma_{NA} = \frac{(2-3\alpha) - \sqrt{3\alpha - 2}\sqrt{16(2\alpha-1)c^2 + 16(1-2\alpha)c+(3\alpha-2)}}{8(2\alpha-1)(c-1)}.$$

Similarly, denoting $\pi^*_aa$ as the profit of each firm when both firms use autoscaling, which is the same as what we had with the efficient rationing rule, we have $\pi^*_nn > \pi^*_aa$ for
$$\gamma > \gamma_{AA} = \frac{(3\alpha - 2)\sqrt{(\alpha-1)^2((15\alpha^2+16\alpha-4)c^2+2(6\alpha^2-7\alpha+2)c+(1-2\alpha)^2)} + (3\alpha^2 - 5\alpha + 2) c + (-2\alpha^2 + 3\alpha - 1)}{2(2\alpha-1)(3\alpha-2)(c-1)}.$$

Thus, autoscaling decreases entry for $\gamma > \gamma = Max[\gamma_{AA}, \gamma_{NA}]$ and $Max[\pi^*_aa; \pi^*_na] < F < \pi^*_nn$.

**A.10 Proofs of Section 5.3**

When autoscaling is available, the condition for at least one firm entering the market is $F < (1-c)\gamma(1-\alpha)$. This means at least one firm enters as long as $c \leq \frac{F}{(\alpha-1)\gamma} + 1$. The cloud provider’s profit, given the single entrant’s probability of success, is $\gamma c(1-\alpha)$. Thus, the provider maximizes $c$ and sets $c = \frac{F}{(\alpha-1)\gamma} + 1$. The provider’s expected profit is $\pi_a^{CP} = \gamma(\frac{F}{(\alpha-1)\gamma} + 1)(1-\alpha)$.

Both firms enter the market using autoscaling for $F < (1-c)(\alpha\gamma^2 + (1-\alpha)(1-\gamma)\gamma)$. In this region, the total capacity purchased is 1, if both entrants realize high value. If only one firm realizes high value, the total capacity purchased is 1 $- \alpha$. Thus, the provider’s profit is $c(2(1-\alpha)(1-\gamma)\gamma + \gamma^2)$ and $c$ is set to its maximum, $1 - \frac{F}{\gamma(\alpha(2\gamma-1)-\gamma+1)}$. Thus, the provider’s expected profit is $\pi_{aa}^{CP} = (1-\frac{F}{\gamma(\alpha(2\gamma-1)-\gamma+1)})(2(1-\alpha)(1-\gamma)\gamma + \gamma^2)$.

We know one firm uses autoscaling and the other sets $k_{na} = \frac{c(\gamma^2+1)+\gamma(\alpha(-\gamma)+\alpha-1)}{2\gamma(c-1)}$, when $c < \hat{c}$.

The cumulative distribution functions of price for the firms are $F_{na}(x) = \frac{k_{na}(c(k_{na}-\alpha)-k_{na}+1+(\alpha-1)x)}{(\alpha-1)x(k_{na}-\alpha)} + 1$ and $F_{an}(x) = \frac{k_{an}(c(k_{an}-\alpha)-k_{an}+1+(\alpha-1)x)}{(\alpha-1)x(k_{an}-\alpha)} + 1$. The purchased capacity of the firm using autoscaling is $(1-k_{na})F_{na}(x) + (1-\alpha)(1-F_{na}(x)) = \frac{k_{na}(c(\alpha-k_{na}-1))}{x-\alpha}$. Thus, the capacity of this firm depends on the probability density function of its price, which is $f_{an}(x) = \frac{k_{na}(c(\alpha-k_{na}+1)(k_{na}-1))}{(\alpha-1)x^2(k_{na}-\alpha)}$. We derive the expected capacity of the firm using autoscal-
The expected profit of the provider when both firms enter but only one firm uses autoscaling is
\[ \pi_{na}^{CP} = c \left( \frac{(1 - \alpha)(1 - \gamma) + \gamma^2 E[k_{an}] + k_{na}}{2} \right) \]
Maximizing \( \pi_{na}^{CP} \) with respect to \( c \) does not have a closed form solution. However, we can numerically compare the providers’ maximum profit when one firm or both firms use autoscaling and determine what \( c \) would be set by the provider. The results are shown in Figure 14.

