Price-at-Risk: A methodology for pricing utility computing services

by G. A. Paleologo

Whereas most companies use the century-old cost-plus pricing, this pricing method is especially inadequate for services on demand because these services have uncertain demand, high development costs, and a short life cycle. In this paper we propose a novel methodology, Price-at-Risk, that explicitly takes into account uncertainty in the pricing decision. By explicitly modeling contingent factors, such as uncertain rate of adoption or demand elasticity, the methodology can account for risk before the pricing decision is taken. The methodology optimizes the expected "net present value," subject to financial performance constraints, and thus improves on both the cost-based and value-based approaches found in the marketing literature.

Pricing is a crucial business decision in the life of a product. A minor adjustment in price can dramatically affect the profitability of the product, its diffusion in the market, and its ultimate success. Like many corporate decision processes, pricing is driven partly by rational reasoning, partly by established practice, and partly by "black magic" (not necessarily in this order). In the information technology (IT) sector, pricing falls mostly in two classes. For an IT product, such as a hardware device or a software license upgrade, the development costs are small compared to the high initial sunk costs. For example, the costs associated with the production of a new CPU are small compared to the cost of a chip manufacturing plant. The pricing of IT services has strong similarities with instances of pricing in the retail industry. Hardware equipment is sold on a per-unit basis, but this simple unit pricing is supplemented by a variety of price schedule modifications, such as quantity discounts, bundling, and market skimming (gradual price reduction) and dealing (temporary price cutting). Conversely, for IT services, such as services in outsourcing contracts, a fixed-price contract is dominant.

In recent years, IBM has promoted a third way to provision information services[1] (other companies have made similar proposals). *Utility computing services* deliver information services when needed, in such a way that customers neither incur the high fixed costs of purchasing hardware and software, nor commit to long-term fixed-price outsourcing contracts. Instead, they receive the service they need and pay only for what they use. Utility computing services represent a departure from the current ways of doing business. On one hand, they feature attributes that appeal to customers: short lead times in service provisioning, high reliability and survivability, customized service level agreements, a reduced learning curve in the adoption of a new service, and easy access to new technology. On the other hand, utility computing services have direct financial benefits for the customer. These benefits come about in two distinct ways. First, utility computing services reduce the risk faced by the customer because the costs to the customer are proportional to the volume of transactions performed during a certain time interval (say,

0018-8670/04/$5.00 © 2004 IBM

Table 1 Comparing provider costs and risks among IT products, outsourcing services, and utility-computing services

	Sector		
	Products	Outsourcing services	Utility-computing services
Initial investment	*Low-medium*	*High*	*High*
Demand uncertainty	*High*	*Low*	*High*

a quarter). These transactions are usually correlated with the number of financial transactions performed during the same interval, and therefore with the revenue stream of the customer. Therefore, the cost structure is tied to the revenue. This reduces the downside risk faced by a customer when the revenue falls below target. In this respect, utility computing services represent a risk management instrument for the customer, similar to insurance.

A second financial advantage of utility computing services comes from economies of scale. Utility computing services are designed to run on a shared infrastructure, in which resources can be dynamically shared among customers. As the number of customers grows, the average resource utilization grows because of the statistical multiplexing of customer demand. As a consequence, hardware costs are sublinear in the total volume of transactions. Similarly, labor and software costs do not increase linearly with the size of the infrastructure. The increase in operational efficiency can be translated to lower prices to customers.

While utility computing services deliver distinct benefits to customers of IT services, they pose new challenges for providers:

- **Reduced contract duration.** Contracts for on demand IT services have a minimum duration of one year, and this term could be further reduced in the future. This is in stark contrast with the currently typical terms of five to seven years for outsourcing contracts. Previously, monitoring revenues and ensuring they are in line with forecasts (revenue assurance) was handled within each individual contract. In the new model, the challenge is associated with a portfolio of contracts, and one uncertainty faced by the provider lies in the variable duration of these contracts.

- **Reduced switching costs and customer lock-in.** Although set-up fees and fixed recurring fees are also part of utility-computing service contracts, they constitute a smaller percentage of the cumulative revenue. In turn, this facilitates the migration of customers among providers.

- **Uncertain customer demand.** The core of the realized revenue is variable; that is, it is proportional to customer demand. With a small customer base consisting of few customers, the provider faces the risk associated with fluctuations in demand. If the customer base is sufficiently large, the fluctuations have less impact on the profitability of the offering, and the provider only faces the risks associated with the industry sector in which the customers operate. In either case, the risk faced by the provider is higher than in the previous outsourcing environment.

- **Short life cycles and high sunk costs.** Durations of utility-computing service contracts are already shorter than the life cycles of hardware and software products. Together with low switching costs, the short contract duration allows customers to switch to the newest available technology at little or no cost. As a result, the life cycle of utility-computing service offerings will be short and tightly correlated to technological cycles. Within the cost structure of utility-computing service offerings, sunk costs are much larger than the variable costs. Sunk costs include development costs for instrumentation, provisioning, and monitoring of new services.

Thus, new utility-computing services require significant *ex ante* development and start-up costs in the face of uncertain demand. Compared to the existing pricing practices for IT products and outsourcing services, this is the worst of both worlds. We compare the three sectors in Table 1.

How is pricing affected by the features of an on demand offering? We should point out that three types of decisions are involved: what to price, how to price, and when to price. "What to price" pertains to the

set of attributes associated with the service. Should a content delivery service include guarantees on performance, such as maximum packet loss or maximum latency? Should the service offer HTTP (HyperText Transfer Protocol) and SSL (Secure Sockets Layer) transactions separately, or should these two types of transactions be bundled as a single service? Notice that even a homogenous product can be differentiated by posting prices that depend on volume. This is a special form of bundling, in which multiple units of the same product are bundled together. Therefore, pricing is strongly related to the choice of a product (or service) line. However, assigning prices to each item in the product line ("how to price") is perhaps the most important task in the pricing process and is the only one that is performed by the "pricer" alone.

