Investment in Flexible Resources with Demand Correlation: an application to cloud computing[*]

Wing Man Wynne Lam[†]

January, 2017

Abstract

This paper considers firms' incentives to invest in local and flexible resources when demand is uncertain and correlated. I find that market power of the monopolist providing flexible resources leads to underinvestment in the flexible resource and overinvestment in the local resource. Moreover, investment responds to demand correlation differently under monopoly and social optimum. This implies that to analyze investment efficiency in industries with correlated demand (e.g., cloud computing and sharing economy), we need data on both costs and correlations.

Keywords: Demand correlation; Capacity; Cloud computing

JEL Classification: D4, L8

[*]A previous version of this paper was circulated under the title "Competition in the Market for Flexible Resources: an application to cloud computing". I thank Paul Belleflamme, Giacomo Calzolari, Jacques Crémer, Vincenzo Denicolò, Florian Englmaier, Federico Etro, Neil Gandal, Axel Gautier, Michael Katz, Thomas-Olivier Léautier, Fabio Manenti, Domenico Menicucci, Paul Seabright, Tommaso Valletti, and the participants at the CRESSE 2013 Conference, the Toulouse 2013 Conference on Intellectual Property, Software and Internet, the Munich 2013 ICT Conference, and the European University Institute 2013 Seminar on Communications and Media Markets, as well as those at the seminar at the University of Liege for their valuable comments. I also acknowledge the support of Toulouse School of Economics and University of Bologna at earlier stages of this research. All opinions expressed are strictly my own.

[†]University of Liege (ULg), HEC Management School, Liege Competition and Innovation Institute (LCII). E-mail: wingmanwynne.lam@ulg.ac.be

1 Introduction

For firms in various industries, capacity investment decision involves both local and flexible capacities. In particular, firms often need to invest in their own capacity before demand for their products is realized, and such investments are difficult to reverse. After the demand is realized, firms have the option to further invest in a flexible resource to accommodate excess demand. Cloud computing—the leasing of computer services, computing power and storage on an unprecedented scale—is one such flexible resource. While local computing capacity can support a firm's average demand, cloud computing is able to scale services on demand and accommodate the workload that exceeds what the local capacity can handle.[1] Accordingly, firms can use cloud computing as a flexible resource for business continuity and disaster recovery plans. In addition to IT, firms have to decide whether to rent or buy office space; and people, whether to use public or private means of transport. In recent years, the sharing economy—like Uber, Lyft, and Airbnb—has further increased people's choices of flexible resources in transport and property. Another example is electricity, where firms have to decide whether to buy electricity from public utilities, which deliver instant services, or to invest in their own private generators.

Another important characteristic of these markets is that demand is uncertain and correlated. For instance, in the cloud computing market, computing demand varies daily and is correlated globally and seasonally. Geographically, computing demand from countries lying close to each other is positively correlated, whereas demand from countries that are located in different time zones is negatively correlated. Seasonally, retailers increase computing demand during the holiday season.[2] As argued by Harms and Yamartino (2010), even the largest cloud provider will not be able to fully resolve issues related to uncertainty and correlation by aggregating demand. Similarly, time and weather are common drivers of correlated demand in electricity and transport: household consumption of electricity tends to peak in the evening, while workplace consumption tends to peak in the daytime, and people most want a taxi during rush hours, holidays, and in

[1] The U.S. National Institute of Standards and Technology provides five defining characteristics of cloud computing: on-demand service, broad network access, resource pooling, rapid elasticity, and measured service. This paper focuses on the definition of on-demand service and rapid elasticity.

[2] See "Black Friday: Websites swamped by record demand," *BBC News*, 28 November 2014, available at `http://www.bbc.com/news/business-30242124`.

bad weather conditions.

Despite demand correlations being very common, the effect of them on investments is unknown. And providers of flexible resources often need to make their capacity investments under such demand uncertainties. For example, on 25 August 2013, Amazon struggled to keep up with the growing computing demand, and an IT problem at one of its cloud datacenters had caused many users of major web services such as Instagram, Vine, and Netflix to experience lengthy delays and reduced data transfer speeds for several hours.[3] This raises the interesting questions of whether investments in local and flexible resources are efficient and how they are shaped by demand correlations.

I consider two firms, whose demand is uncertain and correlated, that make their investment decision in local resources. A monopoly provider of flexible resources (e.g., Amazon, Google or Microsoft), observing firms' local investment, decides how much to invest in capacity and sets the price for its flexible resources (e.g., Amazon Web Services (AWS), Google Cloud Platform or Microsoft Azure).[4] After demand is realized, firms can buy flexible resources if demand exceeds their local capacity. One should keep in mind the "firm" here refers to the seller of final services to the consumers, and the "provider" to the supplier of flexible resources. Thus, investment occurs on both sides: the firm choosing between local and flexible resources, and the provider investing in its flexible capacity.

