Roland T. Rust & Ming-Hui Huang

Optimizing Service Productivity

To increase service productivity, many companies utilize automation extensively to reduce the use of labor. However, greater use of automation does not always result in higher service quality, and the effectiveness of automation in providing service hinges on how advanced the technology level is. Departing from the standard perspective in which productivity is simply treated as an output measure of firm performance, the authors propose service productivity as a strategic decision variable; that is, the firm manages the service productivity level to maximize profits. They develop a theory of optimal service productivity that explains when the optimal productivity level will be higher or lower and distinguishes between short-term effects of service productivity due to labor–automation trade-offs and long-term effects due to the advance of technology. The theory, together with the existing literature, inspires the development of three testable empirical hypotheses, which are confirmed using data from more than 700 service companies in two time periods. The research shows that service productivity should be lower when factors (e.g., higher profit margin, higher price) motivate the provision of better service quality and that service productivity should be higher when factors (e.g., higher market concentration, higher wages) discourage the provision of better service quality. The empirical results also provide preliminary evidence that large service companies may tend to be too productive relative to the optimal level and, if so, should place less emphasis (in the short run) on cost reduction through automation and more emphasis on service quality.

Keywords: service productivity, financial impact, marketing metrics, self-service, customer satisfaction, service quality, technology

As service becomes an ever larger part of developed economies, service productivity is increasingly the focus of attention. The management of service productivity requires consideration of both efficiency (productivity) and effectiveness (service quality and customer satisfaction) (Grönroos and Ojasalo 2004; Rayport and Jaworski 2005, p. 1). All else being equal, it is always better to have service that is both more efficient and more effective.

Unfortunately, in reality, increasing service productivity often involves a trade-off, with better service typically requiring more labor intensity, lower productivity, and higher cost. Top executives are continually struggling with the trade-off between improving service to customers and cutting costs by using less (or less expensive) labor. For example, for years, firms have outsourced their call centers overseas and decreased the use of more expensive domestic labor because of cost consideration, but now many firms, such as Dell and United Airlines, have reestablished their domestic call centers to achieve higher-quality service and increased customer satisfaction (Goolsby 2009; McCue 2004). Researchers in marketing have shown that this trade-off between customer satisfaction and productivity exists (Anderson, Fornell, and Rust 1997; Marinova, Ye, and Singh 2008; Singh 2000) and is especially pronounced in the service sector, unlike the goods sector, in which increasing customer satisfaction and increasing productivity often go hand in hand (Deming 1986).

Automation has played a significant role substituting for labor to increase service productivity. Consider telephone customer service. Originally, obtaining customer service over the telephone meant calling the company and talking to a customer service representative. Eventually, though, automated telephone systems became cost-effective and were implemented in many firms because they were cheaper to operate than a fully labor-based system. Yet the menus that must be navigated often frustrate the customer. The seemingly universal sense among customers is that such systems typically provide worse service than would be provided by a fully labor-based customer service system. If the service provided is poor enough, this may reduce sales revenues so much that the cost savings are not justified. For example, Winch (2011) notes that the runner-up for daily irritations in people's lives is automated phone systems. An Accenture (2005) survey finds that despite the increased use of new technologies intended to improve customer service, such as automated telephone systems, poor customer service is the primary reason that consumers switch service

Roland T. Rust is Distinguished University Professor and David Bruce Smith Chair in Marketing and Executive Director of the Center for Excellence in Service and the Center for Complexity in Business, Robert H. Smith School of Business, University of Maryland (e-mail: rrust@rhsmith.umd.edu). Ming-Hui Huang is Distinguished Professor of Electronic Commerce, Department of Information Management, College of Management, National Taiwan University (e-mail: huangmh@ntu.edu.tw). Both authors contributed equally to the article. This research was supported by grants (97-2410-H-002-001-MY2 and 97-2410-H-002-091-MY3) from the National Science Council, Taiwan, and the E. Sun Academic Award from the College of Management, National Taiwan University, and by support from the Center for Excellence in Service at the University of Maryland. The authors thank the Marketing Science Institute; conference attendees at the Frontiers in Service Conference, the Marketing Science Conference, and the INFORMS Conference; and seminar participants at several universities for their helpful comments. Steve Shugan served as guest editor for this article.

ISSN: 0022-2429 (print), 1547-7185 (electronic)

providers. Survey respondents in the United Kingdom reported the most frustrating aspect of interacting with customer service representatives was being kept on hold too long (82%). The survey also indicates that the average consumer in the United Kingdom spends an average of nearly six minutes on hold when seeking assistance over a telephone help line. As a consequence, on average, approximately 10% of respondents reported they had switched service providers because of customer service issues.

This type of trade-off is typical in the service world. More customer service representatives imply better service (meaning, in this case, that it is easier for a customer to contact a person) but lower productivity and higher cost. For a given level of technology, a firm must determine how to trade off service quality versus productivity.

The advance of technology facilitates the effectiveness of automation[1] in providing service.[2] New technology makes possible new service methods that are both more efficient and more effective. For example, airline passengers hated airport check-in kiosks when they were initially employed, but over time, as the technology became more mature, passengers began to enjoy the efficiency benefit it offers. Part of this increase in effectiveness and efficiency is due to improvements in the technology, and part is due to customers becoming more experienced in using it. The Internet, the use of self-service technologies (Baily and Lawrence 2001; Meuter et al. 2000; Meuter et al. 2005), and having customers perform some of the service themselves (i.e., coproduction; Bendapudi and Leone 2003) have all provided great opportunities to reduce the use of frontline service personnel and cut costs. In the best case, if technology has advanced sufficiently, such efficiency improvement efforts may also increase service quality. However, for a given level of technology at a given point in time, there is typically a trade-off between service quality and productivity.

The preceding discussion points out the dilemma that service firms face: At a given technology level, a greater use of automation may increase productivity at the cost of service quality, and the effectiveness of automation in providing service hinges on how advanced the technology level is. To resolve this dilemma, we borrow from different streams of literature and multiple perspectives to develop a theory to explore the trade-off between productivity and service quality, while taking into consideration both the short-term efficiency of automation and the long-term effectiveness of technology.

[1]We refer to "automation" as the use of technology to reduce the need for labor in the provision of service so as to increase input efficiency and "technology level" as the effectiveness of automation in providing quality service.

[2]Our discussion on the distinction of automation versus technology does not exclude other factors that also affect the efficiency and effectiveness of service. Intelligent process design and process improvement methods (e.g., quality programs such as Six Sigma) can improve efficiency and effectiveness simultaneously, resulting in higher productivity and higher service quality (Deming 1986; Juran 1970). How the frontline employees are managed can also have a significant effect on the relationship between productivity and service quality (Marinova, Ye, and Singh 2008).

We first discuss the case in which the firm's goal is to maximize profit by selecting the most profitable level of service productivity. The traditional way of thinking about service productivity is that it is an output measure of performance, and thus higher productivity is better because greater productivity (all else being equal) produces greater profits at the firm level and expands the economy at the aggregate level (Banker, Chang, and Natarajan 2005; Brown and Dev 2000; Brynjolfsson and Hitt 1994). We build on the marketing literature that reveals a trade-off between productivity and service quality for achieving profitability in the service sector (Anderson, Fornell, and Rust 1997). If a trade-off between productivity and service quality exists, the cost savings from productivity may be offset by revenue losses from decreased customer satisfaction, given that better service quality and higher customer satisfaction lead to more customer loyalty, better customer retention, and stronger market performance (e.g., Anderson, Fornell, and Lehmann 1994; Anderson, Fornell, and Mazvancheryl 2004; Fornell et al. 2006; Fornell, Rust, and Dekimpe 2010; Gruca and Rego 2005; Homburg, Koschate, and Hoyer 2005; Keiningham, Perkins-Munn, and Evans 2003; Mittal et al. 2005; Rust and Zahorik 1993; Rust, Zahorik, and Keiningham 1995; Zeithaml, Berry, and Parasuraman 1996). Finally, as the behavioral view (e.g., Cyert and March 1992) suggests, firms may make decisions to satisfy multiple goals in addition to maximizing profit, such as achieving market differentiation and growing market share.

The current work makes several important contributions. We propose a theory of optimal service productivity that conceptualizes the level of service productivity as a strategic decision variable to be optimized by the firm. This theory distinguishes between short-term effects, based on decisions about the trade-offs between the use of service personnel and the use of automation, and long-term structural effects, based on level of technology. Our theory also reveals the conditions under which the optimal level of productivity should be higher or lower. Motivated by the theory and prior literature, we formulate testable hypotheses and observe them empirically using multiple years of data from public firms. The results from the theory and empirical analysis are largely convergent.

Next, we build a theory of when service productivity should be higher or lower and derive several managerially relevant results from the theory. Then, we describe the development of testable hypotheses and the empirical analysis used to test our hypotheses. Following this, we present our results, and we finish with a discussion and conclusions. The Appendix provides details of the derivations underlying the theoretical development.

Theoretical Model

Overview

Productivity is typically defined as units of output divided by units of input. This means that for a manufacturing firm, the goal of increasing productivity is often a reasonable one, given that the quality of the output can be maintained more or less constant, if not improved (Deming 1986). In service, however, there are often trade-offs between productivity and service quality (Anderson, Fornell, and Rust 1997). For example, a call center might have few customer representatives and a long average waiting time, resulting in high productivity and low service quality. Alternatively, the call center might have many customer representatives and a short average waiting time, resulting in low productivity and high service quality. Our aim is to build a theoretical framework that incorporates these effects, placing service productivity decisions within a profit-maximizing decision framework.

We define "service productivity" as dollar sales divided by number of employees (e.g., Basker 2007; Bertschek and Kaiser 2004; Converse 1939).[3] The firm seeks the level of service productivity that will maximize profits, considering the trade-off between labor and automation as productivity inputs and conditional on the level of technology. The service productivity level results from the trade-off between labor and automation, with greater labor intensity often resulting in better service quality and greater value to the customer. The effectiveness of automation in replacing labor depends on how advanced the technology level is.[4] We show that for a given technological level and automation costs, an optimal service productivity level exists. We further explore how the optimal service productivity level is affected by several important determinants, including profit margin, price, market concentration, wages, and factors other than service quality that affect sales.

We explicitly consider the firm's choice of service productivity level as a strategic decision variable to be optimized. We conclude that a firm's profit is affected by the relationship between its actual service productivity level and its optimal productivity level, with the best performance when the service productivity equals the optimal productivity. The optimal productivity level is determined by firm- or industry-specific variables in the marketplace, such as profit margin, price, market concentration, wages, and other factors. Other covariates, as well as unobserved heterogeneity due to industry, may also affect profit.

We first list and support our major assumptions and then provide a more formal development of our theory. The key marketing problem in service productivity is how to most profitably serve the customer—that is, what level of service productivity should be sought to maximize profits.

Assumptions

Assumption 1: Wage rates, cost of automation, and level of technology are fixed in the market in the short run and are known to the firm.

In market economies, wage rates tend to result from supply and demand, over which any individual firm has limited control.[5] Thus, from the firm's point of view, the wage rate for a particular class of employee may be usefully viewed as fixed. Likewise, at any particular point, the cost and effectiveness of automation tend to be exogenous to the firm's short-term decision making. It is generally considered in macroeconomics that wage rates and other input costs adjust slowly to macro shocks (Blinder et al. 1998).

Assumption 2: The firm chooses its level of service productivity to maximize profit.

Assumption 2 reflects the behavioral view of the firm by treating level of service productivity as a strategic decision variable, embedded within a standard microeconomic setup of a profit-maximizing firm. Behavioral theories of the firm predict firm behavior with respect to decisions such as price, output, product lines, product mix, and resource allocation—for example, how a firm chooses and allocates resources to achieve organizational goals (Cyert and March 1992). Reflecting this view, we assume that all service firms must decide the level of productivity to seek, considering that there is likely a trade-off between productivity and service quality. We assume that the objective of the firm is profit maximization and that the level of service productivity is a managerial decision variable whose level is selected to maximize profits. Subsequently, we relax this assumption and consider the case in which the firm has multiple objectives.