In the following, we address the problem of pricing a service with on demand attributes. We focus on how to price a single unit of a service delivered to multiple customers by a shared infrastructure. We leave the attributes of specific utility-computing service offerings in the background; that is, we take on the problem of pricing with a given set of attributes. Whereas the attributes are important and affect the pricing decision, their impact is indirect and is captured in the price elasticity. There are additional features of a complete pricing strategy that are missing in our analysis. Most importantly, we consider neither nonlinear-pricing nor dynamic-pricing strategies. Although these aspects are important, we believe that their impact on the pricing decision is secondary when compared to price point setting for a unit of service.

First, we observe that nonlinear pricing approaches (also known as second-degree price discrimination), such as bundling and quantity discounts, are not allowed when service resale is permitted[2]—a very real possibility in the case of IT services. Second, there is circumstantial evidence that the demand level of an individual customer is not sensitive to price. For example, the traffic to a popular Web site is independent of how much the Web site owner is paying to the content provider. Similarly, the load on a corporate database system is generated by the company employees, and is insensitive to the price paid by the company to the IT provider. Given that demand is mostly exogenous, the impact of quantity discounts on the pricing strategy is likely to be less important than the selection of the unit price. Moreover, it should be noted that one of the most distinctive attributes of on demand services is the high level of contractual standardization: prices are publicly available to customers, thus ruling out first- and third-degree price discrimination that posits different prices for different customers.

Dynamic pricing has been used with success in a number of sectors, the two major areas of application being the travel and hotel industries. Like utility-computing services, both industries exhibit high fixed costs. However, a closer look reveals some significant differences. First, while the sectors mentioned above exhibit increasing marginal costs and long lead times for capacity expansion, IT services show that marginal costs are constant or sublinear (thanks to multiplexing gains) and with short lead times compared to contract duration. Second, the contract duration in the hotel and airline sector ranges from hours (a short one-way flight), to days (round-trip flight, or a long hotel stay), whereas an IT contract spans a much longer interval. The immediate consequence of the different cost structures is that, while rationing is unavoidable in airline and hotel reservations, it is virtually absent in IT contracts because the provider can always find it profitable to expand capacity to accommodate new customers, and this expansion requires short lead times. Therefore, although dynamic pricing techniques may prove useful in the future for managing short-term capacity shortages, the case for their current application in other sectors is not as compelling. Although unit prices do change during the lifetime of utility-computing services, the explanation lies in technological innovations modifying the cost structure of the providers and customers' preferences, rather than a predefined time-varying pricing policy. We choose not to consider these effects because they are specific to each offering and can be treated as additional contingencies.

The remainder of the paper is organized as follows. In the next section we review the pricing methodology currently followed by practitioners and the theory behind it. We then introduce our pricing methodology by presenting a problem formulation for pricing on demand services and an algorithm for obtaining a numerical solution. In the last section we summarize our finding and suggest directions for future research.

Pricing: practice and theory

In their seminal paper, Hall and Hitch suggest the pricing process has three stages.[3] First, the firm es-

timates a certain demand level q_{cp} for a new product. Second, it estimates the production costs $C(q_{cp})$ associated with the assumed production volume. Finally, it generates a price by marking up the unit cost of production, the mark-up being a proxy for the gross profit margin (*GPM*) associated with the product:

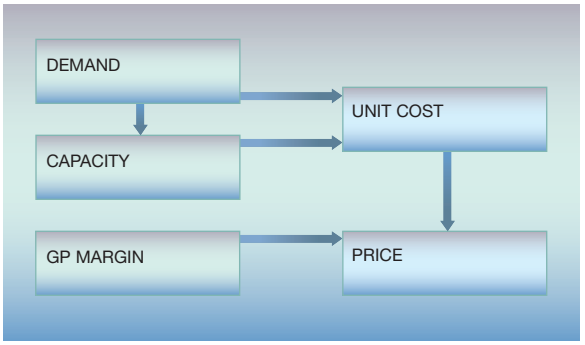
$$p_{cp} = (1 + GPM)C(q_{cp})/q_{cp} \quad (1)$$

The methodology, termed *cost-plus pricing*, is illustrated in Figure 1. Cost-plus pricing is currently widely practiced. Historical evidence suggests that cost-plus pricing has been in use at least since the end of the 18th century. In their 1939 study, Hall and Hitch found that 80 percent of a sample of surveyed companies used this methodology.[3] A study by Skinner found 70 percent of the companies in that sample, and later studies by Shipley found that 59 percent of the companies used cost-plus pricing on all of their product lines, whereas an additional 33 percent of companies used this methodology on some of their product lines.[4,5] A more recent study by Diamantopoulos et al. confirms these findings.[6] All the aforementioned studies concern product pricing. For IT service pricing, a comprehensive study is not yet available. There is ample circumstantial evidence, however, that cost-plus pricing is more often used in service pricing than in product pricing.

Why do companies use cost-plus pricing? Although the question is more than sixty years old, there is still no definitive answer. We list several possible explanations and refer the interested reader to the relevant literature.