I show that the provider of flexible resources tends to underinvest in its capacity with respect to the socially optimal level, whereas firms tend to overinvest in their local capacity. Such inefficiencies are due to the monopolist's market power. Firms invest in local capacity to avoid being exploited by the monopolist, which in turn reduces the monopolist's investment incentives (see Proposition 2).

In addition, I show that both socially optimal and monopoly investments in flexible resources increase with correlation if the investment cost of flexible resources is small

[3]See "Instagram, Vine and Netflix hit by Amazon glitch," *BBC News*, 26 August 2013, available at `http://www.bbc.co.uk/news/technology-23839901`. Note that capacity can be interpreted in two ways: number of physical servers or service quality. In the former case, there is a maximum traffic that each server can handle. In the latter case, even if the capacity does not hit the limit, high demands can put a costly strain on servers, which results in poor quality of service.

[4]The motivating example is cloud computing, but as mentioned at the beginning, this model can be applied to many other industries with flexible resources.

enough, and decrease with correlation if the flexible resource is costly. The reason is that as correlation increases, the social planner and the provider either "win big" when the demand realization of the two firms is high or "lose big" when the demand realization of the two firms is low. If the flexible resource is cheap, the social planner and the provider need not worry about losing. Rather, they will focus on reaping the benefits from the winning outcome, and therefore invest more as correlation increases. On the contrary, if the flexible resource is expensive, losing is costly, and thus they invest less as correlation increases (see Propositions 1 and 3). I also find that the monopolist's investment in flexible resources is more likely to be decreasing in demand correlation due to excessive local investments by firms (see Proposition 4). These results suggest that the cost condition and the degree of demand correlation have important consequences for investments in markets, e.g., for cloud computing and the sharing economy (see the conclusion for a more detailed discussion).

1.1 Paper Contribution and the Literature

This paper's main contribution to the literature is that it identifies the type of investment inefficiencies that might arise in the presence of multiple resources and demand uncertainty. Although the study of investment in local versus flexible resources and investment under demand uncertainty are by no means unheard of, their combination has not been explored yet.

More specifically, this paper relates to several areas of the literature. The first is capacity-pricing games, which dates back to Kreps and Scheinkman (1983), and was extended by Reynolds and Wilson (2000) and de Frutos and Fabra (2011) to incorporate demand uncertainty. Building on their results, I study firms' choices between two types of resources and introduce both demand uncertainty and demand correlation. This also distinguishes my work from Goyal and Netessine (2007) and Anupindi and Jiang (2008), where firms cannot invest in both local and flexible resources.

The second area concerns operational management. For example, Lee (2009) considers the optimal investment of a computing service provider when there is a single resource, and Niyato et al. (2009) study the optimal choice between private and public computing services in a monopoly and an oligopoly market, yet in a setting without demand correlation. By incorporating demand correlation, an interesting finding of this paper is that investment can increase with correlation, which is in contrast to the common belief that only negative correlations are valuable because providers can aggregate demand and reduce the risk.[5] The reason is that when capacity is cheap, the "win big" effect prevails; thus, the provider may invest more as correlation increases. Other papers in operational management, e.g., Van Miegham (1998), and Bish and Wang (2004), have failed to explain why investment inefficiencies may arise in the presence of demand uncertainty and demand correlation.

The third area concerns real options (RO) in finance, which focuses on the role of RO in providing flexibility to management decisions. However, unlike financial assets, capacity investments in IT are not tradable, and they are therefore not valued based on levels of risk. Instead, they are priced by a third party, which is the resource provider in this model. Thus, the study of the investment incentives of the resource provider is important here, but not in RO literature.[6]

Finally, this paper relates to the literature on cloud computing. Cloud computing has emerged as a new business model for managing computing and storage resources for firms and a new source of entertainment and communication services for consumers. As the cloud market is still in its infancy, many classic economic issues such as pricing, investment strategies, and the appropriate market structure are still unclear. Recently, there has been a flurry of research on the opportunities and obstacles for the adoption of cloud computing; see, for example, Armbrust et al. (2009), Harms and Yamartino (2010), and Marston et al. (2011). The existing literature, however, has mostly tackled the problem from an engineering rather than an economic perspective, and much less so from a theoretical one.[7] This paper is a first attempt at understanding the impact of demand correlation on investments in this industry.

[5]See, for instance, p. 218 of Bayrak et al. (2011).

[6]See, for instance, Angelou and Economides (2005), Benaroch and Kauffman (1999) and Kauffman et al. (2002) for details on the limitations of RO's applicability in IT investments.

[7]Recent theoretical work includes Wang (2014) and this paper. However, Wang focuses on user adoption of cloud computing without any investment on the provider's side. See also Fershtman and Gandal (2012) for a survey of the literature on the economic issues of cloud computing.