Assumption 3: Better service quality results in higher demand.

Better service quality increases demand by increasing customer acquisition (Liu and Homburg 2007), customer retention (Kordupleski, Rust, and Zahorik 1993; Zeithaml, Berry, and Parasuraman 1996), and customer loyalty (Caceres and Paparoidamis 2007). For example, it has been found that there is a positive and significant relationship between customers' perceptions of service quality and their intentions to purchase (e.g., Cronin and Taylor 1992; Parasuraman, Zeithaml, and Berry 1988; Zeithaml, Berry, and Parasuraman 1996). This assumption also closely follows the literature on linking service quality to customer satisfaction and profitability (Anderson, Fornell, and Rust 1997; Reinartz, Thomas, and Kumar 2005; Rust, Moorman, and Dickson 2002; Rust, Zahorik, and Keiningham 1995; Wangenheim and Bayon 2007).

Assumption 4: At a given level of technology, less labor intensity in service decreases service quality.

Existing theory and empirical research supports this generalization (Brown and Dev 2000; Oliva and Sterman 2001). For example, Berry, Parasuraman, and Zeithaml (1988) show that labor input plays a key role in determining customers' perceptions of service quality. Four of the five important service quality dimensions in their research—responsiveness, assurance, empathy, and reliability—all

[3]Many operationalizations of output per input exist, such as sales per employee, sales per labor cost, units per employee hour, and others. Which operationalization should be chosen is often a matter of data availability and comparability.

[4]We thus differentiate between long-term automation effectiveness from an advancing technological level (which can, with proper design, result in provision of higher service quality given a higher level of automation) and short-term automation effects, which generally substitute imperfectly for labor. For example, early airport check-in kiosks resulted in poor service quality because of immature technology, but over time the advance of technology enabled check-in kiosks to provide much better service quality.

[5]It is possible that this assumption could be violated for very large firms (e.g., Wal-Mart) that have disproportionate influence on the wage market.

result most naturally from labor performance. Service quality suffers when firms are unwilling or unable to have labor in position to supply service.

We further base this assumption on some preliminary exploratory data analysis. We examined all service firms that were in the American Customer Satisfaction Index database (Fornell et al. 1996) and the COMPUSTAT North America database for 2002 and 2007 (for a more detailed description of the data, see the "Data" section). Service productivity is inversely related to labor intensity. That is, as labor is used more intensively, productivity goes down. For these firms, we regressed the change in customer satisfaction (from the American Customer Satisfaction Index) from 2002 to 2007 against the change in service productivity over the same time period. The resulting regression predictor equation was

(P1) Change in customer satisfaction = 1.984 – .088 (change in service productivity),

with the slope significant at the .01 level. This confirms that lower labor intensity (inversely related to higher productivity) tends to reduce service quality.

Assumption 5: Automation is more cost-effective than labor in providing service.

One of the main reasons for increasing automation use is that it can reduce costs. For example, it has been found that information technology generates excess returns relative to labor input and is increasingly used to substitute labor for production, which is particularly salient for the financial sector (Dewan and Min 1997). Numerous industry applications of automation have shown the cost savings resulting from substituting automation for labor in providing service, such as the use of automated teller machines in the banking industry (Dewan and Min 1997), sales force automation in marketing to enhance cost competitiveness (Hunter and Perreault 2007), and automation of customer service operations in call centers to cut costs of sales (Karimi, Somers, and Gupta 2001).

Formal Theory

We consider a profit-maximizing firm that produces units of a service product. The firm decides the level of service productivity to pursue to maximize profit. (Subsequently, we discuss the situation in which the firm has multiple objectives.) For simplicity of exposition, we assume that the firm must charge a fixed price, which is determined by the market. (Subsequently, we also consider the situation in which price is not fixed.[6]) The service quality level is a function of the labor per unit (resulting from the selected productivity level) and the automation per unit (with automation per unit weighted by the relative effectiveness of automation, which reflects the level of technology). Sales are a function of service quality, with better service producing more sales.

To formalize these assumptions, let

P = service (labor) productivity;
θ = proportion of labor per unit;
$1 - \theta$ = proportion of automation per unit;
N_S = employee-years[7] by service providers to produce a unit of service;
W_S = wage rate per year by service providers;
A = automation cost;
N_A = employee-years to manage the automation for a unit of service;
W_A = wage rate per year for labor to manage the automation;
α = level of technology (relative effectiveness of automation, compared with labor), $0 \leq \alpha$;
m = profit margin;
R = price per unit;
Q = unit sales; and
Z = factors other than service quality that drive sales.

Then, the total labor cost per unit is $\theta N_S W_S$ (the labor cost of service providers) plus $(1 - \theta)N_A W_A$ (the labor cost for managing the automation). The automation cost per unit is $(1 - \theta)A$. The service provider employee-years to produce a unit of service is θN_S, and the employee-years to manage the automation for a unit of service is $(1 - \theta)N_A$. The service (labor) productivity is sales divided by number of employees, or

$$(1)\quad P = (QR)/\{Q[\theta N_S + (1 - \theta)N_A]\} = R/[\theta N_S + (1 - \theta)N_A],$$

from which we note that the associated proportion of labor per unit can be viewed as a function of the productivity P:

$$(2)\quad \theta(P) = (R - N_A P)/[P(N_S - N_A)].$$

Given that purely human service provision is more labor intensive than automation-assisted service provision, it is evident from Equations 1 and 2 that an inverse relationship exists between productivity P and labor intensity θ.

For emphasis, in the remainder of this section, we express θ as $\theta(P)$ to show that the productivity decision is what determines the level of labor usage. We assume that the service quality, S, results from the firm's labor–automation trade-off, taking into account the relative effectiveness of automation, α (level of technology)[8]:

$$(3)\quad S = \theta(P) + \alpha[1 - \theta(P)] = \alpha + (1 - \alpha)\theta(P).$$

Unit sales,[9] from Assumption 3, are a function of the service quality, S, as well as other factors Z:

$$(4)\quad Q = \gamma S + \eta Z,$$

[6]Industries in which price is highly variable across competitors (e.g., education or management consulting services) may not satisfy the fixed price assumption. However, even in those industries, segmenting the industry by price may approximate the fixed price assumption.

[7]For most services, N_S would typically be a small fraction of an employee-year. The purpose of using a year as the time unit is to be consistent with our definition of service productivity, which is dollar sales (in a year) divided by number of employees (for that year).

[8]This implicitly assumes that customers are homogeneous with respect to service quality. Significant heterogeneity might be addressed by aggregating the effects on more homogeneous market segments.

[9]For this section, we assume that capacity constraints are not binding, in which case sales and demand are equivalent.

where $\gamma > 0$ is the degree to which service quality drives sales and $\eta > 0$ is the relative influence of factors other than service quality in driving sales. We implicitly assume that capacity constraints do not limit sales, and thus the firm can supply the quantity demanded. (Subsequently, we relax this assumption.)

The profit per unit, PPU, is the gross contribution per unit less the labor costs and the automation costs:

$$(5) \quad PPU = mR - \theta N_S W_S - [1 - \theta(P)]N_A W_A - [1 - \theta(P)]A.$$

Thus, the total profit is

$$(6) \quad \Pi = PPU \times Q.$$

The firm wants to maximize Π with respect to service productivity P. Assuming the typical case in which a mix of labor and automation is necessary to provide service (Rayport and Jaworski 2005), the optimal productivity, P*, can be derived (see the Appendix) as follows:

$$(7) \quad P^* = R/[\theta^* N_S + (1 - \theta^*)N_A],$$

where

$$(8) \quad \theta^* = (mR - N_A W_A - A)/(2(N_S W_S - N_A W_A - A) - \{(\gamma\alpha + \eta Z)/[2\gamma(1 - \alpha)]\}).$$

The quantity θ^* can be interpreted as the optimal degree of labor to use in service provision. Table 1 summarizes the notations and theoretical definitions of key variables. In the next section, we summarize the main results that arise from the formal theory.

Theoretical Results

From the preceding expression, comparative statics on P* (see the Appendix) yield a set of profit maximization guidelines for managers with respect to the productivity strategy, with all other factors being held constant. The main results that arise from Equations 7 and 8 are as follows:

Result 1: For a given level of technology, there is an optimal productivity level. This suggests that service productivity can be either too low or too high.

Result 2: As profit margin increases, optimal productivity decreases. Firms in low-margin industries should pay particular attention to productivity, and firms in high-margin industries should be more concerned about service quality.

Result 3: As price increases, optimal productivity decreases. Firms in low-price industries should pay more attention to productivity, and firms in high-price industries should attach more weight to service quality.

Result 4: As wages rise, optimal productivity increases. Firms faced with a high-wage environment should seek higher productivity.

Result 5: As factors other than service quality become more influential in driving sales, optimal productivity increases. When factors other than service quality, such as advertising, convenience, and brand equity, are more influential in driving sales, firms should seek greater productivity.

In addition to these results, which can all be directly and rigorously derived from the theoretical model, we can derive the following tentative results by making a few additional (and perhaps more controversial) assumptions:

Tentative Result 1: The optimal productivity level increases over time. This can be derived from the model (see the Appendix) if we assume that the level of technology increases over time. (We supply additional literature support for this assumption in the "Optimality Hypothesis" section.)

Tentative Result 2: As market concentration increases, optimal productivity increases. We obtain this

TABLE 1
The Notation and Theoretical Definition of Key Variables

Variables	Notation	Theoretical Definitions
Service productivity	P	Service (labor) productivity, which is a managerial decision variable whose level is to be selected to maximize profits and other firm objectives
Proportion of labor per unit	θ	Degree of labor uses in service provision, resulting from the selected productivity level
Proportion of automation per unit	$1 - \theta$	Weighted by the relative effectiveness of automation, which reflects the level of technology
Service quality	S	Resulting from the firm's labor–automation trade-off, taking into account the relative effectiveness of automation
Employee year for service provision	N_S	Using a year as the time unit to be consistent with service productivity as dollar sales (in a year) divided by number of employees (for that year)
Wage rate	W_S	Wage rate per year by service providers
Level of technology	α	Relative effectiveness of automation compared with labor
Profit margin	m	The proportion of a firm's net sales to its gross sales, indicating how much of a firm's sales dollars are profit
Price	R	Price per unit
Market concentration	H	The size of firms in relation to the industry that indicates the amount of competition among them
Optimal productivity	P*	The optimal value of the strategic decision variable, service productivity

result if we assume that service quality has less impact on demand in a more concentrated industry (for literature supporting the underlying assumption, see the discussion of H_3 in the "Determinants of the Optimality Hypothesis" section).

Extensions

Multiple firm objectives. Consistent with modern behavioral theories of the firm, firms may have objectives in addition to profit maximization. We can incorporate this idea by building a multiobjective objective function. For example, suppose that the firm has three objectives: profit, service quality, and unit sales (or market share). We can express this firm's objective function as follows:

$$G = \lambda_1\Pi + \lambda_2 S + \lambda_3 Q, \tag{9}$$

with λ_1, λ_2, and $\lambda_3 > 0$. The firm's objective is now to maximize G with respect to service productivity, P. We can show (see the Appendix) that the optimal productivity level becomes

$$P^* = R/[\theta^{**}N_S + (1 - \theta^{**})N_A], \tag{10}$$

where θ^{**} is the new optimal labor usage level and

$$\theta^{**} = \theta^* + \{(\lambda_2 + \lambda_3\gamma)/[2\lambda_1\gamma(N_SW_S - N_AW_A - A)]\}, \tag{11}$$

where θ^* is the profit-maximizing level of labor usage. From Equations 9–11, we can show (see the Appendix) that more emphasis on the service quality objective results in a lower optimal productivity level, and an increased emphasis on the sales/market share objective also results in optimal productivity being lower.

