Price Elasticity in the Enterprise Computing Resource Market

Qingye Jiang, University of Sydney
Young Choon Lee, Macquarie University
Albert Y. Zomaya, University of Sydney

A study shows that consumers in different segments of the enterprise computing resource market vary in their sensitivity to prices changes, leading to different purchasing behaviors.

Pricing strategies are crucial business decisions in a product's life cycle, especially when multiple competitors exist in the market. In a perfect competition market, multiple suppliers offer similar products, and none of them are powerful enough to control the product's price. To compete for market share, suppliers often use price reduction in promotional campaigns. To maintain competitiveness in the market, competitors are then forced to match these prices, resulting in a series of pricing wars. Many of these "matches" in price reduction are the outcome of business intuition rather than rigorous reasoning, and often occur at the cost of lost revenue and profit. Indeed, it's not uncommon to see enterprises go out of business as a result of pricing wars.

The enterprise computing resource market seems to be a perfect competition market with multiple players. Traditionally, enterprises can buy servers from hardware vendors or rent servers from hosting service providers. In recent years, infrastructure as a service (IaaS) solutions, such as Amazon Web Services (AWS), have been gradually making their way into the enterprise computing resource market. Inspired by the dramatic growth of AWS, many followers have quickly emerged—including Microsoft Windows Azure and Google Compute Engine—and engaged in pricing wars to win customers.

Our study of the enterprise computing resource market focuses on three questions:

- Is the enterprise computing resource market a perfect competition market?
- How sensitive are consumers to price changes in the enterprise computing resource market?
- Is price reduction an effective way to win in the enterprise computing resource market?

 2325-6095/16/$33.00 © 2016 IEEE

Table 1. Segments within the enterprise computing resource market.

Market Segment	Server Sales Business	Server Rental Business	Public Clouds
Major vendors	IBM, HP, Dell, Fujitsu, Oracle	Rackspace, OVH, SoftLayer	Amazon Web Services, Google Compute Engine, Microsoft Windows Azure
Acquisition time	Weeks	Days	Minutes
Expense type	Capital (CapEx)	Operations (OpEx)	OpEx
Unit cost	US$2,000–50,000 (one-time payment)	US$100–2,000 (monthly charge)	US$0.02–6.82 (hourly charge)
Ownership	Consumer	Supplier	Supplier
Operations	Consumer	Supplier	Supplier
Life cycle	3–5 years	1–36 months	Hourly
Contracts	Optional service contract with service-level agreement	Long-term contract with service-level agreement	Service-level agreement
Market size	10 million servers annually/40 million total	2 million servers (total)	1 million servers (total)
Market share	93%	5%	2%

To answer these questions, we study the supply and demand relationships in the global enterprise computing resource market from 2006 to 2013. Our data sources include Gartner's quarterly reports on worldwide server shipments, as well as quarterly and annual reports from Amazon and Rackspace (references to all of our data sources are available at www.qyjohn.net/?p=3507). We quantitatively measure the consumer's sensitivity to price changes using the concept of price elasticity of demand.[1] We analyze the impact of different pricing strategies on the performance of the business, using as case studies the server sales business of IBM, Hewlett Packard (HP), and Dell. This research provides new insights into the microeconomic characteristics of the enterprise computing resource market and can help participants in various market segments design better pricing strategies.

The Enterprise Computing Resource Market

Enterprises have three major options when acquiring computing resources: they can buy new servers from hardware vendors, lease servers from hosting service providers, or use public clouds. As Table 1 shows, these different ways to acquire computing resources represent different segments of the enterprise computing resource market—namely, server sales business, server rental business, and public clouds.

Server Sales Business

The major suppliers in the server sales business are hardware vendors, including IBM, HP, and Dell. Consumers can order servers directly from the vendors or through third-party distributors. The cost to acquire new servers is denoted as a capital expense (CapEx) and is usually paid to the supplier in a single payment before or after hardware delivery. The acquisition time typically ranges from a few weeks for low-end servers up to a few months for high-end servers. When the servers are delivered to the consumer, their ownership is also transferred to the consumer. The server life cycle is three to five years.