2 The Model

Demands. Two firms, 1 and 2, offer services to consumers, who have inelastic demand and a willingness to pay of r for either firm's service. I assume that the two firms are not competing against each other, so each firm sets the monopoly price r to extract all consumer surplus.[8] The demand (e.g., the total number of consumers) of each firm is uncertain, and the demands of the two firms are correlated. More specifically, demands for firms 1 and 2, denoted by x and y respectively, are drawn from a joint density $h(x, y)$ with support $[0, \infty) \times [0, \infty)$. The density of the demand for firm 1's service, x, is given by a marginal distribution $f(x) = \int_0^\infty h(x, y)dy$. Similarly, the density of the demand for firm 2's service, y, is given by $g(y) = \int_0^\infty h(x, y)dx$. In the following analysis, I focus on the case where both x and y follow an exponential distribution with a scale parameter of 1.[9] More particularly, the distribution functions of x and y, $F(x)$ and $G(y)$, and the corresponding density functions, $f(x)$ and $g(y)$, are respectively

$$\begin{aligned} F(x) =& 1 - e^{-x}, \\ G(y) =& 1 - e^{-y}, \\ f(x) =& e^{-x}, \\ g(y) =& e^{-y}. \end{aligned}$$

To characterize the correlation between the two firms' demands, I follow Gumbel (1960), where the joint distribution function $H(x, y)$ and the joint density function $h(x, y)$ are given by

$$\begin{aligned} H(x, y) =& (1 - e^{-x})(1 - e^{-y})(1 + \alpha e^{-x-y}), \\ h(x, y) =& e^{-x-y}[1 + \alpha(2e^{-x} - 1)(2e^{-y} - 1)], \end{aligned}$$

where $-1 < \alpha < 1$ is a measure of correlation.[10]

[8] Since the two firms are not competing, the qualitative results of the analysis with inelastic demand go through even if demand is elastic.

[9] The main results carry through with a uniform distribution, the proof of which is available upon request.

[10] Strictly speaking, $\rho = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}$ is the coefficient of correlation, but since α and ρ move in the same direction (more precisely, $\rho = \frac{\alpha}{4}$, see Equation (3.10) on p. 706 of Gumbel 1960), there is no loss of generality in saying that α is a measure of correlation.

Investment technologies. To serve the demands, each firm needs to build a capacity, and serving a customer takes up one unit of capacity of the firm. To do this, firms can either invest in their own local resource L or they can buy flexible resource K from the market. The difference lies in that investments in local resources are irreversible and these resources are for the exclusive use of the investing firm, while flexible resources can be bought from the market instantly when needed and released when not needed. An example of flexible resources is cloud computing, which is an on-demand service. The unit cost of a local resource and a flexible resource are denoted by c_L and c_K respectively. I assume that the local resource is supplied competitively, so that firms can buy L at a price c_L, while the flexible resource is supplied by a monopoly. Hereafter, I use the "firm" to refer to the seller of final services to the consumers, and the "provider" to refer to the supplier of flexible resources.

I consider the following game:

- Stage 1: firms 1 and 2 invest in their own local capacity L_1 and L_2 simultaneously;
- Stage 2: the provider invests in its flexible resource capacity K;
- Stage 3: the provider sets a per unit price for flexible resources p (a simple linear tariff);[11]
- Stage 4: demands x and y are realized and firms decide whether and if so, how much to buy the flexible resource.[12]

I make the following assumptions. First, $r > c_L$, so that there is incentive to purchase local resources. Second, I focus on the case where $c_K < c_L$. This could be reasonable

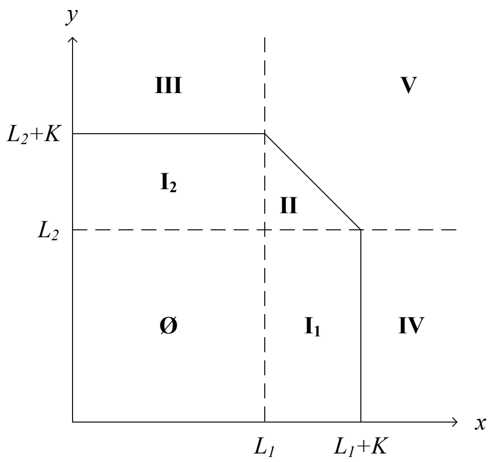
[11] In practice, a non-linear tariff is also possible. For example, Amazon and Dropbox provide basic storage services for free and additional storage for a fee. It would be interesting to analyze the optimal pricing scheme under demand uncertainty but such a topic would require substantial extension, which is beyond the scope of this paper.