Without autoscaling, at least one firm enters the market for \( F < (1 - \alpha - 1)(\gamma - c) \). The provider can increase the price to \( c = \frac{F}{\alpha - 1} + \gamma \), so only one firm enters and purchases \( 1 - \alpha \) capacity. The provider makes a profit of \( (1 - \alpha)\gamma - F \). Note that this profit is equal to \( \pi_{na}^{CP} \), the provider’s profit when autoscaling is offered and only one firm enters the market.

When \( \gamma (1 - \gamma) < c < \gamma \), both firms enter the market, each firm purchasing a capacity half the size of the market, if \( F < \frac{\gamma^2 - c}{2} \). The provider sets \( c = \gamma - 2F \) and makes a profit of \( \gamma - 2F \).

Next, consider the overlapping capacities equilibrium, where Firm i has \( 1 - \alpha \) capacity and Firm j has a capacity of \( k_j^* = \frac{\alpha(\gamma^2 - 1)}{2} \). Firm j’s profit equals \( \frac{(\alpha(\gamma^2 - 1) + 1)^2}{4(1 - \alpha)^2} \gamma \) and the highest c for which Firm j enters the market in this equilibrium is \( c = \frac{\gamma(\alpha(\gamma^2 - 1) + 1)}{2(1 - \alpha)^2} \). The profit of the provider is \( c(k_j^* + (1 - \alpha)) \). This profit is maximized at \( c = \frac{\gamma(\alpha(\gamma^2 - 1) + 1)}{2(1 - \alpha)^2} \), resulting in a profit of \( \frac{(\alpha(\gamma^2 - 1) + 1)^2}{8(1 - \alpha)} \) for the provider.

Finally, for \( c < \frac{2(\alpha^3 - 1) - 2\gamma + 1}{\alpha - 1} \), both firms purchase \( (1 - \alpha) \) capacity and earn a profit of \( (\alpha - 1)c + (2\alpha - 1)\gamma^2 - (\alpha - 1)\gamma \). The provider’s profit is \( 2c(1 - \alpha) \), where the maximum c for which both firms enter is \( c = \frac{F + \gamma(2\alpha(\gamma - \alpha - 1) - F)}{\alpha - 1} \). Thus, the provider’s expected profit is \( 2(\gamma(2\alpha(\gamma - \alpha - 1) - F) - F) \)

We compare the provider’s maximum profit for different \( \gamma \), \( F \), and \( \alpha \) and present the results in Figure 15.
Separate Endogenous Prices for Autoscaling Capacity and Fixed Capacity

We assume the provider charges $c_k$ for a unit of fixed capacity and $c_a$ for a unit of capacity purchased through autoscaling. This distinction between the two capacity prices only affects the provider’s optimal profit when one firm uses autoscaling and the other enters the market without autoscaling. The capacity of the firm not using autoscaling is calculated in the same way as before and equals $k_{na} = -\alpha c_k + c_k + \gamma \left( \alpha \gamma c_a + (\alpha(\alpha) + \alpha - 1) \right)$. The expected capacity of the firm using autoscaling is derived as $E[k_{an}] = \frac{(c_a - 1)k_{na}(c_a - \alpha) + (\alpha c_a) + (k_{na} - 1)}{(\alpha c_a) - (k_{na} - \alpha)}$. 

Thus, the expected profit of the provider when only one firm uses autoscaling becomes $\pi_{na}^{CP} = c_0 \left( (1 - \alpha)(1 - \gamma) + \alpha^2 E[k_{an}] \right) + c_k k_{na}$. We solve numerically for the maximum $\pi_{na}^{CP}$ with respect to $c_k$ and $c_a$, accounting for a possible deviation from the firm using autoscaling to purchasing fixed capacity.

The results are shown in Figure 16. Region 4 is where the provider sets prices such that both firms enter but only one firm uses autoscaling, resulting in autoscaling increasing the provider’s profit. Regions 1-3 represent similar equilibria as in Figure 13.