Server Rental Business

The major suppliers in the server rental business are hosting service providers, including Rackspace, OVH, and SoftLayer. Consumers usually establish a server rental contract directly with the service provider. According to the contract, the consumer pays a recurring monthly fee to maintain access to designated computing resources through a network. This cost is denoted as an operations expense (OpEx). The time needed to acquire access to computing resources after signing the contract is usually a few days. The server rental contract doesn't change the ownership of computing resources, but some service providers give ownership to the consumer after the successful execution of a long contract (such as three years).

Public Clouds

Major cloud providers are AWS, Google Compute Engine, and Microsoft Windows Azure. Consumers can directly request and gain instant access to computing resources through Web portals or APIs. Similar to public utilities such as gas and electricity,

RELATED WORK ON PRICING STRATEGIES IN THE COMPUTING RESOURCE MARKET

An abundance of literature exists on pricing strategies in the computing resource market, but most discussions focus on maximizing revenue and profit for a particular vendor or service provider within its own customer base. Such research usually assumes a certain resource consumption pattern from the customers, without considering the impact from its competitors and the overall business environment. Only a few researchers study the supply–demand relationship in the computing resource market, which is the focus of our work here.

Gregory Chow studies the technological change and the demand for computers with data on the price and quantity of computers in the US from 1955 to 1965.[1] Kar Yan Tam and Kai Lung Hui extend Chow's work with annual computer hardware spending data in the US from 1955 to 1984.[2] Their findings indicate that the price elasticity of demand for computer hardware is inelastic, but that the degree of elasticity changes over time. The data sources used in both of these research efforts are extremely old, whereas technology advancements have dramatically changed the horizon of the computing resource market in general. On the technology side, expensive mainframes are being replaced by commodity x86-64 hardware. On the business side, server rental and public clouds are becoming increasingly popular among enterprises. Such changes in the demand for computing resources, as well as the relationship between supply and demand, are largely underexplored.

Joanna Stavins studies the price elasticity of demand for the PC market through annual personal computer sales data in the US from 1976 to 1988.[3] Austan Goolsbee compares the price sensitivity on computer purchases from online stores versus in retail stores.[4] Both of these researchers focus on PCs rather than enterprise computing resources.

Finally, Owen Rogers and Al Sadowski use game theory to analyze the effect of price reduction in the public cloud market.[5] They conclude that there's no Nash equilibrium in this game, resulting in a constant stalemate. Cloud service providers practice price reductions to gain media attention and up-sell value-added services at the cost of reduced margin from commoditized services such as storage and compute. The problem with this approach is that unless the profit from the value-added services exceeds the losses from down-pricing, lowering price always results in reduced profit and can threaten the survival of a business.

References

1. G.C. Chow, "Technological Change and the Demand for Computers," *Amer. Economic Rev.*, vol. 57, no. 5, 1967, pp. 1117–1130.
2. K.Y. Tam and K.L. Hui, "Price Elasticity and the Growth of Computer Spending," *IEEE Trans. Eng. Management*, vol. 46, no. 2, 1999, pp. 190–200.
3. J. Stavins, "Estimating Demand Elasticities in a Differentiated Product Industry: The Personal Computer Market," *J. Economics and Business*, vol. 49, no. 4, 1997, pp. 347–367.
4. A. Goolsbee, "Competition in the Computer Industry: Online Versus Retail," *J. Industrial Economics*, vol. 49, no. 4, 2001, pp. 487–499.
5. O. Rogers and A. Sadowski, "Prices: To Cut or Not to Cut? That Is the Question Facing Cloud Providers," 2014, http://tinyurl.com/ozjeozh.

consumers pay only for the amount of computing resources they use for the duration of their usage. Ownership of the physical computing resources always belongs to the service providers. Consumer responsibility is essentially confined to a pay-as-you-go expense.