[12] The timing of this model, where firms invest first, fits the scenario where some flexible resources such as cloud computing offers more flexibility in managing demand uncertainty than local resources. Alternatively, one could consider the case where firms observe the provider's investment in flexible resources before deciding their own local investment. Even in this setting, the main results go through provided price is chosen after the capacity decision, as the monopoly price will emerge as long as demand is inelastic. Another alternative is to consider the case where p is chosen prior to L, but the underinvestment problem will still occur because the provider will never price at marginal cost as its profit will become zero and it will not have any incentive to invest.

to the extent that flexible resources e.g., cloud computing usually exhibit significant economies of scale compared to local resources e.g., on-premise servers. Here I focus on the specification with $r = 1$, $c_L = 0.5$ and $c_K \in [0, 0.5]$.[13] Third, when firms are indifferent to buying and not buying the flexible resource, they will always buy it. The solution concept adopted here is subgame perfect equilibrium.

3 Social Optimum

The social planner chooses L_1, L_2 and K so as to maximize social welfare, which is defined as expected demand served minus investment costs (recall that $r = 1$). Figure 1 illustrates the basic structure of the demand for flexible resources.



Area	Demand for Flexible Resources
Ø	0
I_1	x-L_1
I_2	y-L_2
II	x+y-(L_1+L_2)
III	K
IV	K
V	K

Figure 1: Demand for Flexible Resources.

In *Area* $\emptyset$, both demands x and y are small and hence can be served by local capacities. *Area* I_1 captures the situation where y can be served by firm 2's local capacity L_2, but since x exceeds what firm 1's local capacity L_1 can serve, firm 1 needs to buy $x - L_1$ flexible resources. *Area* I_2 shows the reverse situation where firm 2 buys $y - L_2$ flexible resources. In *Area II*, both x and y exceed what the two firms' local capacities can serve, but by buying additional flexible resources, all the demands can still be served. Notice that in all the cases above, all demands are served. *Area III* represents the situation where x can be served by local capacity alone, but y is so large that not even the combination

[13] These assumptions are innocuous for two reasons. First, setting $r = 1$ is only a normalization. Second, the main results hold more generally as long as the flexible resource is more efficient, i.e., $c_K < c_L$.

of local and flexible resources can serve them all. *Area IV* shows the reverse situation where $y < L_2$, but x is very large. Finally, *Area V* captures the situation where both x and y are so large that they are not fully served even when all the capacity of flexible resources is exhausted. Thus the social welfare is given by

$$\begin{aligned}\max_{L_1,L_2,K} S = &\int_{\emptyset+I_1+I_2+II}(x+y)h(x,y)dydx + \int_{III}(x+L_2+K)h(x,y)dydx\\ &+ \int_{IV}(L_1+y+K)h(x,y)dydx + (L_1+L_2+K)\int_V h(x,y)dydx\\ &- c_K K - c_L(L_1+L_2). \end{aligned} \tag{1}$$

Let $\Omega(L_1, L_2, K)$ denote the probability of x and y falling in areas $\{III\}+\{IV\}+\{V\}$. We can show that the social planner only invests in flexible resources, the level of which is given by

$$\Omega(0,0,K) = 1 - \int_0^K \int_0^{K-x} h(x,y)dydx = c_K,$$

which equates the social marginal benefit with the marginal cost. Let L_1^*, L_2^* and K^* denote the solutions to the maximization problem (1). Then we have

Proposition 1. *The social planner only invests in the flexible resource (i.e., $L_1^* = L_2^* = 0$ and $K^* > 0$). Moreover, there exists a cutoff level $\bar{c}$ such that the socially optimal investment in flexible resources K^* increases with demand correlation α if $c_K \leq \bar{c}$, but decreases with demand correlation if $c_K > \bar{c}$.*

Proof. See Appendix A. □

The social planner only invests in flexible resources because flexible resources can be used for both firms and they are less expensive than local resources (i.e., $c_K < c_L$). Moreover, as the degree of demand correlation increases, the probability of getting either high or low demand realizations increases, which means that the social planner either "wins big" or "loses big". The impact of an increase in demand correlation therefore depends on the cost of the flexible resource. If the investment cost is sufficiently low, then losing is cheap and the social planner will focus on reaping the benefits of high demand realizations. Therefore, investment increases with correlation. On the contrary, if the investment cost is large enough, the social planner will aim at minimizing the risk of losing, so investment decreases with correlation.

4 Monopoly

Suppose now that there is a monopoly provider for the flexible resource that chooses p and K to maximize its expected profit. Since each consumer has inelastic demand, each firm will set the price equal to r. This means that a firm will buy flexible resources to serve its customers whenever the price for the flexible resource is less than or equal to r—in other words, firms' demand for flexible resources is also inelastic. Given L_1, L_2, K and the monopoly price p^m, the demand for flexible resources is the same as that in Figure 1 as long as $p^m \leq r$, which means that we must have

$$p^m = r \tag{2}$$

in Stage 3. This is because the monopolist can extract all the value of the flexible resource. The investment of the provider is then determined by

$$\Omega(L_1, L_2, K) = 1 - \int_0^{L_1} \int_0^{L_2+K} h(x,y)dydx - \int_{L_1}^{L_1+K} \int_0^{L_1+L_2+K-x} h(x,y)dydx = c_K, \tag{3}$$

which equates the monopolist's private benefits from investment with the costs.