- **Bounded rationality.** Companies, like individuals, are not entirely rational agents. Their preferences are not always fully formed, and their ability to process information may be limited. In particular, they often ignore the demand elasticity for a given product. Their mode of operation, then, is not to maximize their profit or a similar objective function, but to obtain a solution that is good enough.
- **Fairness.** Companies set low prices because these prices are perceived as "fair" by customers.[7,8] By foregoing profits in the short run, the company secures the customers' allegiance and trust and thus benefits in the long run.
- **Organizational constraints.** Corporate decisions are the outcome of interactions among organizational entities with possibly conflicting objectives.[9] For example, if the pricing decision follows the investment decision, the person making that pricing decision might not be driven by profit maximization. Instead, the pricer tends to validate the already settled investment decision by assuming that the investment will yield a certain return.

Figure 1 Cost-plus pricing process



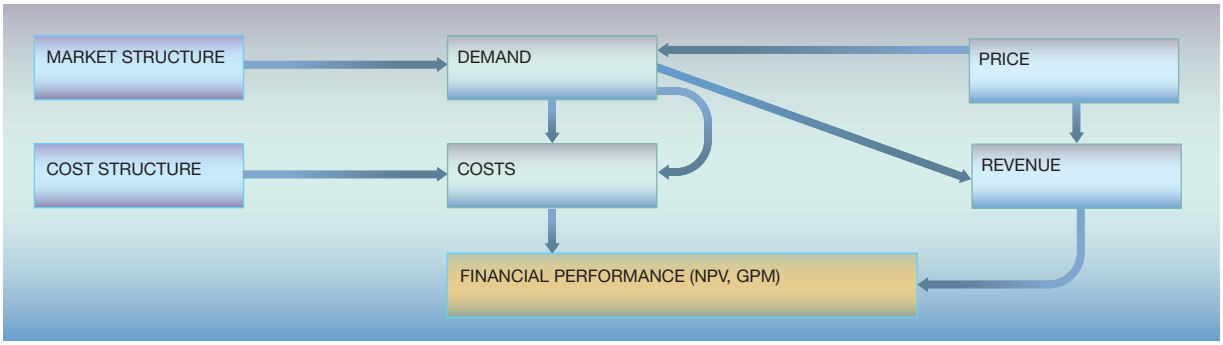
The cost-plus pricing methodology is at odds with the prescriptions suggested by Walrasian economic theory. According to this theory, the two basic tenets of industrial organizations are that prices affect demand, and that decision makers are rational; that is, their goal is to maximize their expected profit. The difference between the solutions prescribed by the cost-plus methodology and the utility-maximization methodology is best illustrated by analysis. We consider the case of a new service that has some differentiating features when compared to substitute services already offered in the market. Furthermore, for the sake of simplicity, we assume that the competitive environment is static; that is, no other services enter the same market, and the prices of the existing services are fixed. In this case, assume there exists an inverse demand function $P(q)$ that maps a certain demand q to the price per unit of service that would generate this demand level. If the average cost of provisioning q service units is $C(q)$, the problem to be solved by the provider is

$$\begin{aligned} &\text{Maximize} \quad qP(q) - C(q) \\ &\text{Subject to} \quad q \geq 0 \end{aligned} \quad (2)$$

We call the solution to this optimization problem the *rational pricing solution* because it maximizes the value of the investment in the case of constant unit prices (see Figure 2).

The differences between cost-plus and rational pricing are best illustrated graphically. Consider first Fig-

Figure 2 Rational pricing process



ure 3. The red curve describes the average unit costs to provide an SU (service unit), say, one GB of storage capacity for the duration of the contract. The curve has a saw-tooth shape, since capacity cannot be increased in continuous increments, but only by adding servers with a fixed minimal capacity. The blue curve describes the inverse demand curve. In the cost-plus methodology, the pricer assumes a demand level q_{cp} and computes the resulting price p_{cp}. The demand at this price, however, is much higher than anticipated, and additional investment is made to increase production. The production output q^* corresponding to this price is much higher than q_{cp}. The average unit costs are higher than forecasted, and the net result is a loss for the provider (the yellow area in Figure 3). Compare the previous analysis to the rational pricing solution illustrated in Figure 4. The optimization problem (2) is equivalent to maximizing the area of the yellow area in Figure 4. The optimal production output is q_{opt}. Interestingly, this value is close to the demand level q_{cp}, but the optimal price p_{opt} is greater than p_{cp}. The graph offers a qualitative interpretation of the most effective pricing strategy. In this example, the provider should take advantage of the insensitivity of demand when price is greater than \$9/SU. Based on the (inverse) demand curve and on the provider's cost structure, the correct positioning of the service is as a premium service. The strategy is to command such a price that the output level is below 300 SU, and thus avoid additional investments in capacity.

Price-at-Risk: A pricing methodology

Although the concept of a demand curve is an elementary and powerful one, it has not been extensively used by practitioners. The main obstacle to its adoption is the difficulty of estimating the demand curve, or equivalently, the price elasticity of demand. The problem is exacerbated in the area of IT services delivered over networks. The two major methods for demand estimation from transactional data, one based on time series and the other based on cross-sectional data, find little application here.[10] Long-term historical data on demand are not available, given the high pace of innovation in this sector. IT services usually cater to few large geographical markets with very different tastes and needs. Thus cross-sectional analysis does not help either. The only available alternative is therefore marketing analysis, such as panel studies, survey research, and focus groups. The resulting estimates of market size, price elasticity, and rate of adoption have a significant margin of error. Considering only the expected demand curve would underestimate the risk associated with the investment and could lead to undesirable decisions. For example, imagine that, based on the available market information, the *expected GPM* is 35 percent. If the distribution of market demand is ignored, these indicators of financial performance might be deemed acceptable by the decision maker. But, what would be the assessment if a more accurate analysis would reveal that with high probability, say 20 percent, the *GPM* could be negative? The outlook on the price and on the investment would change significantly. The goal of the Price-at-Risk (P@R) methodology is to incorporate the essential features of rational pricing, and, at the same time, quantify the uncertainty associated with the market assessment of the demand and incorporate it in the decision process. P@R improves upon the current pricing methodology in three ways.