Price Elasticity of Demand

Here, we introduce the microeconomics concept of *price elasticity of demand*. We then describe its application to the enterprise computing resource market and present our analysis on price elasticity of the three market segments.

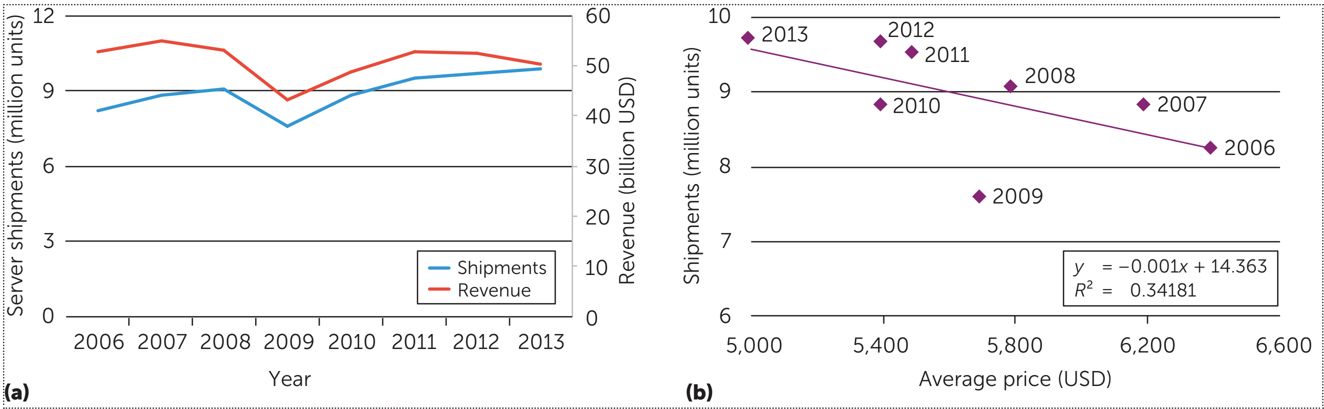

FIGURE 1. The server sales business: (a) the volume of server shipments and revenues on a global scale, and (b) the demand curve of the global server sales business. Both factors were calculated based on data in Gartner's quarterly reports on worldwide server shipments.

Price elasticity of demand is defined by the percentage of change in demand over the percentage of change in price. More formally,

$$E = \frac{(dQ/Q)}{(dP/P)}, \tag{1}$$

where E represents the price elasticity of demand, P represents the price, Q represents the quantity sold, and dP and dQ represent the percentage changes (in absolute values) in price and demand, respectively. When E is greater than 1, the percentage change in demand is greater than the percentage change in price, which is classified as *elastic demand*. When E is smaller than 1, the percentage change in demand is smaller than the percentage change in price, which is classified as *inelastic demand*. When E equals 1, the percentage change in demand equals the percentage change in price, which is classified as *unitary demand*. When demand is inelastic, a rise in price leads to a rise in revenue. When demand is elastic, a fall in price leads to a rise in revenue.

In general, two factors contribute to a product's price elasticity of demand. The first factor is the availability of close substitutes. If a buyer has greater choice among close substitutes in consumption, the price elasticity of demand is greater. The second factor is the proportion of a buyer's budget that is devoted to a product. The larger the proportion of budget, the more responsive the quantity demanded is to price changes, and the greater the price elasticity of demand.

To identify the price elasticity of demand in this study, we carry out linear regression with history price and demand data. We can thus express the relationship between price and demand as

$$Q = b_0 + b_1 P.$$

The slope of a linear curve b_1 can be expressed as

$$b_1 = dQ/dP.$$

Therefore,

$$E = -b_1(P/Q). \tag{2}$$

In the enterprise computing resource market, vendors and service providers offer computation units—such as servers and Elastic Compute Cloud (EC2) instances—with different capacities. In this study, we carry out price elasticity analysis from a statistics viewpoint. That is, we ignore the differences in the actual computation units, focusing instead on the total amount and the average price of the computation units.