In Stage 1, firm 1, expecting that $p^m = r$, will choose its local capacity L_1 so as to maximize its profit:

$$\max_{L_1} \int_0^{L_1} xf(x)dx + \int_{L_1}^{\infty} L_1 f(x)dx - c_L L_1.$$[14]

The first term shows that the whole demand is served when demand is below local capacity. The second term shows that capacity is saturated when demand exceeds local capacity. Finally, the third term represents the total spending in local capacity. Then, the equilibrium investment in L_1 is determined by

$$1 - F(L_1) \leq c_L. \tag{4}$$

The second-order condition is also satisfied.

Similarly for firm 2, the equilibrium investment in L_2 is determined by

$$1 - G(L_2) \leq c_L. \tag{5}$$

The market equilibrium is characterized by Equations (2), (3), (4) and (5). Let L_1^m, L_2^m and K^m denote the solutions to this set of equations. It should be clear that the

[14]The firm only gets positive profit from investing in its local capacity because the surplus of the consumers, who are served by utilizing the flexible resource, is extracted entirely.

monopoly outcome diverges from the social optimum because firms invest in a positive amount of local capacities.

Proposition 2. *The monopolistic provider underinvests in the flexible resource relative to the social optimum ($K^m < K^*$), whereas the firms overinvest in their local capacity ($L_1^m = L_2^m > L^*$).*

Proof. See Appendix B. □

The intuition is that the monopolist sells the flexible resource at the monopoly price, which extracts all consumer surplus. Anticipating this, the firm will invest in local resources L even if L is less efficient compared with K, in order to gain part of the consumer surplus. As a consequence, the benefit of investing in flexible resources for the monopolist is lower than that for the social planner, and hence the monopolist underinvests. Notice that Proposition 2 holds more generally for any rationing rule. The reason is that firms pay the monopoly price and, thus, all their profits in serving a customer with the flexible resource are extracted. As a result, rationing rules do not affect firms' profits and hence their investment incentives.

Turning to the impact of correlation, we have the following proposition.

Proposition 3. *Under monopoly, there is a positive amount of local investment by the firms (i.e., $L_1^m = L_2^m > 0$). Moreover, there exists a cutoff level $\underline{c}$ such that the monopolist's investment in flexible resources K^m increases with demand correlation α if $c_K \leq \underline{c}$, but decreases with demand correlation if $c_K > \underline{c}$.*

Proof. See Appendix C. □

The impact of an increase in demand correlation on both socially optimal and equilibrium investments depends on whether the flexible resource is significantly more efficient than the local resource, which is consistent with the intuition in Proposition 1. However, the monopolist's investment is more likely to be decreasing in demand correlation, as shown in the following proposition:

Proposition 4. *The smallest c_K above which investment in flexible resources decreases with demand correlation is larger at the social optimum than that under monopoly, i.e., $\bar{c} > \underline{c}$.*

Proof. See Appendix D. □

The intuition is that local investment is zero at the social optimum but positive in the monopoly case. Thus, while the monopolist has to bear the risk associated with low demand realizations under which the firms will rely on their local capacity only, there is no such risk under the social optimum, which makes the social planner more willing to invest in flexible resources as demand correlation increases compared to the monopolist.

5 Discussion and Conclusion

This paper has analyzed firms' incentives to invest in local and flexible resources when demand is uncertain and correlated. I find that market power of the monopolist providing flexible resources distorts investment incentives—namely, underinvestment in the flexible resource and overinvestment in the local resource. Moreover, the impact of demand correlation on investments depends on the investment cost: if the cost is small, investments in flexible resources under social optimum and monopoly both increase with correlation; if the cost is large, they go in opposite direction. These results have important implications for investments in the market for cloud computing. For instance, it is often argued that the emergence of cloud computing reduces the costs of computing power and storage significantly. While the marginal cost of producing an extra unit of computing power is close to zero, the costs of electricity for powering up thousands of machines and cooling them, as well as managing, maintaining, and implementing the relevant hardware and software used in a large server farm are far from negligible.[15] Moreover, although storing a large amount of data may be inexpensive, managing large data sets and retrieving the relevant data at the right time can be costly. Admittedly, the cloud computing market is growing unpredictably and there is no clear indication or consensus on how it will develop. For the time being, this paper shows that analyzing data both on cost and on demand represents a useful first step towards a fuller understanding of the nascent industry.