Figure 3 Cost-plus pricing

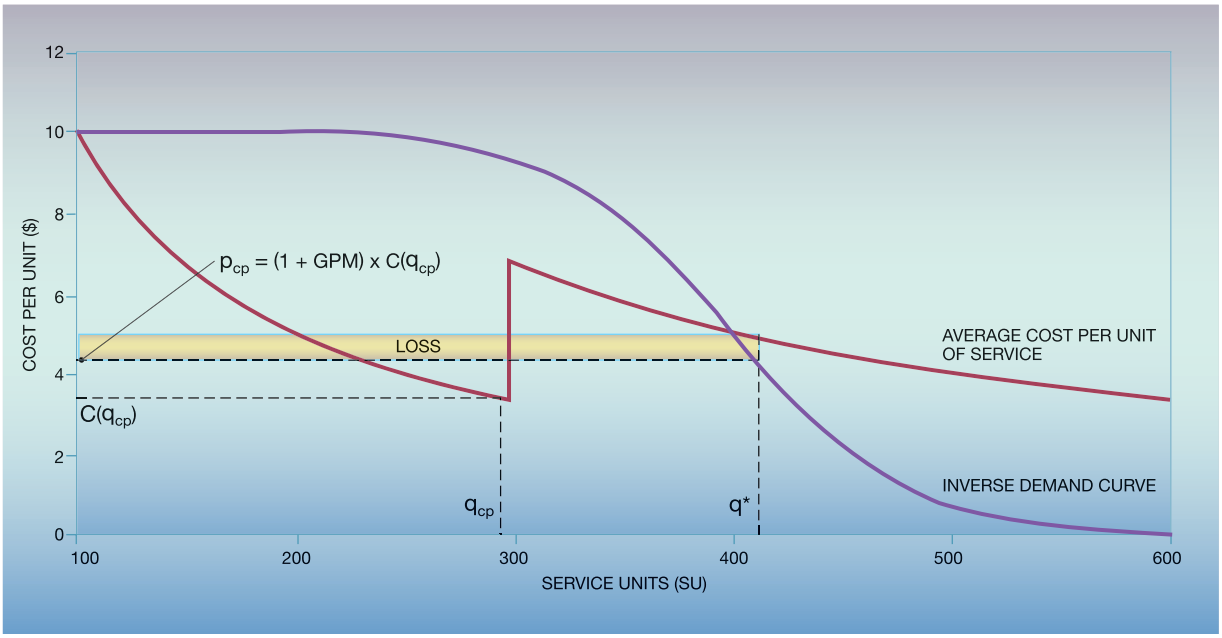


Figure 4 Rational pricing

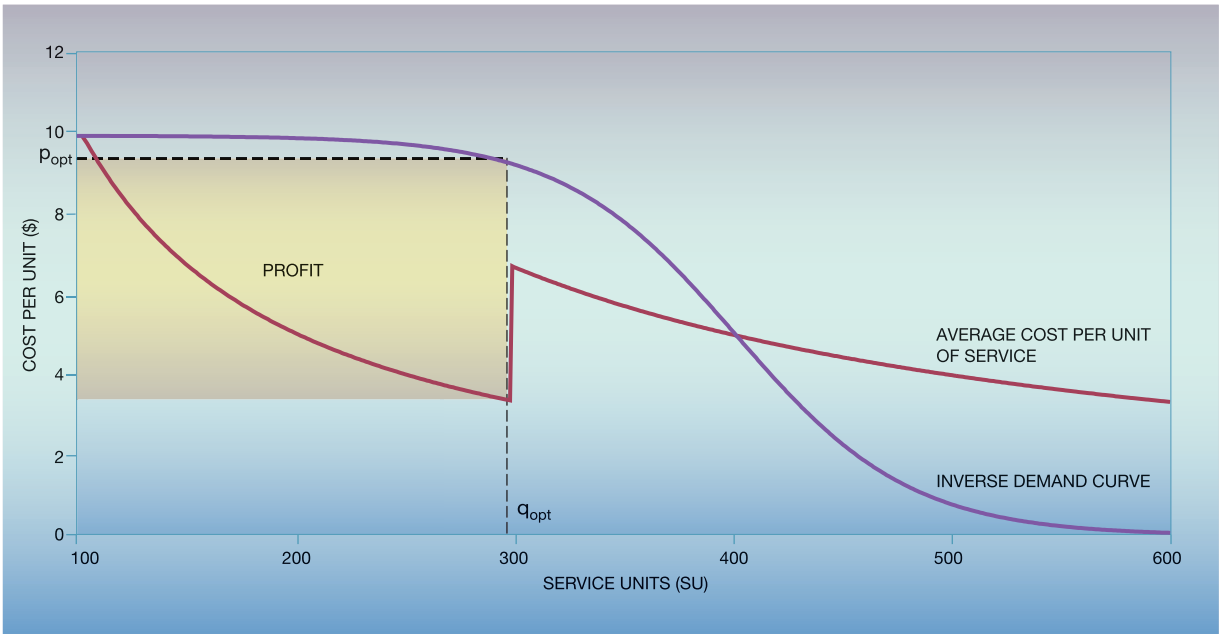
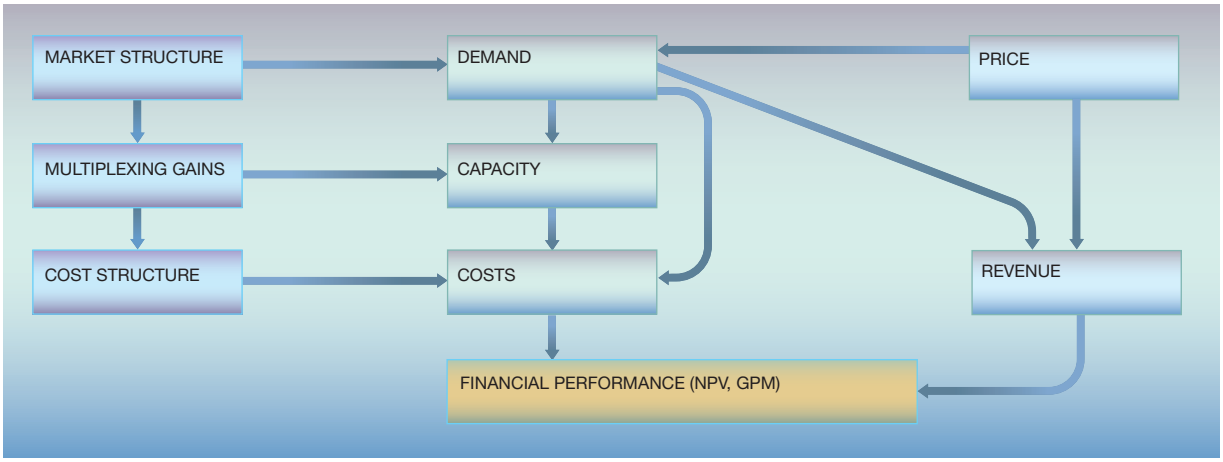