Price decisions. To address this issue, we build a simplified model that is adequate to explore several key implications of including price as a decision variable. We now assume that the firm sets both the productivity level and the price to maximize profit and carry forward the other four assumptions from the original model. We set the sales function as follows:

$$Q = aS - bR^2, \tag{12}$$

where Q, S, and R are defined as previously and a, b > 0. This equation assumes that sales increase with service quality (Zeithaml 2000) and that price has a nonlinear effect on sales, as is standard in the microeconomic theory of demand (Perloff 2011). If we assume, for simplicity of exposition, zero marginal cost and cost of service proportional to service quality, this yields revenue and cost equal to

$$\text{Revenue} = QR = aRS - bR^3 \text{ and} \tag{13}$$

$$\text{Cost} = cQS, \tag{14}$$

where c > 0. For simplicity of exposition and tractability, we work with service quality S rather than productivity directly, which is sufficient for our purposes, understanding that, according to Assumption 4, higher service quality implies lower productivity (for a given level of technology). We can derive (see the Appendix) the optimal service quality level and optimal price as follows:

$$S^* = a/(8bc^2), \text{ and} \tag{15}$$

$$R^* = a/(2bc). \tag{16}$$

As in Assumption 4 in the original model, we assume that more labor (worse productivity) increases service quality. From this, we can affirm (see the Appendix) that the main findings that we can test are all consistent with those of the original model. In particular, we replicate Result 3 (higher price is associated with lower optimal productivity), Result 4 (higher wage rate is associated with higher optimal productivity), and Tentative Result 2 (in a more concentrated market, optimal productivity is higher).

Capacity constraints. It may be the case that the firm has capacity constraints that imply it may not meet demand. For example, a restaurant, an airliner, or a sports arena may have only so many seats. A doctor or business consultant may have only a limited amount of time. We extend the original model to accommodate this situation. In this case, we replace Equation 4 with a demand equation:

$$D = \gamma S + \eta Z. \tag{17}$$

Whenever the capacity constraint $Q_{max} < D$, the capacity constraint is binding. In that case, sales $Q = Q_{max}$ and Equation 4 shows that the firm wants to provide service quality

$$S_{cap} = (Q_{max} - \eta Z)/\gamma, \tag{18}$$

which implies (using Equations 1 and 3) an optimal productivity level of

$$P_{cap} = R/[\theta_{cap}N_S + (1 - \theta_{cap})N_A], \tag{19}$$

where

$$\theta_{cap} = (Q_{max} - \eta Z - \alpha)/(1 - \alpha). \tag{20}$$

We can show (see the Appendix) from Equations 17–20 that whenever a capacity constraint is binding (demand exceeds capacity), the optimal productivity level will be increased, less labor will be used, and the service quality level will be reduced, given that price is fixed. Intuitively, this occurs because the firm does not need to employ as much labor to boost sales.

We now consider the effect of a binding capacity constraint in the model from the previous section, in which we also permit price to be a decision variable. In that case, with Q_{max} as the capacity constraint, we can show (see the Appendix) that the price is unchanged from the unconstrained case, but from Equation 12, it is evident that service quality will be lower:

$$S_{cap} = (Q_{max} + bR^2)/a, \tag{21}$$

and the lower the capacity constraint, the lower is the optimal level of service quality.

However, there may be cases in which it is in the firm's interest to maintain service quality even in the presence of a capacity constraint. Consider, for example, a time period of limited duration in which a capacity constraint may emerge (e.g., area hotels during the Olympics). A high season (e.g., winter in a tropical resort location) may produce similar effects. In such a case, we consider that the firm may want to keep its service quality constant but adjust price. This produces no changes in the firm's labor usage per unit or its service productivity. We show that in such a situation the result is that, as expected, price should be increased (see the Appendix).

Empirical Analysis

Optimality Hypothesis

Figure 1 presents a broad overview of the conceptual framework underlying our empirical analysis. Inspired by the theoretical derivations of the previous section, as well as prior research and literature, we build three empirically testable hypotheses. The optimality hypothesis describes the nature of optimal productivity, and the two determinants of optimal productivity hypotheses identify the directional effect (positive or negative) of the impact of several variables on the optimal service productivity.

> H_1: For a given level of technology, (a) there is an inverted U-shaped relationship between productivity and profitability for each firm, and (b) the optimal productivity level increases as technology advances.

The optimality hypothesis differentiates between short-term productivity optimization (automation effect) and long-term productivity optimization (technology effect). It predicts that a unique optimal productivity level exists for each firm at any given technology level; as technology advances, this optimal productivity level moves up as well. Result 1 from the "Theoretical Results" section suggests that a unique optimal productivity level exists, and Tentative Result 1 states that the optimal productivity level increases over time. Intuitively, too much productivity harms sales by damaging service quality, and too little productivity makes costs too large, but the point at which this happens may change over time.

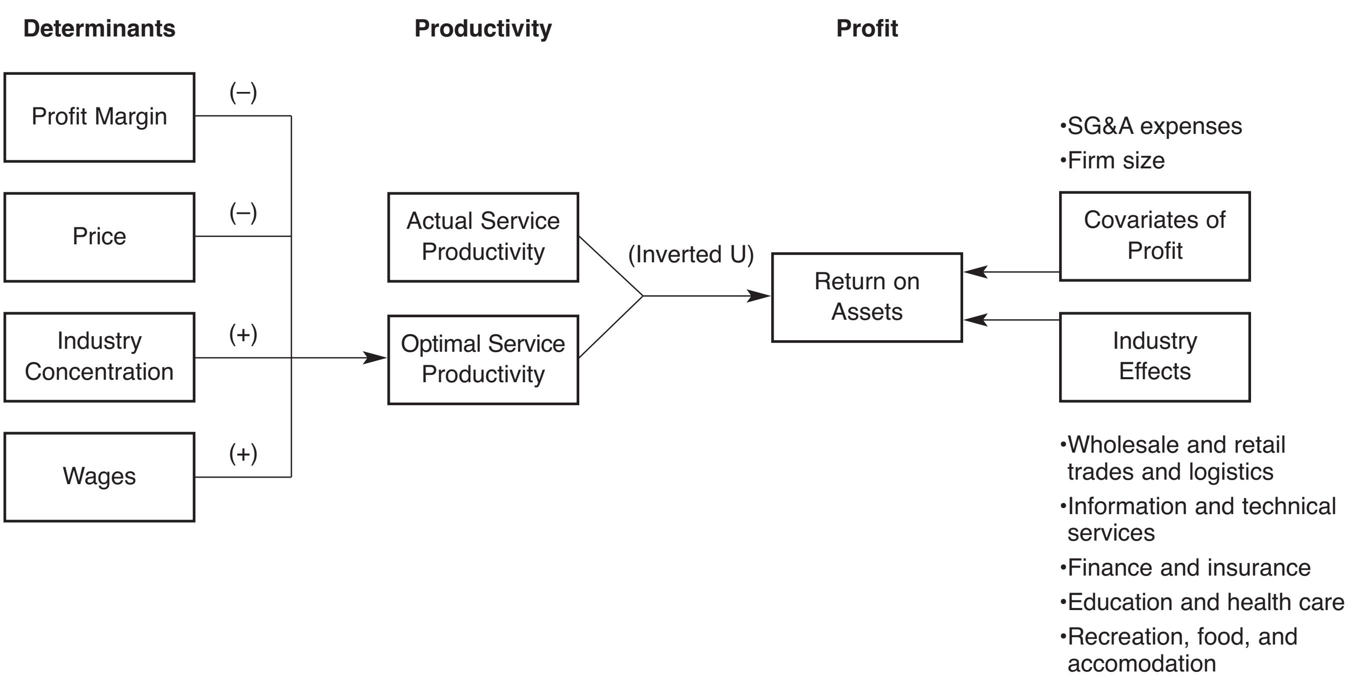
The effectiveness of automation in serving customers is limited by the current technology level. Over time, technological advances make automation more effective at providing service, leading to more use of automation (Dewan and Min 1997). Science and technology are cumulative, which means that, except for unusually backward periods (e.g., the Dark Ages), technological capability advances over time. For example, Moore's Law, which has been shown to be an excellent predictor of technological advance in the computer industry, contends that the number of transistors an integrated circuit can hold doubles every two years, a rate that has kept pace for the past 50 years (Intel Corporation 2005). Virtually without exception, the level of technology continually advances.

Determinants of Optimality Hypotheses

In the following two hypotheses, we explore managerially relevant factors that increase or decrease the optimal productivity level. Given data availability, we empirically examine profit margin and price as the two factors that decrease optimal productivity and market concentration and wages as the two factors that increase optimal productivity. Factors other than the four variables empirically examined here can be expected to have similar effects in determining optimal productivity, as long as they are the driving forces for the firm to provide better (or worse) service quality.

> H_2: Firms decrease their level of optimal productivity when factors (e.g., higher profit margin, higher price) motivate the provision of better service quality.

FIGURE 1
Conceptual Framework for the Empirical Analysis



Notes: No specific prior predictions are made for the industry effects and covariates of profit.

In this hypothesis, service quality is implicit but central in linking market drivers to level of optimal productivity. All else being equal, we expect that higher levels of variables that motivate higher service quality (e.g., profit margin, price) will lead the firm to lower its level of service productivity by using more labor.

The theoretical results we offered previously indicate that as margins rise, unit sales are worth more, and therefore a higher level of service quality is justified to increase unit sales. The higher level of service quality implies that the optimal productivity level should be lower. Similarly, higher prices mean sales become more valuable because the absolute profit margin increases, even as the percentage margin remains constant. This provides incentive for better service quality and implies a lower optimal productivity level.

This hypothesis can be viewed as a combination of Assumptions 3 (better service quality results in higher demand) and 4 (less labor intensity in service decreases service quality). Thus, the firm will use more labor to provide better service quality when factors strengthen the service quality–demand relationship, which results in lower optimal productivity. (For literature support, see the discussion of Assumptions 3 and 4.)

Specifically, high-margin and high-price services encourage better service quality. This can typically be found in the high-end service market, such as first-class flights, five-star hotels, and Michelin star restaurants, which requires the provision of superior service levels to justify the price premiums that customers must pay. The extant literature posits positive relationships between both profit margin and price and service quality (e.g., Zeithaml 2000). Thus, we expect higher margins and higher prices to result in lower productivity:

> H_3: Firms increase their level of optimal productivity when factors (e.g., higher market concentration, higher wages) discourage the provision of better service quality.

In this hypothesis, we examine two factors, market concentration and wages, that drive the firm to reduce service quality and thus to increase the level of service productivity by using more automation.

In the "Theoretical Results" section, we assume that service quality would have less impact on sales in a more concentrated market. The tentative theoretical result derived from this assumption suggests that in a more concentrated market, optimal productivity is higher. In addition to the theoretical result, the literature and common observation suggest that in a concentrated market, economies of scale would entice firms (which have higher average market share, by the definition of market concentration, and thus should be larger on average) to maintain a higher level of optimal productivity by using more automation to provide service because (1) automation can often provide a tremendous efficiency gain and (2) in such a market customers have fewer options to switch, even if the firm's efficiency gain is at the cost of service quality (Storbacka, Strandvik, and Grönroos 1994). In addition, a concentrated industry may have a well-established level of service quality, which would result in service quality being less of a factor in driving sales.

Similarly, a straightforward theoretical conclusion is that when labor is more expensive, there is more benefit (cost savings) from substituting (cheaper) automation for (more expensive) labor, even if the labor required to operate the automation has a higher wage rate than that of the service providers. The firm is motivated to use more automation for cost saving, even if it is at the cost of service quality, which results in a higher optimal productivity level.

This hypothesis can be considered a combination of Assumptions 3 (better service quality results in higher demand) and 5 (automation is more cost-effective than labor in providing service) and the additional assumption that service quality has less effect on demand if the market is more concentrated. When factors weaken the service quality–demand relationship, the firm will use automation to provide service for cost savings (for literature support, see the previous section).