Here are some important topics that lie beyond the scope of this paper but would be appropriate for further work. The first is to consider product differentiation. For

[15] The New York Times reported that datacenters worldwide use about 30 billion watts of electricity, which is roughly equivalent to the output of 30 nuclear power plants. See "Power, Pollution and the Internet," *The New York Times*, 22 September 2012, available at `http://nyti.ms/NFuUvJ`.

example, assuming that cloud computing services (such as Dropbox storage services) and local storage services are differentiated—how, then, would the investment strategy change? Second, it would be interesting to study the consequences of vertical integration. For instance, what will happen if upstream cloud computing firms such as Microsoft and Google also enter the downstream market of software applications? Third, it would be interesting to think about how spot pricing (i.e., state-dependent pricing) can change investment incentives in an environment with both flexible and local resources. Although spot pricing is rarely used in the cloud computing market (except for AWS), electricity is bought and sold at spot prices in electricity markets.[16] And as argued by Carr (2005) and Jeff Bezos in Stone (2013), today's IT supply would likely transform from companies' private capacity into a centralized utility service, just as electricity became a utility a century ago.

Finally, one could examine the potential merits of policies that can remedy the underinvestment problem. Since surplus appropriation originates from market power, it seems reasonable to investigate whether introducing more competition in the market for flexible resources—thereby forcing down the price—would incentivize the providers and the firms to behave optimally. The characterization of the equilibrium under competition, however, remains a challenge because in the related model of capacity-price game of de Frutos and Fabra (2011), the equilibrium in the pricing stage is a mixed-strategy equilibrium and provider capacity choices are asymmetric, even if they consider one type of resources and exogenous demand. In a model with two types of resources, the demand for flexible resources becomes endogenous. More specifically, it is determined by L and α, which, in turn, affect the providers' mixed strategy in prices. I leave the details of such a model for future work.

References

Angelou, G. and A. Economides (2005). Flexible ICT investment analysis using Real Options. *International Journal of Technology, Policy and Management 5*(2), 146–166.

Anupindi, R. and L. Jiang (2008). Capacity Investment under Postponement Strategies,

[16]For analyses of spot pricing in electricity markets, see Borenstein and Holland (2005) and Murphy and Smeers (2005).

Market Competition, and Demand Uncertainty. *Management Science 54*(11), 1876–1890.

Armbrust, M., A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia (2009). Above the clouds: A Berkeley view of cloud computing. Technical Report No. UCB/EECS-2009-28. available at `http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html`.

Bayrak, E., J. Conley, and S. Wilkie (2011). The Economics of Cloud Computing. *The Korean Economic Review 27*(2), 203–230.

Benaroch, M. and R. Kauffman (1999). A Case for Using Real Options Pricing Analysis to Evaluate Information Technology Project Investments. *Information Systems Research 10*(1), 70–86.

Bish, E. and Q. Wang (2004). Optimal Investment Strategies for Flexible Resources, Considering Pricing and Correlated Demands. *Operations Research 52*(6), 954–964.

Borenstein, S. and S. Holland (2005). On the Efficiency of Competitive Electricity Markets with Time-invariant Retail Prices. *RAND Journal of Economics 36*(3), 469–493.

Carr, N. (Spring 2005). The End of Corporate Computing. *MIT Sloan Management Review*, 67–73.

de Frutos, M. and N. Fabra (2011). Endogenous Capacities and Price Competition: The role of demand uncertainty. *International Journal of Industrial Organization 29*, 399–411.

Fershtman, C. and N. Gandal (2012). Migration to the Cloud Ecosystem: Ushering in a New Generation of Platform Competition. *Communications & Strategies Digiworld Economic Journal 85*(1), 109–123.

Goyal, M. and S. Netessine (2007). Strategic Technology Choice and Capacity Investment under Demand Uncertainty. *Management Science 53*(2), 192–207.

Gumbel, E. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association 55*(292), 698–707.

Harms, R. and M. Yamartino (2010). The Economics of the Cloud. Microsoft White Paper.

Kauffman, R., H. Lucas, P. Tallon, A. Whinston, and K. Zhu (2002). Using Real Options Analysis for Evaluating Uncertain Investments in Information Technology: Insights from the ICIS 2001 Debate. *Communications of the Association for Information Systems 9*, 136–167.

Kreps, D. and J. Scheinkman (1983). Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes. *The Bell Journal of Economics 14*(2), 326–337.

Lee, I. (2009). A Model for Determining the Optimal Capacity Investment for Utility Computing. In D. Neumann, M. Baker, J. Altmann, and O. Rana (Eds.), *Economic Models and Algorithms for Distributed Systems*. Basel, Switzerland: Birkhäuser Verlag.

Marston, S., Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi (2011). Cloud Computing - The Business Perspective. *Decision Support Systems 51*, 176–189.

Murphy, F. and Y. Smeers (2005). Generation Capacity Expansion in Imperfectly Competitive Restructured Electricity Markets. *Operations Research 53*(4), 646–661.

Niyato, D., S. Chaisiri, and B.-S. Lee (7-11 December 2009). Economic Analysis of Resource Market in Cloud Computing Environment. Paper presented to the 2009 IEEE Asia-Pacific Services Computing Conference, Singapore.