Figure 5 P@R functional model



1. It uses a detailed model specification that takes into account multiplexing gains and decreasing average costs;
2. It takes into account uncertainty by formulating the pricing problem as a stochastic model;
3. As in rational pricing, it uses optimization, and it takes advantage of market information.

Next we describe each of these aspects of the methodology in more detail.

Model specification. The functional model of P@R shown in Figure 5 differs from the model of Figure 2 in that a Capacity entity is present, together with Multiplexing Gains as input needed to determine system capacity. We now examine in detail the functional relationship between the elements of the model.

Market structure. Market structure determines how the market responds to price and how rapidly a new service is adopted. There are three parameters in the model: (1) the market size M, defined as the cumulative number of potential customers during the lifetime of the service; (2) the demand elasticity $\eta(p) = -p/D \times dD/dp$, where $D(p)$ is the percentage of customers subscribing to the service at price p; and (3) the rate of adoption of the service. We use the Bass model of diffusion of a new product to capture the adoption of the service over time.[11,12] Let $F(t)$ be the fraction of customers that have purchased the service for the first time.

The Bass model postulates that the evolution of F is described by the differential equation

$$dF = (a + bF)(1 - F)dt,$$

where a is termed the coefficient of innovation, and b is the coefficient of imitation. This simple model admits an intuitive interpretation and has been consistently validated by empirical studies.

Demand. The number of customers S in the system at time t is given by the formula

$$S(p, t) = M \times D(p) \times F(t), \tag{3}$$

where

$$D(p) = \exp(-\int_0^p \eta(p)/p\, dp)$$

and

$$F(t) = (1 - e^{-(a+b)t})/(b/a \times e^{-(a+b)t} + 1)$$

follow from the definitions of demand elasticity and the Bass diffusion model.

Multiplexing gains. A common feature to all utility-computing service offerings is that the infrastructure is shared dynamically among customers. This means that resources can be allocated to a customer, released when the customer does not need them anymore, and reallocated to another customer. Therefore the total capacity of the system must accommodate the peak of the sum of customers' demands. In the case of static allocation of resources,

or of dedicated resources, the capacity must be higher and equal to the sum of customers' peak demand. As a result, when the infrastructure is shared, the average utilization increases with the total demand served by the system. For example, suppose that a content delivery system serves a single customer whose load oscillates between 5 and 10 SU/Day, averaging 7.5 SU/Day. In order to guarantee service to this customer, we need to allocate a capacity of 10 SU/Day, which is the peak demand rate. The average utilization of the system is 7.5/10 = 75%. However, if there are eight customers with the same but independent demand requirements served by the same infrastructure, then the total demand smoothes out, and we need a total capacity of 66 SU/Day, to serve an average demand of 60 SU/Day. The average utilization is 60/66 = 91%, a 16 percent increase over the single-customer case. As an effect of the multiplexing of demands, as more customers get on board the same delivery system, the peak of the aggregated demand gets closer to the average demand.

In general, economies of scale due to demand aggregation can be expressed as increases in average system utilization $G(S) \in (0,1]$ as the total number of customers S increases. This function is used in the next subsection to derive the system capacity.

Capacity. From the definition of multiplexing gains, and from Equation (3), it follows that the system capacity L needed at time t is given by

$$L(p, t) = S(p, t)/G(S(p, t))$$

Cost structure. Costs are classified in three classes:

- *Sunk costs* are fixed costs that are incurred at the beginning of the investment and are independent of capacity;
- *Fixed avoidable costs* are optional and usually correspond to investments in capacity. After they are incurred, these costs are fixed; that is, they are independent of the actual demand served by the provider;
- *Variable costs* are proportional to the demand served by the provider over a certain horizon.

Below this broad classification lies a rather complex cost structure. For example, all costs may have a one-time and a reoccurring component. One-time charges can be financed over a certain horizon. Variable costs can be one-time costs (such as customer set-up costs) or reoccurring costs (such as customer service costs, costs of failures due to high utilization, and penalties due to infringement of Service Level Agreements). In general, the cost incurred at time t is a function of the demand levels (and associated capacity upgrades) up to time t. We write $K(p, t)$ for

$$K(S(p, 1), S(p, 2), \cdots, S(p, t))$$

Revenue. The revenue accrued in period t is $R(p, t) = p \times S(p, t)$.