In addition to the theoretical result, the literature provides more support regarding the relationship between wages and service productivity. For example, Napoleon and Gaimon (2004) use mathematical models to analyze information technology worker systems and find that firms offering mass-produced services often invest more heavily in automation, such as cash registers with automatic change makers, to reduce the wage rate of the workforce they must pay to lower costs. Filiatrault, Harvey, and Chebat (1996) find that large service firms automate tasks and hire part-time personnel more than smaller firms to enhance service productivity. Brown and Crossman (2000) state that hotels often deploy automation as a means to absorb the heightened labor costs in the face of the introduction of the national minimum wage. Dewan and Min (1997) observe that the earliest applications of information technology were directed toward the reduction of personnel costs in such labor-intensive operations as accounting, purchasing, and payroll. In summary, these studies show that there is a stronger substitution effect of automation for labor for the provision of service, and thus more motivation to increase productivity, when wages are higher.

Method

Empirical Model

We build an empirical model that can test the three hypotheses developed in the previous section. We first propose the empirical model in general terms and then show how the model is operationalized with respect to data and specific measures and how the hypotheses from the previous section may be tested. We can express the general empirical model as follows:

$$\Pi_j = \delta_0 - \delta_1(P_j - P^*_j)^2 + \Sigma_c D_c Y_{cj} + \eta_m + \varepsilon_j, \quad (22)$$

where Π_j is the profit for firm j; P_j is firm j's labor productivity; P^*_j is the optimal level of productivity for firm j; Y_{cj} is covariate c of profit for firm j; δ_0, δ_1, and the D_cs are model parameters to be estimated; η_m is a fixed effect for firm j's industry, industry m; and ε_j is the error term, assumed to be independently and normally distributed. The quadratic relationship of optimal productivity to profit is motivated by the theory we presented in the previous section. We chose the covariates from variables typically used to model profitability (discussed in more detail subsequently). We include the industry fixed effect to control for unobserved heterogeneity due to industry.

This formulation can directly test the automation effect of the optimality hypothesis (H_{1a}). It is general enough to include the case in which there is an optimal value of productivity ($\delta_1 > 0$), supporting the automation effect, as well as the cases in which productivity should instead be maximized ($\delta_1 < 0$) or makes no difference ($\delta_1 = 0$) (either way rejecting H_{1a}). This formulation also facilitates testing the technology effect of H_1, namely, the shift in optimal productivity over time (H_{1b}). If the model is estimated at two different time periods and if H_{1b} is correct (that optimal productivity increases as technology advances), we expect the average P^*_j across firms j to increase from the first period to the second, given that the level of technology only increases over time.

The determinants of the optimal productivity hypotheses (H_2 and H_3) involve predictions about how various variables (e.g., profit margin, price) affect the predicted optimal productivity level. This is captured in the empirical model by expressing the optimal productivity level, P^*_j, as a function of the levels of these determinants, X_{ij}. If X_{ij} is determinant i for firm j, the expression for P^*_j is

$$P^*_j = \beta_0 + \Sigma_i \beta_i X_{ij}, \tag{23}$$

where the βs are coefficients to be estimated.

When using Equation 23 to test H_1, we assume that Equation 23 accurately reflects true optimal productivity; that is, all variables that will affect the optimal productivity are included. Although the predictors in Equation 23 are consistent with theory, the current data lack sufficient information to completely validate the equation empirically. Subsequently, we show that the test for H_1 is robust to any systematic error in this estimation.

Combining Equations 22 and 23, we obtain a nonlinear equation that enables us to estimate all the model parameters at once, using nonlinear estimation methods:

$$\Pi_j = \delta_0 - \delta_1[P_j - (\beta_0 + \Sigma_i \beta_i X_{ij})]^2 + \Sigma_c D_c Y_{cj} + \eta_m + \varepsilon_j. \tag{24}$$

Straightforward hypothesis tests on the β coefficients directly test the determinants of optimal productivity hypotheses. Given data availability, we examine the following four determinants for optimal productivity:

$$\begin{aligned} X_{1j} &= \text{profit margin,} \\ X_{2j} &= \text{price,} \\ X_{3j} &= \text{market concentration, and} \\ X_{4j} &= \text{wage rate.} \end{aligned} \tag{25}$$

We use a nonlinear least squares estimation with the Gauss–Newton iterative method to estimate the model parameters, which regresses the residuals onto the partial derivatives of the model with respect to the parameters until the estimates converge.

Data

We tested the hypotheses empirically using data from COMPUSTAT North America that include more than 30,000 active and inactive publicly held U.S. and Canadian firms. We focus our analysis on service firms with North American Industry Classification System (NAICS) codes of 42–92 for 2002 and 2007, so we can investigate both the robustness of the results over time and the change over time in optimal productivity. The U.S. Census Bureau adopted NAICS in April 1997 to replace the Standard Industrial Classification system because it better reflects the current economy structure by including new service sectors such as information and professional, scientific, and technical services. Our universe of service firms is comprehensive, in that it includes all service sectors: wholesale and retail trade, transportation and warehousing, information, finance and insurance, real estate and rental and leasing, management of companies and enterprises, administrative and support services, educational services, health care, entertainment and recreation, accommodation and food services, and other services. After removing firms with missing values, we had 741 firms in 2002 and 751 firms in 2007.

Many firms in the COMPUSTAT data set do not report wage. We used two tests to determine whether there is a sample selection bias due to this nonreporting practice. First, we performed multivariate t-tests to check whether there were significant differences between wage reporting and nonreporting firms with respect to dependent variables and determinants in our optimality service productivity equation (i.e., return on assets [ROA], service productivity, profit margin, Herfindahl–Hirschmann index [HHI], number of employees, and selling, general, and administrative expenses [SG&A] expenses). The results indicate that, in general, wage-reporting firms are not significantly different from nonreporting firms for most of the variables in the equation with one exception. For 2002, reporting firms tended to have a higher number of employees.

Second, we carried out Heckman two-step models to estimate a firm's ROA, conditioned on whether the firm reports wage. We estimated the probability of reporting wage as a function of number of employees, SG&A expenses, and a set of industrial sector dummy and the ROA model as a function of profit margin and HHI. The insignificant correlation estimate ($\rho = -.197$, $p = .103$ for 2002; $\rho = -.114$, $p = .459$ for 2007) indicated that selection bias is not a significant problem in the estimation of the ROA equation. Table 2 summarizes the firm characteristics for the two data years.

Measures

Our empirical model (Equation 24) predicts profit as the inverted U-shaped relationship between productivity and optimal productivity, covariates of profit, and industry fixed effects to capture unobserved heterogeneity. Optimal productivity is represented in the empirical model by a linear combination of its determinants. The determinants of opti-

TABLE 2
Summary of Firm Characteristics

	2002		2007	
	Number of Firms	Percentage	Number of Firms	Percentage
Industry				
Wholesale and retail trades and logistics	46	6.21%	47	6.26%
Information and technical services	98	13.23%	67	8.92%
Finance and insurance	530	71.52%	574	76.43%
Education and health care	12	1.62%	12	1.60%
Recreation, food, and accommodation	55	7.42%	51	6.79%
Sales ($ Millions)				
<10	45	6.07%	22	2.93%
10–100	326	43.99%	316	42.08%
100–500	189	25.51%	196	26.10%
500–1000	68	9.18%	61	8.12%
>1000	113	15.25%	156	20.77%
Number of Employees				
<50	35	4.72%	27	3.60%
50–100	59	7.96%	52	6.92%
101–400	227	30.63%	270	35.95%
>400	420	56.68%	402	53.53%

mal service productivity were inspired by both the theoretical model and the existing literature and include profit margin, price, market concentration, and wage rate. In the following subsections, we discuss the specific operationalization of the variables in the empirical model.

Profit. We operationalize profit as ROA, one of the most frequently used indicators for assessing firm financial performance in the marketing literature (e.g., Aksoy et al. 2008; Anderson, Fornell, and Mazvancheryl 2004; Fornell et al. 2006; Narver and Slater 1990; Noble, Sinha, and Kumar 2002) and an effective way of normalizing profits across firms of different size.[10] We calculate ROA as net income divided by total assets. The popularity of using ROA to gauge profit is because it is relatively more stable than other return indexes (e.g., return on equity), can be calculated for companies with negative shareholder's equity, is useful for analyzing competing companies in the same industry (Anderson, Fornell, and Mazvancheryl 2004), and is useful for gauging the profit of a company on an absolute basis. High-ROA firms are more profitable than low-ROA firms.

Service (labor) productivity. Following prior research (e.g., Basker 2007; Bertschek 2004; Converse 1939; Datta, Guthrie, and Wright 2005; Guthrie 2001; Huselid 1995), we calculate labor productivity as the log of sales per employee where dollar sales captures total output and number of employees measures labor input.[11] Researchers consider the sales per employee metric a good measure of labor productivity, with its greatest use being the comparison of industry competitors and the historical performance of the firm.

Determinants of optimal productivity. We include four variables as the determinants of optimal productivity. We calculate profit margin as the proportion of a firm's net sales to its gross sales, which indicates how much of a firm's sales dollars are profit. Although the firm may have a variety of services, this formulation for profit margin is a weighted average of the profit margins of those services, weighted by sales. We calculate price as the ratio of selling costs to one minus the proportion profit margin.[12] It is the total price summed over all services provided by a firm, not the unit or average price of services. The wide variations in a firm's service offerings would make the unit price, even if it were available, not easily comparable across firms and industries. Market concentration index (HHI) is the sum of the square of market shares (Schmalensee 1977) at the four-digit NAICS level. We define wage rate as the log of labor expenses per employee.[13]

Covariates of profit. We include two variables as the covariates of ROA: firm size and SG&A. Firm size is calculated as the log of a firm's number of employees (e.g., Huselid 1995; Koch and McGrath 1996) with the expectation that the larger the firm, the better the firm can combat competition, regardless of service quality, and the higher the ROA will be. SG&A is a standard reported item in a firm's financial statement that includes all salaries, indirect production, marketing, and general corporate expenses. By normalizing as a percentage of SG&A per dollar of sales,

[10]To test the robustness of our results to choice of profitability measure, we also tested return on sales as an alternative dependent variable. The results were similar, with the exception that we had to delete one of the predictor variables because of collinearity issues.

[11]The log of sales per employee is monotonically related to sales per employee (the other typical operationalization), which means that the optimal value with respect to either definition will result in the same sales per employee as optimum.

[12]As with many of the variables in the empirical analysis, the variable is standardized within industry.

[13]For consistency consideration, all measures that involve the number of employees are log-transformed (e.g., Conti 2005; Haltiwanger, Lane, and Spletzer 2007).

we expect a negative relationship between SG&A and ROA because higher costs directly reduce a firm's profit.

Industry effects. We include a set of industry fixed effects in the estimation to explicitly model industry heterogeneity. Industries are categorized according to the broad one-digit NAICS categories, with the model identified by setting the finance and insurance sector as the reference sector. We estimate the empirical model for each of two data years (2002 and 2007). Table 3 summarizes the descriptive statistics, correlations, and variance inflation factors (VIFs) for all measures for the two data years.

Estimation

We tested the hypotheses using the nonlinear regression equation as specified in Equation 24. We first explored possible multicollinearity using ordinary least squares regression analyses with ROA as the dependent variable and with all ROA covariates and optimal productivity determinants in the equation as the predictors to calculate VIFs. The VIF statistics (see Table 3) are reasonable, with a majority of them below 3.00, indicating that multicollinearity is not a concern. We then adjusted all variables to their industry means by dividing their mean-centering scores by their respective two-digit NAICS industry average.[14] This approach is similar to that in Rao, Agarwal, and Dahlhoff (2004). We Winsorized the 1% and 99% outliers of each variable to reduce the impact of extreme values. We standardized all variables in the equations to have a mean of 0 and a standard deviation of 1 to ensure direct comparability.

Results

All the tested hypotheses received support for both the 2002 and 2007 data years. The results from the empirical analysis are shown in Table 4, with findings summarized in Table 5.

