Scholarly Article

Optimizing Service Productivity With Substitutable and Limited Resources

Journal of Service Research
2023, Vol. 0(0) 1–16

DOI: 10.1177/10946705231213118
journals.sagepub.com/home/jsr
Sage

Jens Hogreve[1], Alexander Hübner[2], and Mirjam Dobmeier[3]

Abstract

This article develops a decision model which enables service firms to optimize their productivity. Companies must efficiently determine the necessary resource input to increase service productivity to meet customer demand. In so doing, managers face service-specific challenges: They must select the appropriate type and quantity of limited resources to deliver services efficiently, consider the volatility of demand to provide services effectively, and integrate the interaction effects of resources in terms of substitution to utilize constraint resources optimally. In addressing these challenges, we develop an interdisciplinary approach by combining insights from service research and operations research to create a decision model that helps managers select the optimal type and quantity of resources available to overcome the abovementioned challenges. We validate our model in several case studies and further generalize our findings by applying it to different data settings. Ultimately, we prove that productivity can be increased significantly if firms optimize resource selection by considering stochastic demand, the effects of substitution among resources, and resource constraints.

Keywords

service productivity, service efficiency, decision theory, stochastic optimization, simulation, resource selection, capacity constraint

The growing importance of technology and artificial intelligence (Huang and Rust 2018, 2021) and the shortage of service employees (e.g., Bhattarai and Penman 2023; Horowitz 2021) in service encounters forces service firms to address a significant challenge: Companies must efficiently determine the necessary resource input (e.g., level of human resources and computer-aided support) to meet customer demand. This optimization of service productivity directly impacts the bottom-line profitability of service firms. However, while technological advancements progress and services became pivotal for economic growth, service productivity is declining in many developed countries, suggesting that productivity-enhancing approaches at the firm level have yet to materialize (Hofmeister, Kanbach, and Hogreve 2023a). Therefore, in his call for action, Andreasen (2021) highlights the necessity to research service productivity so service firms can achieve sustainable growth.

The managerial issue of obtaining high service productivity is present in nearly all industries. Healthcare industries further exemplify the productivity issues service firms face in input- and output-related decisions: The COVID-19 pandemic has shown that limited hospital resources such as nurses and beds must be used as efficiently as possible. At the same time, a sufficient level of patient care and service quality must be maintained. Therefore, hospital managers need to optimize the workload for medical staff, even in unexpected emergencies, while operating with specific capacities and limited idle times to ensure cost-efficient operations. These challenges in managerial practice are widely acknowledged in academic literature focusing on service productivity (e.g., Anderson, Fornell, and Rust 1997; Mittal et al. 2005; Rust and Huang 2012; Wirtz and Zeithaml 2018).

Optimizing service productivity requires strategic and efficient planning of capacities as resources cannot be scaled up or down on a short-term basis (Mittal et al. 2005; Rust and Chung 2006). Thus, our research develops a strategic decision support system to optimize service productivity based on the seminal work of Rust and Huang (2012). Adopting a profit-optimization perspective, Rust and Huang (2012) demonstrate how firms can obtain higher service productivity by optimizing resource selection. We add to this research by extending the optimization model by Rust and Huang (2012) and contribute in several

[1]Professor and Chair of Service Management, Ingolstadt School of Management, Catholic University of Eichstaett-Ingolstadt, Germany
[2]Professor and Chair of Supply and Value Chain Management, School of Management and Campus for Biotechnology and Sustainability, Technical University of Munich, Germany
[3]Postdoctoral Researcher, Ingolstadt School of Management, Catholic University of Eichstaett-Ingolstadt, Germany

Received: 20 March 2023; accepted: 11 October 2023

Corresponding Author:
Jens Hogreve, Ingolstadt School of Management, Catholic University of Eichstaett-Ingolstadt, Auf der Schanz 49, Ingolstadt 85051, Germany.
Email: jens.hogreve@ku.de

ways. First, Rust and Huang (2012) are the first to research service productivity within a decision context. In their model, the authors show that productivity increases can be obtained by the optimal selection of two main resources: automation and labor. In managerial practice, however, firms usually deal with multiple resources. Furthermore, each resource type is available at different quality levels. For example, in a hospital, physicians with different seniority and qualification levels are essential for patient care; other resources include nurses, administrators, medical devices, and technologies. Thus, our model considers multiple resources and different resource qualities to increase managerial applicability. We deal with a strategic decision problem where managers need to make decisions in the long-term context. In this context, the decision-maker determines the resources on a longer planning horizon.

Second, customer demand for services is not fully known in advance and is subject to deviations. However, whenever customer demand exceeds the capacity of a service firm, profits are lost; the opportunity to sell a service will either disappear, or high investments will be needed to win back lost customers. An emergency room, for instance, must provide sufficient health care levels, even in unexpected emergencies. Therefore, patients must be transferred to other hospitals if demand exceeds capacity. We address this critical issue by considering stochastic demand, thus revealing how capacities can be managed strategically and on a long-term basis best in the face of demand volatility. The current state-of-the-art in the service literature is to model known deterministic demand. From a scientific perspective, we contribute to the literature by following a stochastic approach capable of handling uncertainties in the inputs applied.