Price Elasticity of Server Sales Business

Figure 1 shows the volume of server shipments and revenues, as well as the demand curve for the server sales business. The significant drop in both server shipments and revenue in 2009 was the result of the global economic recession. Using Equation 2, we calculate the price elasticity of demand for the worldwide server market in 2013 as 0.51—that is, extremely inelastic.

For most enterprise consumers who need computing resources, the first option that comes to mind is to purchase server hardware. Other forms of computing resources aren't close substitutes for

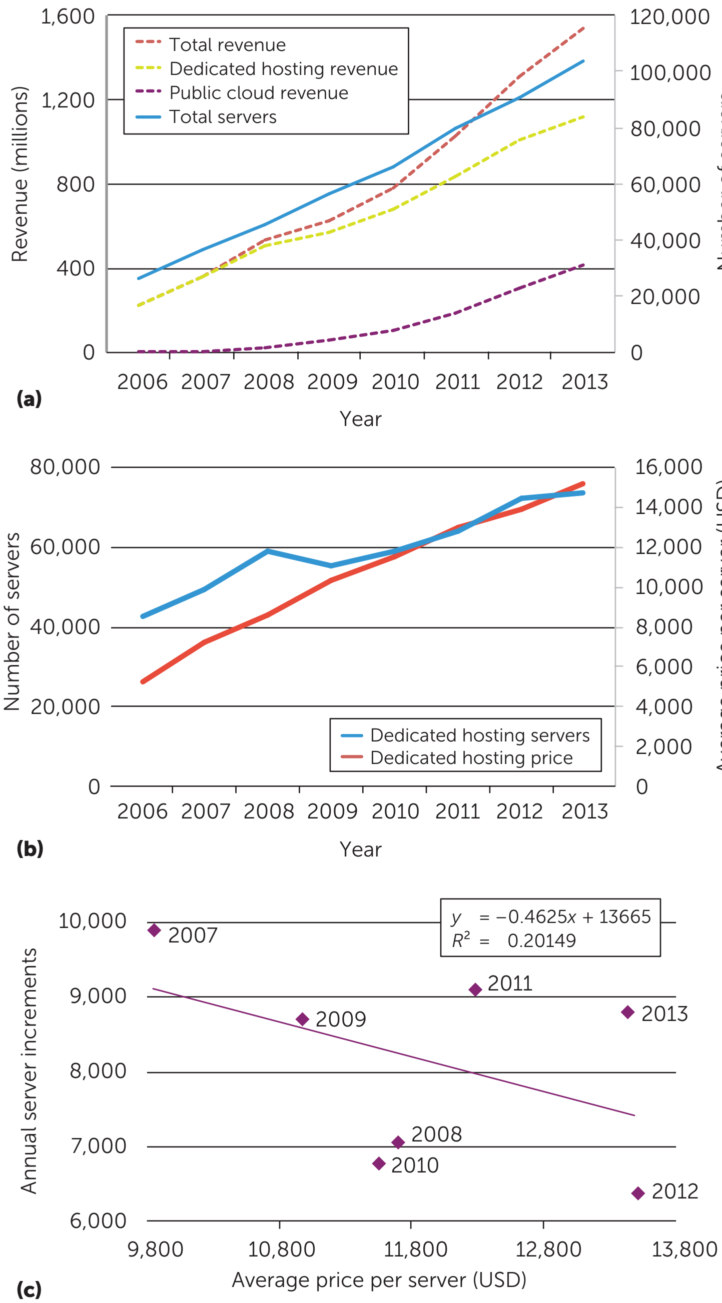

FIGURE 2. Server rental business at Rackspace: (a) total servers and revenues, (b) annual server growth and average server price, and (c) demand curve for the company's server rental business. All figures are calculated based on the data in Rackspace's annual reports to the US Securities and Exchange Commission.

various reasons; for example, renting server hardware from hosting service providers might pose security problems, while public clouds have yet to be widely adopted. From a budget perspective, server hardware purchases, as fixed-asset investments, are usually decided at the organization level and are only a small proportion of an organization's annual budget. The lack of close substitutes and the small budget proportion combined contribute to the inelastic demand in server hardware. In other words, organizations invest in server hardware based on their business plans, regardless of price changes in the market.