Hypothesis Testing

Optimality hypothesis (H_1). We test the automation effect (i.e., for a given level of technology, there is an optimal productivity level for each firm; H_{1a}) using the δ_1 parameter. A positive δ_1 parameter indicates that there is an optimal value of productivity ($\delta_1 > 0$), a negative value indicates that productivity should be maximized ($\delta_1 < 0$), and a value of 0 indicates that it makes no difference between automation and labor for service productivity ($\delta_1 = 0$). As we predicted, the δ_1 parameter is positive and significant ($p < .001$) for both data years, suggesting an inverted U-shaped relationship between productivity and profitability and implying that there is a unique level of optimal productivity associated with a given technology, in support of H_{1a}. If the δ_1 parameter had been negative, that would have instead provided support for the idea that increasing productivity always produces higher profitability. Although the true optimal productivity level for each firm remains unknown, because our model does not (and cannot) include all explanatory variables, the estimated optimal productivity for each firm may be considered a reasonable proxy for the true value.[15]

We can test the technology effect (i.e., optimal productive level increases as technology advances; H_{1b}) (in proxy) using a t-test on the estimated mean optimal productivity levels between the two data years. We found that the average estimated optimal productivity level in 2007 (M = .514, SD = .771) was significantly higher than the average estimated optimal productivity in 2002 (M = –.153, SD = .978; t = 14.63, $p < .001$). This corresponds to a mean log optimal productivity level (not normalized and not standardized) of 6.004 in 2007 and 5.008 in 2002, in support of H_{1b}.[16, 17]

Together, the results of H_1 substantiate the necessity of separating the short-term automation effect from the long-term technology effect, as they operate differently on service productivity. Automating too much hurts service quality, and the extent to which automation can substitute labor for service effectiveness depends on the given level of technology.

Negative determinants hypothesis (H_2). H_2 predicts that there is a negative relationship between margin (price) and optimal productivity. This argument received consistent support for the two data years. As we predicted, as margin increases, optimal productivity decreases (for 2002, β = –.306, t = –3.517, $p < .001$; for 2007, β = –.205, t = –3.203, $p < .001$), and as a service firm's selling price increases, optimal productivity decreases (for 2002, β = –.217, t = –3.678, $p < .001$; for 2007, β = –.296, t = –4.698, $p < .001$).

Positive determinants hypothesis (H_3). H_3 predicts a positive relationship between market concentration (wage rate) and optimal productivity. We received strong and consistent support for the prediction. As we predicted, both the market concentration index (HHI) (for 2002, β = .228, t = 3.563, $p < .001$; for 2007, β = .313, t = 6.956, $p < .001$) and wage rate (for 2002, β = .928, t = 13.070, $p < .001$; for 2002, β = .713, t = 12.964, $p < .001$) were significant, positive determinants of optimal productivity.

Additional Analyses

Average firm productivity levels. Using the estimation results, we calculated the ratio of a firm's actual service productivity to its estimated optimal service productivity in the two data years to give a preliminary indication of whether firms' actual productivity levels tend to be too high or too low, compared with the estimated optimal level, in

[14]The HHI did not need to be normalized by industry because the predictor was calculated using industry-specific sum and sum of squares.

[15]We test the sensitivity of the findings to perturbations in the estimated optimal productivity by decreasing or increasing by 1%, 5%, and 10% the optimal productivity estimates to determine the robustness of the conclusions to different levels of error. The results show the changes in δ_1 are only in the third decimal place, and the t-values are essentially unchanged, indicating that the test for H_1 is robust to error in estimating optimal productivity.

[16]These values are interpretable as the average P* in Equation 22.

[17]To test the robustness of the time trend in optimal productivity, we also calculated the mean estimated optimal productivity for 2005 (.272) and 2009 (1.136). Together with the results from 2002 and 2007, the four years have significantly different means (F = 504.67, $p < .001$), the time ordering of the productivity levels is as expected, and the trend is in the hypothesized direction.

TABLE 3
Descriptive Statistics of Variables

Variables	M	SD	VIF	1	2	3	4	5	6	7	8
2002 (N = 741)											
1. ROA	–.03	.24		1.00							
2. Log labor productivity	5.15	.90	4.18	.06	1.00						
3. Profit margin	.52	.48	2.71	.42**	.22**	1.00					
4. Price	1470.67	5212.81	1.65	.02	.14**	.02	1.00				
5. Market concentration	.02	.02	1.21	–.20**	–.33**	–.19**	.00	1.00			
6. Log wage rate	3.60	.97	3.65	–.09*	.81**	.03	.12**	–.23**	1.00		
7. SG&A expense	.29	.42	3.11	–.55**	–.14**	–.76**	–.02	.02	.13**	1.00	
8. Log firm size	–.19	2.03	1.82	.17**	–.27**	.02	.54**	.15**	–.24**	–.14**	1.00
2007 (N = 751)											
1. ROA	–.00	.21		1.00							
2. Log labor productivity	5.52	.92	3.22	.11**	1.00						
3. Profit margin	.48	.23	1.10	.41**	.06	1.00					
4. Price	3579.58	13,300.45	1.74	.03	.25**	–.03	1.00				
5. Market concentration	.03	.03	1.25	–.11**	–.28**	–.30**	.06	1.00			
6. Log wage rate	3.92	.83	2.97	–.07*	.79**	.04	.21**	–.27**	1.00		
7. SG&A expense	.26	.26	1.24	–.72**	–.14**	.03	–.08*	–.01	.08*	1.00	
8. Log firm size	–.17	2.14	1.88	.20**	–.17**	–.06	.54**	.25**	–.19**	–.25**	1.00

*$p < .05$.
**$p < .01$.
Notes: All variables are firm-level variables, except for HHI, which measures market concentration in a firm’s industry. We obtained VIF using ordinary least squares regression with ROA as the dependent variable. Log firm size is the log transformation of employees (1000s). Price is the total price summed over all services provided by a firm, not the unit or average price of services. We normalized SG&A expense as a percentage of sales.

TABLE 4
Results of Model Estimation

Parameter	2002 (N = 741) Estimate (SE)	2002 (N = 741) t-Value	2007 (N = 751) Estimate (SE)	2007 (N = 751) t-Value
δ_0	.237 (.055)	4.309**	.227 (.042)	5.405**
Optimal Productivity				
δ_1	.195 (.027)	7.222**	.177 (.023)	7.696**
Determinants of Optimal Productivity				
β_0: Intercept	–.153 (.190)	–.805	.514 (.149)	3.450**
β_1: Profit margin	–.306 (.087)	–3.517**	–.205 (.064)	–3.203**
β_2: Price	–.217 (.059)	–3.678**	–.296 (.063)	–4.698**
β_4: Market concentration	.228 (.064)	3.563**	.313 (.045)	6.956**
β_5: Wage rate	.928 (.071)	13.070**	.713 (.055)	12.964**
Covariates of Profit				
D_1: SG&A	–.287 (.049)	–17.755**	–.457 (.034)	–13.441**
D_2: Firm size	–.166 (.034)	–4.882**	–.131 (.031)	–4.226**
Industry Effects				
η_1: Wholesale and retail trade and logistics	–.091 (.038)	–2.395*	.108 (.029)	3.724**
η_2: Information and technical services	–.314 (.059)	–5.322**	–.106 (.035)	–3.029**
η_3: Education and health care	–.083 (.035)	2.371*	.006 (.029)	.207
η_4: Recreation, food, and accommodation	.092 (.058)	1.586	.288 (.052)	5.538**

*$p < .05$.
**$p < .001$.
Notes: Dependent variable is profit, measured as ROA.

TABLE 5
Summary of Findings

Topics	Operational Definitions	Theoretical Results (Derived Relationship)	Empirical Hypotheses (Predicted Relationship)	Hypothesis Testing Results 2002	Hypothesis Testing Results 2007
Optimal productivity level	Relationship between productivity and profit	R_1: Optimal level	H_{1a}: Inverted U (optimum implied)	Supported	Supported
Optimal productivity level over time	The estimated optimal productivity difference between 2002 and 2007	TR_1: Positive	H_{1b}: Positive	Supported	Supported
Profit margin and productivity	Net sales ($MM)/gross sales ($MM)	R_2: Negative	H_2: Negative	Supported	Supported
Price and productivity	Selling costs ($MM)/(1 – profit margin)	R_3: Negative	H_2: Negative	Supported	Supported
Market concentration and productivity	HHI = the sum of the squared market shares of the firms in the industry	TR_2: Positive	H_3: Positive	Supported	Supported
Wage rate and productivity	Wages ($MM)/per 1000 employees	R_4: Positive	H_3: Positive	Supported	Supported
Other factors and productivity	Sales are driven by factors other than service quality	R_5: Positive	N.A.	N.A.	N.A.

Notes: R denotes results from the analytical theory, TR denotes tentative results from additional assumptions, and H denotes predictions drawn from the empirical hypotheses.

the two data years. A ratio greater than 1 may be interpreted as too productive, whereas a ratio less than 1 is not productive enough. A one-sample t-test showed that in 2002, firms were perhaps slightly overproductive (M = 1.028, SD = .109), which was significantly higher than 1.0 (t = 6.89, $p < .001$). In 2007, firms (with the exception of large firms, as we discuss subsequently) may have been underproductive (M = .938, SD = .126), which was significantly lower than 1.0 (t = –13.49, $p < .001$). This result suggests that in 2002, when the economy was still strong, firms may have been belt tightening slightly too much on average, but in 2007, when the economy was teetering on the brink of recession, firms (with the exception of large ones) may not have been belt tightening enough.

Large firm productivity. A further inspection of the distribution pattern of the productivity ratio points to to the preliminary conclusion that large firms (in terms of annual sales) may have been too productive, when comparing their actual productivity with their estimated optimal productivity in both data years. For 2002, the average productivity (productivity divided by estimated optimal productivity) of large firms (M = 1.088, SD = .143) was, on average, 8.8% too high, significantly higher than the estimated optimum (t = 5.30, p < .001), and significantly higher than that of smaller firms (M = 1.021, SD = .103; comparison of means test $1.088_{\text{large firms}}$ vs. $1.021_{\text{small firms}}$; t = 5.12, p < .001). In 2007, the average productivity of large firms (M = 1.068, SD = .164) was 6.8% too high (relative to the estimated optimum), while small firms (M = .924, SD = .111) were, on average, only 92% as productive as they should have been (compared with the estimated optimum). Large firms were, on average, significantly too productive (t = 3.62, p < .001) and significantly more productive than small firms ($1.068_{\text{large firms}}$ vs. $.924_{\text{small firms}}$; t = 10.15, p < .001).

Discussion

Implications from the Major Findings

We built testable hypotheses, motivated by the theoretical model and prior literature, and tested them using data from more than 700 firms in two data years, five years apart. The empirical results are largely convergent with those of the formal theory. Combined, the theory and empirical analysis yield important findings about the nature of service productivity. We recognize that the managerial implications derived from these important findings might be sensitive to some of the assumptions on which we build our analytical theory and empirical analysis. Our theory assumes that several important variables (e.g., wage rates) are fixed in the market and are known to the firm, that better service quality increases demand, and that greater increase in labor increases service quality. In addition, our empirical analysis assumes a parsimonious set of variables that determine optimal service productivity, as well as a quadratic relationship between productivity and profit.[18]

Optimal productivity level. We provide a theoretical approach that provides new insight into how a firm should manage its level of service productivity as a strategic decision variable. This approach draws from both the economic view of the firm and behavioral theories of the firm in that it recognizes that a firm should manage its level of service productivity in pursuing profit maximization, among other strategic goals. This involves a trade-off between using labor to provide service, which increases service quality (effectiveness), and using automation, which increases productivity (efficiency). By distinguishing between a technology effect (increasing effectiveness of automation and decreasing automation costs over time) and a labor–automation substitution effect, we provide new insight into how service productivity works. Our theoretical work also advances theory on the relationship between service quality and productivity. We show how the service productivity level and the labor–automation trade-off relate to service quality, sales, and profitability, indicating how a firm might use service productivity as a strategic decision variable.

In the empirical analysis, we show that for a given level of technology, there is an inverted U-shaped relationship between productivity and profitability, suggesting the existence of an optimal level of productivity for each firm. This functional form suggests that when a firm achieves a level of productivity equal to its optimal level, the firm's profit is maximized, all other factors being equal. When service productivity is either too low or too high, profitability suffers. As a result, the firm should seek an appropriate level of service productivity, using the insights obtained from the set of determinants of optimal service productivity.