Financial performance. The profit accrued in period t is $\pi(p, t) = R(p, t) - K(p, t)$. From the net cash flow stream, we can compute the net present value and the annual gross profit margins.

Stochastic model. In the pricing decision we must take into account the uncertainty inherent in the estimation of input parameter values. Information on the demand is assessed through marketing analysis. Similarly, economies of scale, an essential feature of utility computing services, can only be estimated based on benchmarks and historical data. The P@R approach does not require perfect knowledge of price elasticity, rate of market adoption of a new offering, or economies of scale. Rather, it allows the decision maker to input *worst-case* and *best-case* values of the relevant parameters. For dealing with the demand curve, we use two curves, corresponding respectively to the estimated lower and upper bound for demand (Figure 6). For each price point in Figure 6, the decision maker sets confidence intervals on the market share that can be captured.

Figure 6 Demand curve

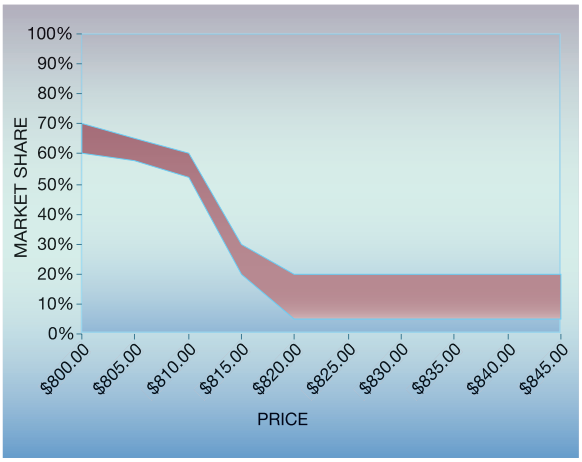
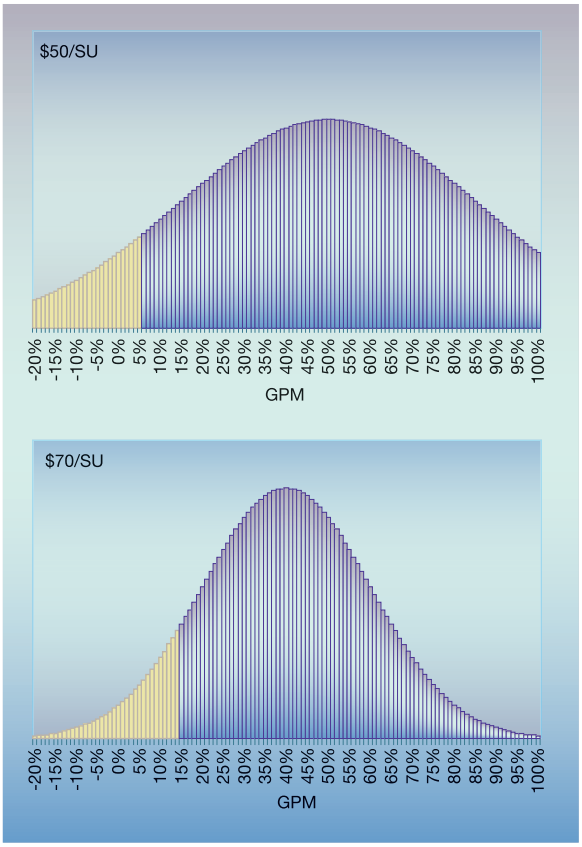


Figure 7 GPM probability density function for two price points



The set of parameter values compatible with the confidence intervals constitutes the *sample space*; a value is sampled for each parameter used in the model, and this set of samples constitutes a scenario ω. P@R computes the financial performance associated with each scenario. By generating a large number of scenarios, P@R develops a probability distribution of the financial performance of a price point. The last step of the methodology, after input specification and scenario generation, is the performance evaluation.

Market information and optimization. P@R considers two indicators of financial performance: gross profit margin (*GPM*) and net present value (*NPV*). *GPM* is used to decide whether a price point is *viable* or not, whereas *NPV* is used to assess the *value* resulting from an offering for a certain price point.

Gross profit margins are used in cost-plus pricing to determine the price at which the offering yields an acceptable return. Over a three-year investment horizon, the decision maker specifies target *GPM* values for each year. P@R retains the concept that *GPMs* should be used as benchmarks of financial performance but uses the entire distribution of *GPM* values, obtained through automatic generation, instead of a single value. To explain how this is accomplished, we give a concrete example. Suppose we are evaluating an offering with a lifetime of one year. We would like to evaluate the relative performance of two price points: \$50/SU and \$70/SU. To do this, we set a *target GPM* (*TGPM*) of 15 percent and a *maximum risk* (*MR*) of 10 percent. The *TGPM* is the minimum profit margin that we target for our offering. We choose a *TGPM* of, say, 15 percent. The maximum risk is the maximum probability we accept for not meeting our *TGPM*. We choose $MR = 10\%$.

The probability density functions of the gross profit margin for the two price points are shown in Figure 7. On the top, the \$50/SU *GPM* density function has an average return of 50 percent, but the *GPM* is less than the target in 17 percent of cases. The yellow area represents 10 percent of cases, and the 10th percentile of the *GPM* is 6 percent, much less than the target. The \$70/SU density function shown on the bottom has an average *GPM* of 40%, sensibly less than in the \$50/SU case (area below the 15% *TGPM* value). However, the 10th percentile of the distribution in this case is exactly 15 percent, which means that we meet the target at least in $(1 - MR) = 90\%$ of cases. Therefore, the \$70 price point is viable, while the \$50 is not.