Reynolds, S. and B. Wilson (2000). Bertrand-Edgeworth Competition, Demand Uncertainty, and Asymmetric Outcomes. *Journal of Economic Theory 92*, 122–141.

Stone, B. (2013). *The Everything Store: Jeff Bezos and the Age of Amazon.* United States: Little, Brown and Company.

Tong Wang (2014). Migration to the Cloud. Working Paper. available at `https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIOC2014&paper_id=162`.

Van Miegham, J. (1998). Investment Strategies for Flexible Resources. *Management Science 44*(8), 1071–1078.

Appendices

A Proof of Proposition 1

The social optimum is obtained by differentiating Equation (1) with respect to L_1, L_2 and K.

The F.O.C. with respect to L_1 is given by

$$\{IV\} + \{V\} \leq c_L.$$

Similarly, the F.O.C. with respect to L_2 is

$$\{III\} + \{V\} \leq c_L.$$

Finally, the F.O.C. with respect to K is

$$\{III\} + \{IV\} + \{V\} \leq c_K.$$

As $\{III\} + \{IV\} + \{V\} > \{IV\} + \{V\}$ or $\{III\} + \{V\}$, the marginal benefit of investing in flexible resources is always higher than that in local resources. Furthermore, the marginal cost of investing in flexible resources is lower since $c_K < c_L$. Then we must have $L_1^* = L_2^* = 0$. Since $c_K < r$, the F.O.C. with respect to K is satisfied with equality.

The socially optimal investment in flexible resources is then determined by

$$J(K, \alpha, c_K) = \int_0^K \int_0^{K-x} h(x, y) dy dx - 1 + c_K = 0.$$

By implicit function theorem,

$$\frac{\partial K}{\partial \alpha} = -\frac{\frac{\partial J}{\partial \alpha}}{\frac{\partial J}{\partial K}}.$$

We can show that

$$\frac{\partial J}{\partial K} = \int_0^K e^{-K}[1 + \alpha(2e^{-x} - 1)(2e^{x-K} - 1)]dx > 0$$

is positive. Moreover, we have

$$\begin{aligned}\frac{\partial J}{\partial \alpha} &= \int_0^K \int_0^{K-x} e^{-x-y}(2e^{-x} - 1)(2e^{-y} - 1) dy dx \\ &= -e^{-K}[K + 3e^{-K} + 2Ke^{-K} - 3].\end{aligned}$$

It can be shown that there exists a cutoff $\bar{K}^* \approx 2.15$ such that $\frac{\partial J}{\partial \alpha} < 0$ when $K > \bar{K}^*$, and $\frac{\partial J}{\partial \alpha} > 0$ when $K < \bar{K}^*$. Because K^* is decreasing in c_K, if c_K is small such that $K > \bar{K}^*$, then $\frac{\partial K}{\partial \alpha} > 0$. On the contrary, if c_K is large such that $K < \bar{K}^*$, then $\frac{\partial K}{\partial \alpha} < 0$. Moreover, because

$$\int_0^{\bar{K}^*} \int_0^{\bar{K}^*-x} h(x,y)dydx \approx 0.63 > 0.5,$$

we must have $K^* < \bar{K}^*$ when $c_K = 0.5$.

B Proof of Proposition 2

For firm 1, its equilibrium investment is determined by

$$1 - F(L_1) = c_L.$$

As $1 - F(L_1) > \{IV\} + \{V\}$, we must have $L_1^m > L_1^* = 0$, and hence there is overinvestment. The same happens for firm 2.

For the flexible resource provider, its equilibrium investment K^m is determined by

$$\begin{aligned}
\max_K \Pi = & \int_0^{L_1} \int_{L_2}^{L_2+K} (y - L_2)h(x,y)dydx + \int_{L_1}^{L_1+K} \int_0^{L_2} (x - L_1)h(x,y)dydx \\
& + \int_{L_1}^{L_1+K} \int_{L_2}^{L_1+L_2+K-x} (x + y - L_1 - L_2)h(x,y)dydx \\
& + K \left[\int_0^{L_1} \int_{L_2+K}^{\infty} h(x,y)dydx + \int_{L_1+K}^{\infty} \int_0^{L_2} h(x,y)dydx \right. \\
& \left. + \int_{L_1}^{\infty} \int_{L_2}^{\infty} h(x,y)dydx - \int_{L_1}^{L_1+K} \int_{L_2}^{L_1+L_2+K-x} h(x,y)dydx \right] - c_K K,
\end{aligned}$$

which gives us

$$\Omega(L_1^m, L_2^m, K^m) = c_K = \Omega(0, 0, K^*).$$

Suppose that the flexible resource provider invests K such that $L^m + K = K^*$, Since $L^m > 0$, it must be $\Omega(L_1^m, L_2^m, K) < \Omega(0, 0, K^*)$, which means such K cannot be the equilibrium. Therefore, the flexible resource provider must invest K^m such that $L^m + K^m < K^*$, which implies that $K^m < K^*$ (underinvestment).