A fundamental characteristic of inelastic demand is that the percentage change in demand is less than the percentage change in price. In this case, price reduction might produce more demand, but will also result in lost revenue. In a perfectly monopoly industry, the vendor would always do the opposite—that is, the vendor would raise the price and harvest more revenue. The problem is that the competition in the server hardware market is extremely intensive, and leading server vendors are cutting prices to win market share. Regardless of the steady growth in worldwide server shipments, the overall revenue from server sales seems to be decreasing, as Figure 1a shows. This results in a decreasing margin from server sales and makes it difficult for server vendors to survive.

Price Elasticity of Server Rental Business

Figure 2a shows Rackspace's number of servers and revenue, as well as the revenue breakdown for dedicated hosting (the server rental business) and the public cloud from 2006 to 2013. Figure 2b shows the estimated number of servers and average revenue per server for Rackspace's dedicated hosting business. The number of worldwide server shipments shown in Figure 1a represents only the newly added computing resource every year. Furthermore, the life cycle of server hardware is usually longer than one year; in most cases, it is between three and five years. The number of servers for Rackspace's dedicated hosting business represents the accumulated computing resource since 1998. To facilitate a fair comparison, we calculate the annual server increments for Rackspace's dedicated hosting business and compare that number to the server sales business. Figure 2c shows a supply and demand plot, with the average revenue per server on the horizontal axis and the annual server increments on the vertical axis. The calculated price elasticity of demand for Rackspace's dedicated hosting business is 0.70 in 2013, which is modestly inelastic.

Enterprise consumers choose server rental over buying server hardware for various reasons. For small and medium businesses, the cost and effort to build and maintain a small datacenter with the appropriate networking capacity is overwhelming. Many server hosting service providers are available at both international and local levels. Enterprises

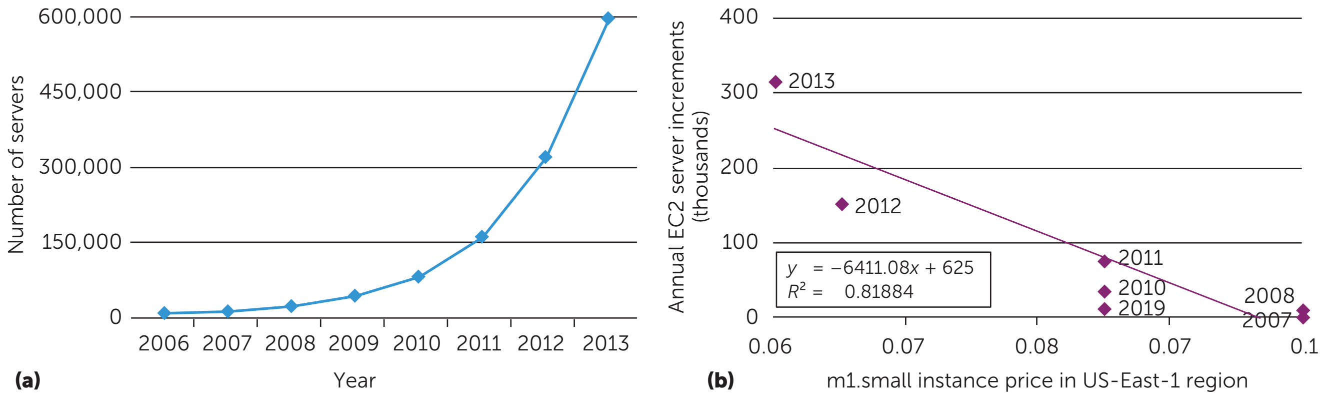

FIGURE 3. Public clouds: (a) number of AWS EC2 servers, physical servers deployed for the AWS EC2 service estimated and adjusted based on Amazon's quarterly earnings report in the second quarter of 2015; and (b) demand curve for AWS EC2 in the US-East-1 region, calculated based on EC2 pricing history and the estimated number of EC2 servers.