In its use of percentiles to quantify financial performance, P@R is similar to the VaR (Value-at-Risk) methodology used to evaluate the risk exposure associated with financial decisions.[13] Although economic theory suggests that an expected utility maximization (or the specialized mean-variance) approach is more justifiable, VaR and P@R allow the decision maker to quantify the potential losses and gains in monetary terms, and are naturally suited to quantify the risk-adjusted performance of an investment decision.

The P@R methodology can be extended to multiyear investments. Gross profit margins are used to assess if a price point is viable. The price point is deemed viable if all the target gross profit margins are met with a prespecified probability. Within the range of viable price points, we use the *NPV* as a metric of absolute value. (The choice of the *NPV* is

somewhat subjective; P@R can be adapted to any metric that is considered useful.) From the set of viable prices, the decision maker selects the price point that generates the highest expected *NPV*, and assigns it as the price of the offering. Alternatively, the range of viable prices can be considered as the starting point of a more complex pricing strategy that includes special contractual terms and conditions, such as quantity discounts that add value to the offering and are still compatible with the price range. The P@R tool analyzes the investment over a variable time horizon, although the choice of a three-year period is the most common one. Instead of computing *GPM* for the overall investment, the tool computes for each year the distribution of *GPM* values. The decision maker specifies three target values α_i, $i = 1, 2, 3$, one for each year, and the maximum acceptable risk r that any of the targets will not be met. The optimization problem can be formulated as follows.

$$\text{Maximize Expected } NPV(p, \omega) \tag{5}$$

$$\text{s.t. to Probability}(GPM_i(p, \omega) > a_i \text{ for } i = 1, 2, \cdots, N_{\text{years}}) \geq r \tag{6}$$

P@R simulates a large number of scenarios for each price point, and computes the probability distributions of the gross profit margins. Then it checks that the target *GPM* is met with probability greater or equal than $(1 - r)$.

Numerical solution. The problem stated in Equations (5, 6) is a stochastic program with nonlinear objective function and percentile constraints. It is well known that in general the feasible region is non-convex.[14] A possible approach to finding one point satisfying the Karush-Kuhn-Tucker conditions for this problem could be to estimate the expected values and percentiles for a fixed price point via simulation, and then to optimize with respect to price by using standard optimization methods. This approach has two drawbacks. First, estimating the values of the objective function and of the constraint can be computationally expensive. Second, the algorithm is not ensured to converge because the output of the simulation is an estimate and not the actual expectation. We present an alternative algorithm that employs stochastic approximation techniques. The method updates the price point p_n after each sampling; correspondingly, it updates the Lagrange multiplier λ_n associated with the inequality constraint and is described as follows:

Parameters

$$K > 0$$

Initialization

$$\lambda_0 = 0$$

$$p_0 > 0$$

For each $n = 1$ to N:

$$\text{Set } d_n = (NPV(p_n + K/n, \omega) - NPV(p_n, \omega)) \times n/K \tag{7}$$

$$\text{Set } f(p_n, \omega) = \begin{cases} 1 - r & \text{if all } GPM \text{ requirements are satisfied} \\ -r & \text{otherwise} \end{cases} \tag{8}$$

$$\text{Set } g_n = (f(p_n + K/n, \omega) - f(p_n), \omega)) + n/K$$

$$\text{Set } p_{n+1} = p_n + K/n \times (d_n - \lambda_n \times g_n) \tag{9}$$

$$\text{Set } \lambda_{n+1} = \begin{cases} \lambda_n + \max\{0, \lambda_n - K/n \times f(p_n, \omega)\} & \text{if } \lambda_n > 0 \\ \lambda_n + r & \text{if } \lambda_{n=0} \end{cases} \tag{10}$$

The convergence result can be stated as follows:

Theorem 1: Assume that, for each $p > 0$ $E(NPV^2(p, \omega)) < \infty$ holds. As $N \to \infty$, p_N converges in probability to a price point p^* that satisfies the Karush-Kuhn-Tucker conditions for problem (5, 6).

We relegate the proof of convergence to the Appendix. Note that the price is allowed to take negative values. When $p_n < 0$ we assume that the demand is equal to zero.

Concluding remarks

There is widespread agreement that the pricing procedures currently in use in corporations are inadequate and that an improved pricing methodology should take into account market information. These shortcomings are particularly evident in IT services

that are offered on demand. Their relative short lifetime and high initial investments do not leave room for price adjustments, while at the same time their adoption is highly uncertain due to the pace of technological innovation. The shortcomings of current methodologies notwithstanding, an alternative is not available yet. The main obstacle to innovation lies in the modeling of market parameters, such as market size, price elasticity, and rate of adoption of a new product. In this paper, we have proposed a novel methodology, named Price-at-Risk, whose goal is to include these parameters and to take into account the impact of uncertainty in the decision process. For each parameter of interest, the decision maker enters a confidence interval. A large number of scenarios are simulated, each of them with parameters compatible with the input confidence intervals. A price point is considered viable if the corresponding financial performance, expressed by annual gross profit margins, exceeds a specified target with high probability. The probability represents the risk aversion of the decision maker, and its use is similar in spirit to the Value-at-Risk methodology used in risk management of financial assets. Within the set of viable points, the methodology instructs to choose the points with the highest expected net present value. By modeling explicitly contingent factors, such as the uncertainty in the rate of adoption of an offering or the associated demand elasticity, the methodology can account for risk *before* the pricing decision has been taken. By optimizing the expected net present value, subject to financial performance constraints, the methodology improves upon the simple cost-plus procedure.