The empirical analysis further demonstrates that the optimal service productivity level increases over time as technology level advances. The optimal service productivity level is a moving target. Firms need to reassess their service productivity decision making over time as technology advances. As the level of technology increases (automation becomes a more effective substitute for labor), optimal service productivity increases, and the firm should use less labor and more automation.

Determinants of optimal productivity. Our work proposes a theoretical structure that links variables characterizing a firm and its industry to its strategy with respect to service productivity. This provides a deeper understanding of the nature of the factors that affect optimal service productivity. In brief, our theoretical results suggest that higher profit margin and higher price decrease optimal productivity, whereas a higher wage rate, factors other than service quality that drive sales, and binding capacity constraints increase optimal productivity. Motivated by theory and prior literature, we develop hypotheses and empirically test a set of variables that help the firm determine when its service productivity level should be higher or lower. As predicted, we find that higher profit margin and higher price have a negative impact on optimal productivity level, whereas higher market concentration and higher wage rate have a positive impact. The data in our empirical analysis are all in the public domain, which means that any firm can replicate our analysis on the most recently available data. By inserting the firm's values of the determinants of optimal productivity into our empirical equation, the firm can get a preliminary idea about whether it is under- or overproductive. This can help give the firm guidance as to how to make appropriate decisions with respect to the firm's level of service productivity. We can provide additional insight by examining the "perfect storm" examples of when a firm should adopt either a high or low service productivity strategy:

- *The perfect high-service-productivity situation*. The firm is positioned as a low-profit-margin, low-price player in a market with few competitors, and the industry's wage rates are high. Sales are heavily influenced by factors other than service quality. This firm should seek to automate heavily and find ways to replace its labor with machines. To visualize this, imagine a fast-food restaurant in an exclusive and expensive city such as Santa Barbara, Calif., where zoning laws restrict the number of restaurants that can exist. The

[18]The functional form does not presume, however, that the quadratic relationship implies a maximum. That depends on the parameter value for δ_1.

firm is low profit margin and low price compared with its restaurant competitors, and wages are high because of the expensive location. This restaurant should implement as much automation as possible.

- *The perfect low-service-productivity situation*. The firm is positioned as a high-profit-margin, high-price player in a very competitive industry, but wages are relatively low. Service quality is a strong driver of sales. This firm should seek to provide "high-touch," labor-intensive service and automate only when it can be done in such a way that service quality is enhanced. To visualize this, imagine a gourmet French restaurant in a highly competitive large city with a low wage structure, such as Shanghai, China. Such a restaurant should be labor intensive to the extreme, with abundant service personnel attending to the customer's every need.

Of course, typical examples will not be so clear-cut and will instead involve shades of gray. In particular, the variables noted in the perfect storm examples do not always vary together in that way. Illustrated with the four determinants, we establish a general pattern that, as long as factors motivate firms to increase service quality, service productivity should be lower; in contrast, as long as factors discourage firms from increasing service quality, service productivity should be higher. This has important managerial implications in that every service firm can find its own set of positive and negative driving forces to help guide whether productivity should be higher or lower.

It may also be useful for a firm to consider the competitive positioning implications of the trade-off between productivity (efficiency) and service quality (effectiveness). For example, if a smaller firm is challenging a larger one in competition for essentially the same market segment, is the optimal productivity level the same? How should managers of these firms adapt and calibrate our results to determine their optimal level? Anderson, Fornell, and Lehmann (1994) provide an insightful discussion regarding how small firms can provide excellent service quality to their niche market to challenge a larger firm. Large firms are more likely to have more heterogeneous customers, making it more costly for the firm to maintain a high level of service quality for all customers. Our findings provide additional insights into how smaller firms can leverage level of optimal productivity for competition. We expect that (1) smaller firms should keep their level of productivity relatively lower than that of larger firms so that their customers can be better served and that (2) larger firms should be cautious about increasing their service productivity too much due to the lure of economies of scale or customer lock-in.

Implications from the Tentative Findings

Findings from our theoretical extensions and additional empirical analyses provide several important implications for a service firm to manage its service productivity. These include managing the level of service productivity for achieving multiple firm objectives, managing service productivity in the presence of capacity constraints, considering how actual productivity relates to estimated optimal productivity, and examining productivity levels for large firms.

First, in the theoretical extensions, we explore how the level of optimal productivity is affected by considering the case in which the firm has other objectives in addition to profit. Specifically, we examine the case in which the firm values service quality and sales/market share. We show that the optimal productivity level becomes lower when the firm chooses to build competitive strength to profit in the long run, rather than seeking only to maximize profit in the short run. In summary, our theoretical results give managers insights into choosing productivity levels not only to maximize profit but also to achieve differentiation (i.e., compete on service quality) and grow sales and market share.

Second, we explore the situation in which price is permitted to vary and find that most of the results from the original theory are replicated. In particular, optimal productivity increases with higher wages or greater influence by factors other than service quality in driving sales. We also show that higher price should be associated with lower optimal productivity. These results show that including price as a decision variable leads to results largely consistent with the original theory.

We then extend the original theory to incorporate capacity constraints. We show that whenever demand exceeds capacity, the optimal productivity level will be increased, less labor will be used, and the service quality level will be reduced. This is because the firm does not need to work as hard to produce the optimal sales level, which is reduced from the capacity-unconstrained case; thus, the firm uses more automation to utilize the capacity more efficiently (but less effectively). We also consider a capacity-constrained model, in which price is permitted to vary, and find that the results are largely consistent with those of the price model: The optimal price is the same, and again, the optimal service quality level is lower. If service quality is instead held constant, a capacity constraint, if binding, results in a higher optimal price.

Third, the additional empirical analysis on average firm productivity levels shows that firms may have been slightly overproductive in 2002 but underproductive in 2007. This may indicate the relevance of the macroeconomic environment in service productivity decisions. In 2002, the economy was strong, while in 2007 the economy was teetering on the brink of recession. We speculate three possibilities for this productivity. First, firms may simply be slow to adapt to changing economic situations. Thus, we observe that in 2002, firms may have failed to take full advantage of the favorable environment by being too slow to employ more labor to drive sales. In 2007, in contrast, firms may have failed to adapt to the declining economic environment by being too slow to reduce labor and boost productivity to reduce costs. Second, when the economy is strong, firms may be enticed to be too productive because higher sales levels lead to greater economies of scale for automation, motivating greater adoption of automation. Third, in strong economic times, the job market drives wages higher, motivating less use of labor, whereas in weak economic times, labor is cheaper, and there is stronger motivation to employ labor. Investigation of these possibilities provides a rich opportunity for further research.

Fourth, the preliminary empirical analysis on large firms' productivity reveals a systematic tendency for large firms to be too productive. If confirmed, this result may be due to large firms' tendency to take advantage of economies of scale in automation. If our empirical findings hold, this suggests that large firms should focus more on providing good service, perhaps automating less, and placing less emphasis on productivity for cost reduction. Such firms might instead benefit from higher labor intensity and more concern for service quality.

Taken as a group, these results can help managers make more informed and strategic decisions about the level of service productivity to maintain and how to make the most profitable labor–automation trade-off in service provision. Although consistent and predictable trends in level of technology and automation costs confirm the wisdom of increasing automation over time, it is possible to automate too much, too fast. For any given level of technology, firms should find an optimal level of service productivity—one that will maximize its profit.

Limitations

As with any study, there are limitations to be kept in mind. Our theory, though formal and rigorous, like any theory, is merely suggestive of reality and is useful only to the extent that it provides insights. We could build alternative theories based on somewhat different assumptions. We could also complicate the theory by such things as (1) considering the decisions of competitors (perhaps asymmetric), (2) allowing nonlinear relationships where there are linear ones, or (3) considering the timing aspects of investment in automation. All these complications would make the theory more realistic but also less tractable (perhaps intractable) and with perhaps limited ability to produce additional insight about optimal service productivity.

The limitations of our empirical analysis should also be considered. Although we analyze a very large sample of service firms (more than 700 firms at each of two points in time), the sample has its limitations. It is limited to one continent (North America), and it is possible that service productivity operates differently in other economic environments. The two samples (2002 and 2007) do not include the same firms, so the nature of the samples could be different across the two time periods. (The alternative, including only those firms that are in both samples, might result in a sample selection bias that could miss newer or less successful firms.) In addition, although it would create severe difficulties for data collection, a more extensive time range would be desirable.

Directions for Further Research

As our theory and empirical analysis suggest, the optimal service productivity level will increase over time, so it is essential to learn more about how to increase service productivity in a way that increases a firm's profit. More research is needed regarding the effect of automation efforts on both current and future revenues and costs. That is, the effect of automation on perceived service quality is of vital importance. For example, what kinds of service technology build or maintain customer satisfaction and customer retention? Which kinds result in damage to those future-oriented measures? How does customer lifetime value relate to receptiveness to automation? Are there ways to usefully segment the market to provide the right level of automation to different segments? How can incentive plans for the executives of large firms be designed to curb the short-term viewpoint that leads to excessive cost cutting and inadequate attention to service quality?

It would be useful to have longitudinal studies of firms that have sought to increase service productivity. How quickly can service productivity be increased without damaging customer satisfaction and long-term profits? What kinds of service productivity efforts have the greatest impact, and which ones are implemented the fastest? What are the most significant problems firms have when trying to increase service productivity?

Another important topic for further research is decisions of how and when to invest in automation. An important issue is how much investment a firm should make to increase the level of automation and how those decisions should be timed. Although these topics are beyond the scope of the current study, they are highly relevant to the service productivity issue.

There may be many additional variables that have an impact on the optimal productivity level, such as firm size, structure of the firm, nature of the service provided, and characteristics of the customer base. Further theoretical and empirical analysis is needed to fully explore all of these factors. It may also be the case that it is more difficult to provide high service quality in volume. That is, the same labor intensity may provide worse service quality when the sales volume is high (Horstmann and Moorthy 2003). Researchers may want to explore this phenomenon and its managerial implications more deeply.

Another direction for further research is how to incorporate future expectations into the theoretical analysis. Currently, our analysis considers only the current period, but a forward-looking analysis might provide different results. For example, a company may make decisions about productivity on the basis of expectations about the future labor market, the nature of future demand, or the pace of technological advance. The time required for customers to learn and adapt to new technologies may also have an impact on productivity decisions.

Conclusions

We propose that service productivity should be managed as a strategic decision variable to be optimized. We present a formal theory of when service productivity should be higher or lower and also conduct an empirical investigation using data from more than 700 service firms for each of two data years. Our theory and empirical analysis produce several important conclusions about how service productivity should be managed. The most important conclusions are that for a given level of technology, there is an optimal level of service productivity and that this level is affected by a set of managerially relevant variables. Either too low or too high a level of productivity damages the firm's profit. Our study also provides specific guidance to managers as to when service productivity should be higher or lower and provides cautionary advice to large firms. These findings have important implications for how marketers should provide service to their customers—how the firm should trade off labor and automation to provide a more profitable level of service productivity.

Appendix
Proofs and Derivations

Optimal Productivity

Equation 2 indicates that there is a one-to-one correspondence between θ and service productivity, and because we find the algebra simpler if we optimize with respect to θ, we proceed along those lines. From Equations 4, 5, and 6, we have

$$\text{(A1)}\quad \Pi = [mR - \theta N_S W_S - (1-\theta)N_A W_A - (1-\theta)A][\gamma\alpha + \gamma(1-\alpha)\theta + \eta Z].$$

Taking $\partial\Pi/\partial\theta$ yields the following:

$$\text{(A2)}\quad \partial\Pi/\partial\theta = \gamma(1-\alpha)(mR - N_A W_A - A) - (N_S W_S - N_A W_A - A)(\gamma\alpha + \eta Z) + \theta[-2\gamma(1-\alpha)(N_S W_S - N_A W_A - A)].$$

For the (typical) case in which the firm uses both automation and labor ($0 \le \theta^* \le 1$), we observe the result in Equation 8. We now check the second order conditions:

$$\text{(A3)}\quad \partial^2\Pi/\partial\theta^2 = -2\gamma(1-\alpha)(N_S W_S - N_A W_A - A)].$$

From Assumption 5, we have $(N_A W_A + A) < N_S W_S$, and from Assumption 4, we have $\alpha < 1$, from which we are assured that the expression in Equation A3 is negative, and θ^* does indeed maximize profit. This shows that P^* from Equation 7 is the optimal productivity level. We also note from Equation 5 that

$$\text{(A4)}\quad \partial P^*/\partial\theta^* = [-R(N_S - N_A)]/\{[\theta^* N_S + (1-\theta^*)N_A]^2\}.$$

With the assumption that automation uses less labor than a total labor approach to service ($N_S > N_A$), which results trivially from the definition of automation, we can see that $\partial P^*/\partial\theta^* < 0$.