must typically complete a comparison and selection process before deciding which provider to work with; then, once a decision is made, a long-term contract with the selected service provider is set up. The contract's length is usually measured in months or even years, depending on the enterprise's business plan. Such a long-term contract effectively eliminates the possibility of migrating to alternative options when a price change occurs in the market. Thus, enterprise consumers invest in server rentals based on business plans, which are usually not sensitive to price changes. This is why the price elasticity of demand for the server rental business is inelastic.

Price Elasticity of Public Clouds

For the public cloud business, we carry out our analysis using the estimated number of physical servers deployed for the AWS EC2 service based on estimations from various cloud computing professionals (see www.qyjohn.net/?p=3507 for a complete list of such estimations). AWS is currently the largest public cloud service provider, and is commonly recognized as far larger than all its competitors combined. Figure 3a shows the estimated number of servers for AWS EC2 on a global scale. Figure 3b shows a supply and demand plot of the estimated amount of annual EC2 server increments and the price of an m1.small EC2 instance in the US-East-1 region from 2008 to 2013, with a linear regression trend line. In 2013, the calculated price elasticity of demand is 1.20, which is modestly elastic.

Many consumers consider EC2 an alternative to traditional server rental services and virtual private server (VPS) services, ignoring advanced features such as CloudWatch, AutoScaling, and Elastic Load Balancing. Furthermore, spending on EC2 as an operations expense is usually decided at the project level—and is usually a big portion of a project's budget. The availability of close substitutes and the large proportion of budget combined contribute to the elastic demand for EC2. In other words, enterprise consumers spend on EC2 based on actual business needs rather than the business plan, and they are sensitive to price changes because of budget limitations.

In general, inelastic demand is closely related to planned buying, whereas elastic demand is closely related to unplanned buying. With the fixed-asset model, when enterprise consumers need computing resources, the budget for the needed computing resources must be secured before placing an order. It then takes weeks, if not months, for the needed computing resources to be delivered and deployed. With the utility model, consumers have instant access to computing resources with no capital expense, zero lead time, and affordable prices. In other words, the utility model makes it possible for customers to react to price changes in a timely manner.

Case Studies

We now analyze the impact of different pricing strategies on the performance of the business, using IBM, HP, and Dell as case studies; together, they represent more than 75 percent of the market share. At the time of our study, other vendors included Sun (which was later acquired by Oracle), Fujitsu, Cisco, and Lenovo, which together represented only 25 percent of the market share.

Figure 4a shows the annual server shipments and average prices for IBM, HP, and Dell. The significant gaps in average prices indicate that the three vendors aren't competing in the same fine-grained market segment. IBM dominates the high-end server market, HP occupies the mid-range server market, and Dell dominates the low-end server market.