Several important features of real-world situations have been not been modeled in this analysis. First, an IT company may have a shared infrastructure with the ability of providing diverse services that are metered and priced independent of each other. In this case, the pricing problem is enriched by a multidimensional demand curve. Second, we ignored interactions among competing firms; the setting is akin to that of a monopolist facing a pricing decision over a finite time horizon. These interactions are of great importance in practice and can take various forms, such as one-time price adjustments, "price wars" leading to small profit margins for all players, and "wars of attrition." It is a challenging research problem to model the strategic interactions among companies in order to provide useful recommendations for the pricing task.

Appendix: Proof of Theorem 1

We sketch here the proof of Theorem 1. We rewrite the problem stated in Equations (5, 6) in a different form:

$$\max E(NPV(p)) \tag{11}$$

$$\text{s. t. } E(h(p)) \geq 0 \tag{12}$$

where

$$h(p) = \begin{cases} 1 - r & \text{if all } GPM \text{ requirements} \\ & \text{are satisfied} \\ -r & \text{otherwise} \end{cases}$$

The stochastic process d_n described by Equation (7) is such that $E(d_n) = (E(NPV(p_n, \omega)) - E(NPV(p_{n-1})) \times n/K$.

From Equation (8) we obtain that $E(f_n(p_n, \omega)) = E(h(p, \omega))$.

Consider the stochastic process (p_n, λ_n) described by Equations (7, 8, 9, 10). We consider the piecewise constant interpolation of this process $(p(t), \lambda(t))$, defined as

$$[p(t), \lambda(t)] = (p_n, \lambda_n) \text{ if } t \in \left(\sum_{m=1}^{n} K/m, \sum_{m=1}^{n+1} K/m \right)$$

Under additional mild technical assumptions (see Reference 15, Ch. 8, Theorem 2.3), one can show that $(p(t), \lambda(t))$ converges in distribution to $(\tilde{p}(t), \tilde{\lambda}(t))$, the solution of the system of ordinary differential equations

$$\frac{dp}{dt} = \frac{d}{dp} E(NPV(\tilde{p})) - \lambda \frac{d}{dp} E(h(\tilde{p}))$$

$$\frac{d\tilde{\lambda}}{dt} = \begin{cases} -E(h(\tilde{p})) & \text{if } \tilde{\lambda} > 0 \\ [-E(h(\tilde{p}))]^+ & \text{if } \tilde{\lambda} = 0 \end{cases} \tag{13}$$

where we use the shorthand $x^+ = \max\{x,0\}$, $x^- = (-x)^+$. The stationary points (p^*, λ^*) of the above system satisfy the conditions:

$$\frac{d}{dp} E(NPV(p^*)) - \tilde{\lambda} \frac{d}{dp} E(h(p^*)) = 0$$

$$E(h(p^*)) \geq 0$$

$$\lambda^* \geq 0$$

$$\lambda^* > 0 \Rightarrow E(h(p^*)) = 0$$

$$[E(h(p^*))]^+ > 0 \Rightarrow \lambda^* = 0$$

These are precisely the Karush-Kuhn-Tucker conditions for problem (5, 6), as formulated by Equations (11, 12). Convergence in distribution to the point (p^*, λ^*) implies convergence in probability.

Cited references

1. *IBM on demand business*, IBM Corporation, http://www.ibm.com/e-business/index.html.
2. R. B. Wilson, *Nonlinear Pricing*, Oxford University Press, Oxford, UK (1993).
3. R. L. Hall and C. J. Hitch, "Price Theory and Business Behaviour," *Oxford Economic Papers* **2**, 12–45 (May 1939).
4. D. D. Shipley, "Pricing Flexibility in British Manufacturing," *Managerial and Decision Economics* **4**, 224–233 (1983).
5. D. D. Shipley, "Dimension of Flexible Price Management," *Quarterly Review of Marketing* **11**, 1–7 (Spring 1986).
6. A. Diamantopoulos, "Pricing: Theory and Evidence—A Literature Review," in *Perspectives on Marketing Management*, M.J. Baker, Editor, John Wiley and Sons, Hoboken, N.J. (1991).
7. R. F. Lanzillotti, "Pricing Objectives in Large Companies," *American Economic Review* **48**, 921–940 (December 1958).
8. A. D. H. Kaplan, J. B. Dirlam, and R. F. Lanzillotti, *Pricing in Big Business*, The Brookings Institution, Washington, DC (1958).
9. R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm*, Prentice Hall, Upper Saddle River, N.J. (1963).
10. L. R. Klein, *An Introduction to Econometrics*, Prentice Hall, Upper Saddle River, N.J. (1962).
11. F. M. Bass, "A New Product Growth Model for Consumer Durables," *Management Science* **15**, 215–227 (January 1969).
12. V. Mahajan, E. Muller, and F. M. Bass, "Diffusion of New Products: Empirical Generalizations and Managerial Uses," *Marketing Science* **14,** No. 3, G79–G88 (1995).
13. D. Duffie and J. Pan, "An Overview of Value at Risk," *Journal of Derivatives*, 7–49 (Spring 1997).
14. A. Prekopa, *Stochastic Programming*, Kluwer Academic Publishers, Dordrecht, Netherland (1995).
15. H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York (1997).

Accepted for publication September 8, 2003.

Giuseppe A. Paleologo *IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown, NY 10598 (gappy@us.ibm.com).* Dr. Paleologo is a research staff member in the Mathematical Department at Watson Research Center and an affiliate of the Netlab Research Group at Stanford University. His research interests are in stochastic processing networks, game theory, and service engineering, with particular emphasis on pricing and evaluation of shared resources. He holds a Ph.D. in management science and engineering and an M.S. degree in engineering-economics systems and in statistics, all from Stanford University.