C Proof of Proposition 3

The monopolist's investment in flexible resources is determined by

$$J^m(K, \alpha, c_K) = \int_0^{L^m} \int_0^{L^m+K} h(x,y)dydx + \int_{L^m}^{L^m+K} \int_0^{2L^m+K-x} h(x,y)dydx - 1 + c_K = 0.$$

By implicit function theorem,
$$\frac{\partial K}{\partial \alpha} = -\frac{\frac{\partial J^m}{\partial \alpha}}{\frac{\partial J^m}{\partial K}}.$$
It is straightforward to show that
$$\begin{aligned}
\frac{\partial J^m}{\partial K} =& \int_0^{L^m} e^{-x-L^m-K}[1+\alpha(2e^{-x}-1)(2e^{-L^m-K}-1)]dx \\
&+ \int_0^{L^m} e^{-y-L^m-K}[1+\alpha(2e^{-y}-1)(2e^{-L^m-K}-1)]dy \\
&+ \int_{L^m}^{L^m+K} e^{-2L^m-K}[1+\alpha(2e^{-x}-1)(2e^{-x-2L^m-K}-1)]dy > 0, \\
\frac{\partial J^m}{\partial \alpha} =& \int_0^{L^m}\int_0^{L^m+K} e^{-x-y}(2e^{-x}-1)(2e^{-y}-1)dydx \\
&+ \int_{L^m}^{L^m+K}\int_0^{2L^m+K-x} e^{-x-y}(2e^{-x}-1)(2e^{-y}-1)dydx.
\end{aligned}$$
Similar to Proposition 1, there exists a cutoff $\bar{K}^m \approx 1.39$ such that $\frac{\partial J^m}{\partial \alpha} < 0$ when $K > \bar{K}^m$, and $\frac{\partial J^m}{\partial \alpha} > 0$ when $K < \bar{K}^m$. Because K^m is decreasing in c_K, if c_K is small such that $K > \bar{K}^m$, then $\frac{\partial K}{\partial \alpha} > 0$. On the contrary, if c_K is large such that $K < \bar{K}^m$, then $\frac{\partial K}{\partial \alpha} < 0$. Moreover, because
$$\int_0^{L^m}\int_0^{L^m+\bar{K}^m} h(x,y)dydx + \int_{L^m}^{L^m+\bar{K}^m}\int_0^{2L^m+\bar{K}^m-x} h(x,y)dydx \approx 0.72 > 0.5,$$
we must have $K^m < \bar{K}^m$ when $c_K = 0.5$.

D Proof of Proposition 4

From Propositions 1 and 3, it suffices to show that $\frac{\partial J}{\partial \alpha}(K^*) < 0$ if $\frac{\partial J^m}{\partial \alpha}(L^m, K^m) < 0$. Notice that both terms integrate the same function,
$$v(x,y) = e^{-x-y}(2e^{-x}-1)(2e^{-y}-1),$$
over the respective area as shown in Figure 2, which can be negative if either x is large or y is large. The difference between $\frac{\partial J}{\partial \alpha}(K^*)$ and $\frac{\partial J^m}{\partial \alpha}(L^m, K^m)$ lies in the shaded area.

If $\frac{\partial J^m}{\partial \alpha}(L^m, K^m) < 0$, we must have $v(x,y) < 0$ at points A and B, where point A has the largest value of y and a small value of x and point B has the largest value of x and a small value of y. Then we must have $v(x,y) < 0$ at all the points inside the shaded triangles: all the points inside the triangle to the left of A have even smaller x and larger y compared to A, and similarly for all the points inside the triangle to the right of B.

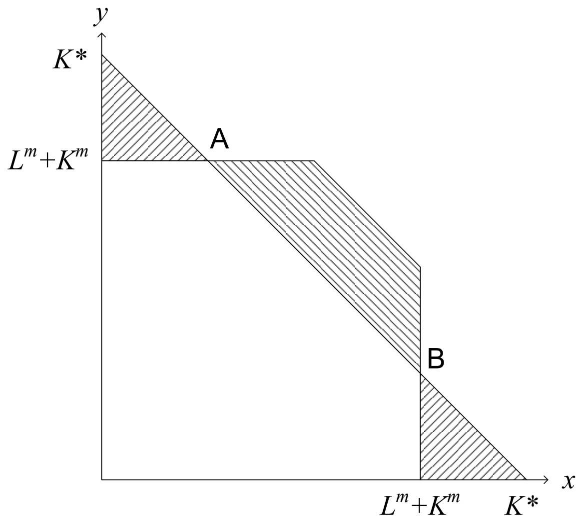


Figure 2: Investments under Social Optimum vs. Monopoly.

Therefore, we must have $\frac{\partial J}{\partial \alpha}(K^*) < 0$. This implies that under social optimum there is a larger range of c_K under which investment increases with correlation as compared to the monopoly case.