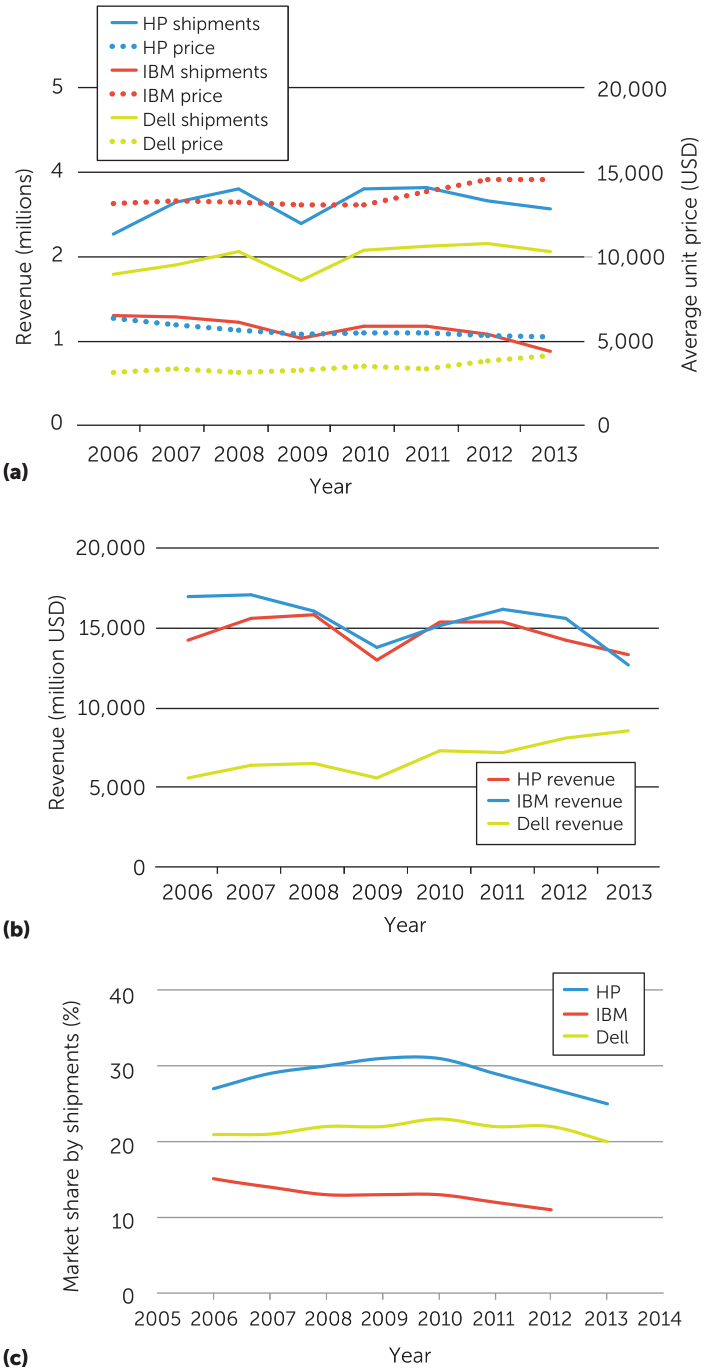

FIGURE 4. Case studies for HP, IBM, and Dell: (a) server shipments and average prices, (b) annual revenue, and (c) market size. All figures were calculated based on data in Gartner's quarterly reports on worldwide server shipments (see www.qyjohn.net/?p=3507).

For IBM, the average price doesn't change much from 2006 to 2010. The slight decrease in server shipments indicates that the market preference is gradually shifting from high-end servers to mid-range and low-end servers. From 2010 to 2012, IBM's average price increases 11 percent, with a corresponding 8 percent decrease in server shipments. The price elasticity of demand for IBM servers is 0.73, which is modestly inelastic. For HP, the average price decreases 18 percent from 2006 to 2013, with a corresponding 13 percent increase in server shipments. The price elasticity of demand for HP servers is 0.72, which is very similar to IBM. For Dell, average price increases 28 percent, with a corresponding 15 percent increase in server shipments. The simultaneous increase in price and shipments indicates that the price elasticity of demand for Dell servers is extremely inelastic; demand increases regardless of price increases.