Result 1: For a given level of technology there is an optimal productivity level.

This is seen from the first-order and second-order conditions for optimality of θ^*, combined with Equation 7, which shows the optimal productivity level P^* has a one-to-one correspondence with θ^*.

Result 2: As profit margin increases, optimal productivity decreases.

$$\text{(A5)}\quad \partial\theta^*/\partial m = R/[2(N_S W_S - N_A W_A - A)].$$

From Assumption 5, we know that $(N_A W_A + A) < N_S W_S$, from which we concluded that $\partial\theta^*/\partial m > 0$. We then obtain the following:

$$\text{(A6)}\quad \partial P^*/\partial m = (\partial P^*/\partial\theta^*)(\partial\theta^*/\partial m) < 0.$$

Result 3: As price increases, optimal productivity decreases.

$$\text{(A7)}\quad \partial\theta^*/\partial R = m/[2(N_S W_S - N_A W_A - A)].$$

Again using Assumption 5, we get $\partial\theta^*/\partial R > 0$, and thus

$$\text{(A8)}\quad \partial P^*/\partial R = (\partial P^*/\partial\theta^*)(\partial\theta^*/\partial R) < 0.$$

Result 4: As wages rise, optimal productivity increases.

Let us assume an increase in wages (W_S and W_A) by a factor of k. Then the optimal labor usage is as follows:

$$\text{(A9)}\quad \theta^* = \{(mR - kN_A W_A - A)/[2(kN_S W_S - kN_A W_A - A)] - [(\gamma\alpha + \eta Z)/(2\gamma(1-\alpha)]\}.$$

$$\text{(A10)}\quad \partial\theta^*/\partial k = (-mRN_S W_S + mRN_A W_A + AN_S W_S)/[2(kN_S W_S - kN_A W_A - A)^2].$$

The sign of this derivative is dependent on the numerator, given that the denominator is positive. To determine the sign of the numerator, we first observe, from Assumption 5, that

$$\text{(A11)}\quad N_S W_S - N_A W_A - A > 0.$$

Multiplying both sides by $N_A W_A$ and then adding $AN_S W_S$, we get

$$\text{(A12)}\quad N_A W_A N_S - N_A^2 W_A^2 - AN_A W_A + AN_S W_S > AN_S W_S.$$

Factoring the left side, we get

$$\text{(A13)}\quad (N_A W_A + A)(N_S W_S - N_A W_A) > AN_S W_S,$$

and given that each unit is profitable, $mR > N_A W_A + A$, we have

$$\text{(A14)}\quad mR(N_S W_S - N_A W_A) > AN_S W_S,$$

from which we can see that the numerator of Equation A10 is negative, and therefore $\partial\theta^*/\partial k < 0$, and

$$\text{(A15)}\quad \partial P^*/\partial W = (\partial P^*/\partial\theta^*)(\partial\theta^*/\partial W) > 0.$$

Result 5: As factors other than service quality are more influential in driving demand, optimal productivity increases.

$$\text{(A16)}\quad \partial\theta^*/\partial\eta = -Z/[2\gamma(1-\alpha)] < 0.$$

$$\text{(A17)}\quad \partial P^*/\partial\eta = (\partial P^*/\partial\theta^*)(\partial\theta^*/\partial\eta) > 0.$$

Tentative Result 1: The optimal productivity level increases over time.

If we assume that the level of technology increases over time, which means that $\partial\alpha/\partial t > 0$, then

$$\text{(A18)}\quad \partial\theta^*/\partial t = (\partial\theta^*/\partial\alpha)(\partial\alpha/\partial t) = (\partial\alpha/\partial t)\{[-\gamma(1-\alpha) - (\gamma\alpha + \eta Z)]/[2\gamma(1-\alpha)^2]\} < 0.$$

$$\text{(A19)}\quad \partial P^*/\partial t = (\partial P^*/\partial\theta^*)(\partial\theta^*/\partial t) > 0.$$

Tentative Result 2: As market concentration increases, optimal productivity increases.

If we assume that a more concentrated market means that service quality has less effect on demand, $\partial\gamma/\partial H < 0$, where H is market concentration and γ is the effect of the service quality on demand, then

$$\text{(A20)}\quad \partial\theta^*/\partial H = (\partial\theta^*/\partial\gamma)(\partial\gamma/\partial H) = (\partial\gamma/\partial H)\{(\eta Z)/[2\gamma^2(1-\alpha)]\} < 0.$$

$$\text{(A21)}\quad \partial P^*/\partial H = (\partial P^*/\partial\theta^*)(\partial\theta^*/\partial H) > 0.$$

Multiple Firm Objectives

The firm wants to maximize the following function with respect to service productivity, P:

(A22) $$G = \lambda_1\Pi + \lambda_2 S + \lambda_3 Q.$$

Again, because it simplifies the algebra, we optimize with respect to θ and then build the expression for the optimal productivity.

(A23) $$\partial G/\partial\theta = \lambda_1(\partial\Pi/\partial\theta) + \lambda_2(\partial S/\partial\theta) + \lambda_3(\partial Q/\partial\theta) = \lambda_1(\partial\Pi/\partial\theta) + (\lambda_2 + \lambda_3\gamma)(1 - \alpha).$$

Checking second-order conditions, $\partial^2G/\partial\theta^2 = \lambda_1(\partial^2\Pi/\partial\theta^2) < 0$, ensuring a maximum. The optimal labor usage level is

(A24) $$\theta^{**} = \theta^* + \{(\lambda_2 + \lambda_3\gamma)/[2\lambda_1\gamma(N_SW_S - N_AW_A - A)]\},$$

where θ^* is the profit maximizing level of labor usage, which corresponds to an optimal productivity level of

(A25) $$P^* = R/[\theta^{**}N_S + (1 - \theta^{**})N_A].$$

We note that

(A26) $$\partial P^*/\partial\theta^{**} = -R(N_S - N_A)/[\theta^{**}N_S + (1 - \theta^{**})N_A]^2 < 0,$$

because $N_S > N_A$ (automation requires less labor).

The impact of more emphasis on service quality is evident from

(A27) $$\partial P^*/\partial\lambda_2 = (\partial P^*/\partial\theta^{**})(\partial\theta^{**}/\partial\lambda_2),$$

where

(A28) $$\partial\theta^{**}/\partial\lambda_2 = [2\lambda_1\gamma(N_SW_S - N_AW_A - A)]^{-1} > 0,$$

because $N_SW_S - N_AW_A - A > 0$ from Assumption 5. Thus, $\partial P^*/\partial\lambda_2 < 0$, which shows that more emphasis on service quality leads to lower productivity.

The impact of more emphasis on sales/market share is evident from

(A29) $$\partial P^*/\partial\lambda_3 = (\partial P^*/\partial\theta^{**})(\partial\theta^{**}/\partial\lambda_3),$$

where

(A30) $$\partial\theta^{**}/\partial\lambda_3 = \gamma/[2\lambda_1\gamma(N_SW_S - N_AW_A - A)] > 0,$$

and thus, $\partial P^*/\partial\lambda_3 < 0$, which shows that more emphasis on sales/market share leads to lower productivity.

Price Decisions

From Equations 13 and 14, we build the profit function as

(A31) $$\Pi = aRS - bR^3 - acS^2 + bcR^2S.$$

We find the partial derivatives with respect to R and S to be

(A32) $$\partial\Pi/\partial R = aS - 3bR^2 + 2bcRS,$$

$$\partial\Pi/\partial S = aR - 2acs + bcR^2,$$

$$\partial^2\Pi/\partial R^2 = 2b(cS - 3R) < 0 \text{ if } R > cS/3,$$

$$\partial^2\Pi/\partial S^2 = -2ac < 0, \text{ and}$$

$$\partial^2\Pi/\partial R\partial S = a + 2bcR.$$

Setting the first partial derivatives equal to zero, two solutions emerge, corresponding to $R = a/(2bc)$ and $R = a/(bc)$. Checking the second order conditions, we need the second partial derivatives to be negative, and also

(A33) $$(\partial^2\Pi/\partial R^2)(\partial^2\Pi/\partial S^2) > (\partial^2\Pi/\partial R\partial S)^2.$$

The inequality in Equation A33 holds only for $R^* = a/(2bc)$, which corresponds to the service quality level $S^* = a/(8bc^2)$. We then verify that $R > cS/3$, as needed, ensuring that the solution maximizes profit.

Because we note that both service quality and price are increased with higher a and decreased with higher b and c, we observe that higher price is associated with higher service quality, and from Assumption 4, which states that increased labor usage (worse productivity) is associated with higher service quality, we observe that higher price is associated with lower productivity, consistent with Result 3. Likewise, higher wage rates, w, increase the cost of service, c, ($\partial c/\partial w > 0$), which enables us to evaluate the impact of wage rates on productivity:

(A34) $$\partial P^*/\partial W = (\partial P^*/\partial S^*)(\partial S^*/\partial c)(\partial c/\partial w) > 0,$$

because we can assume $\partial S^*/\partial P < 0$ as in Assumption 4 of the original model, and $\partial S^*/\partial c = -a/(4bc^3) < 0$, we have $\partial P^*/\partial W > 0$, higher wage rates lead to higher optimal productivity, consistent with Result 4 of the original model.

Assuming that a more concentrated market lessens service quality impact on demand (because it is more difficult for unhappy customers to switch), we explore the effect of impact of service quality on demand, a, on optimal productivity.

(A35) $$\partial P^*/\partial a = (\partial P^*/\partial S^*)(\partial S^*/\partial a),$$

and because $\partial S^*/\partial a = 1/(8bc^2) > 0$, we have $\partial P^*/\partial a < 0$, which says that when service quality has more impact on demand, optimal productivity should be lower, again replicating the result of the original model and providing additional partial evidence for Tentative Result 2.

Capacity Constraints

Equation 18 indicates that as Q_{max} decreases, service quality is reduced, because the effect of other factors on sales is the same, and it makes no sense to drive demand beyond capacity. Equation 3 implies that

(A36) $$\theta_{cap} = (S_{cap} - \alpha)/(1 - \alpha),$$

from which we infer that the service quality reduction implies a reduced use of labor. Equation 1 indicates that a reduced use of labor implies higher productivity. Thus, a binding capacity constraint (capacity less than the otherwise optimal quality) implies increased optimal productivity, less use of labor, and reduced service quality.

In the case in which service quality, S, is held constant, we can solve for the optimal price, R^*, from Equation 12, yielding the following:

(A37) $$R^* = [(aS - Q_{max})/b]^{1/2}.$$

We note that

(A38) $$\partial R/\partial Q_{max} = (-1/2b^2)(aS - Q_{max})^{-1/2} < 0.$$

This indicates that the lower the capacity constraint is, the higher the price should be.

REFERENCES

Accenture (2005), “Poor Customer Service Is Top Reason Consumers Switch Service Providers, Accenture Survey Finds,” (July 26), (accessed May 30, 2011), [available at http://newsroom.accenture.com/article_display.cfm?article_id=4238#rel].

Aksoy, Lerzan, Bruce Cooil, Christopher Groening, Timothy L. Keiningham, and Atakan Yalçın (2008), “The Long-Term Stock Market Valuation of Customer Satisfaction,” *Journal of Marketing*, 72 (July), 105–122.

Anderson, Eugene W., Claes Fornell, and Donald R. Lehmann (1994), “Customer Satisfaction, Market Share, and Profitability: Findings from Sweden,” *Journal of Marketing*, 58 (July), 53–66.

———, ———, and Sanal Mazvancheryl (2004), “Customer Satisfaction and Shareholder Value,” *Journal of Marketing*, 68 (October), 172–85.

———, ———, and Roland T. Rust (1997), “Customer Satisfaction, Productivity, and Profitability: Differences Between Goods and Services,” *Marketing Science*, 16 (2), 129–45.

Baily, Martin H. and Robert Z. Lawrence (2001), “Do We Have a New E-conomy?” *American Economic Review*, 91 (2), 308–312.

Banker, Rajiv D., Hsihui Chang, and Ram Natarajan (2005), “Productivity Change, Technical Progress, and Relative Efficiency Change in the Public Accounting Industry,” *Management Science*, 51 (2), 291–304.

Basker, Emek (2007), “The Causes and Consequences of Wal-Mart’s Growth,” *Journal of Economic Perspectives*, 21 (3), 177–98.

Bendapudi, Neeli and Robert P. Leone (2003), “Psychological Implications of Customer Participation in Co-Production,” *Journal of Marketing*, 67 (January), 14–28.

Berry, Leonard L., A. Parasuraman, and Valarie A. Zeithaml (1988), “The Service–Quality Puzzle,” *Business Horizons*, 31 (September/October), 35–43.

Bertschek, Irene and Ulrich Kaiser (2004), “Productivity Effects of Organizational Change: Microeconometric Evidence,” *Management Science*, 50 (3), 394–404.

Blinder, Alan, Elie Canetti, David Lebow, and Jeremy Rudd (1998), *Asking About Prices: A New Approach to Understanding Price Stickiness*. New York: Russell Sage Foundation.

Brown, Donna and Alf Crossman (2000), “Employer Strategies in the Face of a National Minimum Wage: An Analysis of the Hotel Sector,” *Industrial Relations Journal*, 31 (3), 206–219.

Brown, James R. and Chekitan S. Dev (2000), “Improving Productivity in a Service Business: Evidence from the Hotel Industry,” *Journal of Service Research*, 2 (4), 339–54.

Brynjolfsson, Erik and Lorin Hitt (1994), “New Evidence on the Returns of Information Systems,” working paper, MIT Sloan School of Management, Massachusetts Institute of Technology.

Caceres, Ruben Chumpitaz and Nicholas G. Paparoidamis (2007), “Service Quality, Relationship Satisfaction, Trust, Commitment and Business-to-Business Loyalty,” *European Journal of Marketing*, 41 (7/8), 836–67.

Conti, Gabriella (2005), “Training, Productivity and Wages in Italy,” *Labour Economics*, 12 (4), 557–76.

Converse, P.D. (1939), “Labor Saving Devices in Marketing,” *Journal of Marketing*, 4 (April), 150–56.

Cronin, J. Joseph, Jr., and Steven A. Taylor (1992), “Measuring Service Quality: A Reexamination and Extension,” *Journal of Marketing*, 56 (July), 55–68.

Cyert, Richard M. and James G. March (1992), *A Behavioral Theory of the Firm*, 2d ed. Malden, MA: Blackwell Publishers.

Datta, Deepak K., James P. Guthrie, and Patrick M. Wright (2005), “Human Resource Management and Labor Productivity: Does Industry Matter?” *Academy of Management Journal*, 48 (1), 135–45.

Deming, W. Edwards (1986), *Out of the Crisis*. Cambridge, MA: MIT Press.

Dewan, Sanjeev and Chung-ki Min (1997), “The Substitution of Information Technology for Other Factors of Production: A Firm Level Analysis,” *Management Science*, 43 (12), 1660–75.

Filiatrault, Pierre, Jean Harvey, and Jean-Charles Chebat (1996), “Service Quality and Service Productivity Management Practices,” *Industrial Marketing Management*, 25 (3), 243–55.

Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, and Barbara Everitt Bryant (1996), “The American Customer Satisfaction Index: Nature, Purpose, and Findings,” *Journal of Marketing*, 60 (October), 7–18.

———, Sunil Mithas, Forrest Morgeson, and M.S. Krishnan (2006), “Customer Satisfaction and Stock Prices: High Returns, Low Risk,” *Journal of Marketing*, 70 (January), 3–14.

———, Roland T. Rust, and Marnik G. Dekimpe (2010), “The Effect of Customer Satisfaction on Consumer Spending Growth,” *Journal of Marketing Research*, 47 (February), 28–35.

Goolsby, Kathleen (2009), “Enabling Offshored Call Centers to Move Back on Shore,” *Outsourcing Journal*, (September), (accessed May 16, 2010), [available at http://www.outsourcing-journal.com/sep2009-callcenter.html].

Grönroos, Christian and Katri Ojasalo (2004), “Service Productivity: Towards a Conceptualization of the Transformation of Inputs into Economic Results in Services,” *Journal of Business Research*, 57 (4), 414–23.

Gruca, Thomas S. and Lopo L. Rego (2005), “Customer Satisfaction, Cash Flow and Shareholder Value,” *Journal of Marketing*, 69 (July), 115–30.

Guthrie, James P. (2001), “High-Involvement Work Practices, Turnover and Productivity: Evidence from New Zealand,” *Academy of Management Journal*, 44 (1), 180–90.

Haltiwanger, John C., Julia I. Lane, and James R. Spletzer (2007), “Wages, Productivity, and the Dynamic Interaction of Businesses and Workers,” *Labour Economics*, 14 (3), 575–602.

Homburg, Christian, Nicole Koschate, and Wayne D. Hoyer (2005), “Do Satisfied Customers Really Pay More? A Study of the Relationship Between Customer Satisfaction and Willingness to Pay,” *Journal of Marketing*, 69 (April), 84–96.

Horstmann, Ignatius J. and Sridhar Moorthy (2003), “Advertising Spending and Quality for Services: The Role of Capacity,” *Quantitative Marketing and Economics*, 1 (3), 337–65.

Hunter, Gary K. and William D. Perreault Jr. (2007), “Making Sales Technology Effective,” *Journal of Marketing*, 71 (January), 16–34.

Huselid, Mark A. (1995), “The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance,” *Academy of Management Journal*, 38 (3), 635–72.

Intel Corporation (2005), “Excerpts from a Conversation with Gordon Moore: Moore’s Law,” video transcript, (accessed October 24, 2011), [available at ftp://download.intel.com/museum/Moores_Law/Video-transcripts/Excepts_A_Conversation_with_Gordon_Moore.pdf].

Juran, Joseph M. (1970), *Quality Planning and Analysis*. New York: McGraw-Hill.

Karimi, Jahangir, Toni M. Somers, and Yash P. Gupta (2001), "Impact of Information Technology Management Practices on Customer Service," *Journal of Management Information Systems*, 17 (4), 125–58.

Keiningham, Timothy L., Tiffany Perkins-Munn, and Heather Evans (2003), "The Impact of Customer Satisfaction on Share-of-Wallet in a Business-to-Business Environment," *Journal of Service Research*, 6 (1), 37–50.

Koch, Marianne J. and Rita Gunther McGrath (1996), "Improving Labor Productivity: Human Resource Management Policies Do Matter," *Strategic Management Journal*, 17 (5), 335–54.

Kordupleski, Raymond E., Roland T. Rust, and Anthony J. Zahorik (1993), "Why Improving Quality Doesn't Improve Quality (or Whatever Happened to Marketing?)," *California Management Review*, 35 (3), 82–95.

Liu, Xueming and Christian Homburg (2007), "Neglected Outcomes of Customer Satisfaction," *Journal of Marketing*, 71 (April), 133–49.

Marinova, Detelina, Jun Ye, and Jagdip Singh (2008), "Do Frontline Mechanisms Matter? Impact of Quality and Productivity Orientations on Unit Revenue, Efficiency, and Customer Satisfaction," *Journal of Marketing*, 72 (March), 28–45.

McCue, Andy (2004), "For Dell, Indian Call Center Failure a Lesson," CNET News.com, (March 31), (accessed May 16, 2010), [available at http://news.cnet.com/For-Dell,-Indian-call-center-failure-a-lesson/2100-1001_3-5182611.html].

Meuter, Matthew L., Mary Jo Bitner, Amy L. Ostrom, and Stephen W. Brown (2005), "Choosing Among Alternative Service Delivery Modes: An Investigation of Customer Trial of Self-Service Technologies," *Journal of Marketing*, 69 (April), 61–83.

———, Amy L. Ostrom, Robert I. Roundtree, and Mary Jo Bitner (2000), "Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters," *Journal of Marketing*, 64 (July), 50–64.

Mittal, Vikas, Eugene W. Anderson, Akin Sayrak, and Pandu Tadikamalla (2005), "Dual Emphasis and the Long-Term Financial Impact of Customer Satisfaction," *Marketing Science*, 24 (4), 544–58.

Napoleon, Karen and Cheryl Gaimon (2004), "The Creation of Output and Quality in Services: A Framework to Analyze Information Technology-Worker Systems," *Production and Operations Management*, 13 (3), 245–59.

Narver, John C. and Stanley F. Slater (1990), "The Effect of a Market Orientation on Business Profitability," *Journal of Marketing*, 54 (October), 20–35.

Noble, Charles H., Rajiv K. Sinha, and Ajith Kumar (2002), "Market Orientation and Alternative Strategic Orientations: A Longitudinal Assessment of Performance Implications," *Journal of Marketing*, 66 (October), 25–39.

Oliva, Rogelio and John D. Sterman (2001), "Cutting Corners and Working Overtime: Quality Erosion in the Service Industry," *Management Science*, 47 (7), 894–914.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (Spring), 12–40.

Perloff, Jeffrey M. (2011), *Microeconomics*. Upper Saddle River, NJ: Prentice Hall.

Rao, Vithala R., Manoj K. Agarwal, and Denise Dahlhoff (2004), "How Is Manifest Branding Strategy Related to the Intangible Value of a Corporation?" *Journal of Marketing*, 68 (October), 126–41.

Rayport, Jeffrey F. and Bernard J. Jaworski (2005), *Best Face Forward: Why Companies Must Improve Their Service Interfaces with Customers*. Cambridge, MA: Harvard Business School Press.

Reinartz, Werner, Jacquelyn S. Thomas, and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability," *Journal of Marketing*, 69 (January), 63–79.

Rust, Roland T., Christine Moorman, and Peter R. Dickson (2002), "Getting Return on Quality: Cost Reduction, Revenue Expansion, or Both?" *Journal of Marketing*, 66 (October), 7–24.

——— and Anthony J. Zahorik (1993), "Customer Satisfaction, Customer Retention and Market Share," *Journal of Retailing*, 69 (Summer), 193–215.

———, ———, and Timothy L. Keiningham (1995), "Return on Quality (ROQ): Making Service Quality Financially Accountable," *Journal of Marketing*, 59 (April), 59–70.

Schmalensee, Richard (1977), "Using the H-Index of Concentration with Published Data," *Review of Economics and Statistics*, 59 (May), 186–93.

Singh, Jagdip (2000), "Performance Productivity and Quality of Frontline Employees in Service Organizations," *Journal of Marketing*, 64 (April), 15–34.

Storbacka, Kaj, Tore Strandvik, and Christian Grönroos (1994), "Managing Customer Relationships for Profit: The Dynamics of Relationship Quality," *Journal of Service Industry Management*, 5 (5), 21–38.

Wangenheim, Florian v. and Tomas Bayon (2007), "The Chain from Customer Satisfaction via Word-of-Mouth Referrals to New Customer Acquisition," *Journal of the Academy of Marketing Science*, 35 (2), 233–49.

Winch, Guy (2011), "New Survey: Bad Customer Service Impairs Our Quality of Life," (April 20), (accessed October 24, 2011), [available at http://www.guywinch.com/new-survey-bad-customer-service-impairs-our-quality-of-life/].

Zeithaml, Valarie A. (2000), "Service Quality, Profitability, and the Economic Worth of Customers: What We Know and What We Need to Learn," *Journal of the Academy of Marketing Science*, 28 (1), 67–85.

———, Leonard L. Berry, and A. Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60 (April), 31–46.