The inelasticity in demand and the significant gaps in average prices suggest that enterprise consumers don't perceive IBM, HP, and Dell servers as close substitutes for each other. Each of these vendors dominates in its own fine-grained market segment, and has the market power over pricing. IBM responded to the inelastic demand with up-pricing. However, the market preference was shifting from high-end servers to mid-range and low-end servers, resulting in the decrease in revenue. HP responded to the inelastic demand with down-pricing, resulting in an increase in shipments but a decrease in revenue. Dell responded to the inelastic demand with up-pricing. Coupled with the shift in market preference, Dell achieves a simultaneous increase in server shipments and revenue. The situation for both HP and Dell can be directly derived from Equation 1: when demand is elastic, a fall in price leads to a fall in revenue, while a rise in price leads to a rise in revenue. Therefore, the HP and Dell cases can be considered classical examples of how the principle of price elasticity works in microeconomics.

Figure 4c shows the market share for IBM, HP, and Dell as defined by the percentage of server shipments. As the figure shows, price reduction isn't an effective way to win market share.

Our study of price elasticity of demand of the global enterprise computing resource market reveals that from a microeconomics viewpoint, public clouds are fundamentally different from traditional server sales and server rental businesses. The difference in business model leads to a different purchasing behavior, which is reflected by the difference in price elasticity of demand. The price elasticity of demand is extremely inelastic for both server sales business and server rental business. Since consumers aren't sensitive to price changes, a rise in price leads to a rise in revenue, whereas a fall in price results in a fall in revenue. The price elasticity of demand is modestly elastic for public clouds. Since consumers are sensitive to price changes, a fall in price results in a rise in revenue. As such, vendors and service providers must carefully evaluate the price elasticity of the product or service they offer when developing pricing strategies.

Reference

1. A. Marshall, *Principles of Economics*, vol. III, Digireads.com Publishing, 2004, ch. IV.

QINGYE JIANG *is pursuing a PhD from the School of Information Technologies at the University of Sydney. His research interests include distributed and high-performance computing, open source communities, and the microeconomics behavior of the information technology industry. Jiang has an MS in civil engineering from University of Illinois at Urbana-Champaign. He's a senior member of IEEE. Contact him at qjiang@ieee.org.*

YOUNG CHOON LEE *is an Australian Research Council Discovery Early Career Researcher Award (DECRA) fellow and a lecturer in the Department of Computing at Macquarie University. His research interests include distributed systems and high-performance computing. Lee has a PhD in computer science from the University of Sydney. Contact him at young.lee@mq.edu.au.*

ALBERT Y. ZOMAYA *is the Chair Professor of High Performance Computing and Networking in the School of Information Technologies at the University of Sydney, where he also directs the Centre for Distributed and High Performance Computing. His research interests include parallel and distributed computing and complex systems. Zomaya has a PhD in automatic control and systems engineering from Sheffield University. He's a chartered engineer and a fellow of the American Association for the Advancement of Science, IEEE, and the Institution of Engineering and Technology. Contact him at albert.zomaya@sydney.edu.au.*

Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.

ACM - IEEE CS ECKERT-MAUCHLY AWARD

Call for Award Nominations • Deadline: 30 March 2016 • www.computer.org/awards • awards.acm.org

ACM and the IEEE Computer Society co-sponsor the **Eckert-Mauchly Award**, which was initiated in 1979. The award is known as the **computer architecture community's most prestigious award**.

The award recognizes outstanding contributions to computer and digital systems architecture. **It comes with a certificate and a $5,000 prize**.

The award was named for John Presper Eckert and John William Mauchly, who collaborated on the design and construction of the Electronic Numerical Integrator and Computer (ENIAC), the first large-scale electronic computing machine, which was completed in 1947.

TO BE PRESENTED AT

ISCA 2016
The 43rd International Symposium on Computer Architecture

Nomination Guidelines:
• Open to all. Anyone may nominate. • Self-nominations are not accepted. • This award requires 3 endorsements.

Questions? Write to IEEE Computer Society Awards Administrator at **awards@computer.org** or the ACM Awards Committee Liaison at **acm-awards@acm.org**

IEEE · IEEE computer society · acm Association for Computing Machinery