Service managers must strategically select their resources in light of the constraints they face (e.g., labor shortage, budget constraints, flexibility of technological resources) and of different employee qualification levels (e.g., highly flexible and experienced workers, less experienced workers that can fulfill only specific tasks). These resources may be permanently or temporarily substituted. Again, in the hospital context, take surgery services as an example: robot-assisted surgery technologies partially substitute for the work of physicians, expediting a patient's healing process and shortening expensive hospital stays. We extend previous work by considering resource constraints along with permanently and temporarily substitutable resources. This yields a more sophisticated sense of how service productivity can be optimized through resource allocation, capacity management, and substitution effects.

Finally, we test our model using case studies with industry applications and a general data set. This reveals how service productivity can be optimized across different industries and settings (e.g., resources with varying qualification and quality levels or high degrees of customer coproduction), as well as enhances current literature with general insights about optimizing service productivity given multiple resources, resources constraints, stochastic demand, and permanently or temporarily substitutable resources. We demonstrate that correctly considering stochastic demand enables firms to increase performance significantly, and we identify substituting resources as a significant lever of service productivity management.

The Service Productivity Challenges

Traditionally, productivity is defined as the ratio of inputs and outputs (Deming 1986). Services, however, inherently call for a broader understanding of productivity given that they involve several input- and output-related issues (Grönroos and Ojasalo 2004; Rust and Huang 2012; Wirtz and Zeithaml 2018). We address these issues and define service productivity as a strategic decision variable to optimize firms' profits (Rust and Huang 2012).

In managing service productivity, we identify three main challenges: First, firms need to balance long-term investments and generate a service output that creates external interest and meets operational objectives to get an optimal response for their resources at the same time (Anderson, Fornell, and Rust 1997; Grönroos and Ojasalo 2004; Mittal et al. 2005; Rust and Huang 2012; Rust, Moorman, Dickson 2002; Rust, Zahorik, and Keiningham 1995). Service productivity depends on various resources, including, for example, service employees and their ability as well as a willingness to deliver service quality (Marinova, Ye, and Singh 2008; Singh 2000), customers' willingness to coproduce (Auh et al.2019; Nachum 1999, Xue and Harker 2002), and the information technology supporting the service delivery process (Hogreve and Beierlein 2023; Huang and Rust 2021; Rust and Huang 2012). Each resource adds to the quality of the service outcome and generates revenues. Still, each has its costs, for example, wages, training, and development, or ensuring the functionality of the technical infrastructure. Consequently, we delineate the first challenge firms are confronted with while optimizing their levels of service productivity:

1. *To optimize service productivity, managers must determine on a strategic level the optimal type and quantity of resources from a set of multiple yet limited resources (e.g., due to labor shortage).*

Second, service creation and consumption happen simultaneously (Zeithaml, Bitner, and Gremler 2017). As services cannot be stored, meeting the volatile and unknown customer demand is a powerful lever in obtaining high service productivity by efficient capacity utilization (Armistead and Clark 1994). This is relevant as the resources cannot be scaled up or down on a short-term basis. Companies need to provide these resources from a long-term perspective and can adjust capacities within longer planning periods. In the hospital, for example, other departments within the hospital could send physicians and nurses to compensate for capacity shortages. From a governmental planning perspective, the planners of an entire region could share resources across hospitals. Similarly, the resources of an entire company network (e.g., with different subsidiaries) can be shared across the different entities (e.g., different locations). Limited capacities might mean that certain resources are temporarily or permanently unavailable yet may be substituted; for example, service technologies might substitute for service employees, or customer input might replace employees' input during coproduction (Mills and Morris 1986). This might happen due to manifold reasons, such as short-term breakdown of self-service devices, shortages of human resources, or demand exceeding management expectations. Such substitutions affect service outcomes in terms of quality and costs. For example, less experienced employees might be less costly than experienced employees, yet negatively affect service quality (Marinova, Ye, and Singh 2008; Meyer Goldstein 2003; Nachum 1999; Singh 2000). The same is true for customer coproduction and its effects on service outcomes depending on customers' willingness and ability to coproduce (Auh et al. 2019; Bendapudi and Leone 2003).

Consequently, the decision problem in optimizing service productivity formulated in our first statement is increased as it cannot be done for each resource individually. Additionally, decisions about which type and quantity of resources to invest among a set of multiple, substitutable, and limited resources are required. We thus formulate the second challenge:

2. *To optimize service productivity, managers must consider resources' limited availability and substitutability when deciding on the optimal selection of resources.*

Third, resources might be temporarily or permanently unavailable and might be substituted, for instance, customer input, substituting employees' input during coproduction processes (Mills and Morris 1986). Although we find incidents concerning these issues in current service research (e.g., Haumann et al. 2015), they are not fully addressed in the service management literature. In a service context, supply and demand are interrelated as service production and consumption happen simultaneously in real-time. Controlling service quality, for example, in terms of individualization or achieving economies of scale, are much more complex than products (Chase 1978, 1981; Wirtz and Zeithaml 2018) as demand cannot be backlogged, excess demand or capacity cannot be stored or transferred to another period, and hence the demand and profit of servicing these customers are lost. Customers usually arrive at a service provider and are either served within the period of their arrival or turned away (Armistead and Clark 1994; Armstrong, Morwitz, and Kumar 2000; McLaughlin and Coffey 1990; Wirtz and Zeithaml 2018). Consequently, managing service productivity requires accounting for highly volatile demand (Dobni 2004; Rust and Chung 2006). We thus phase the third challenge:

3. *To optimize service productivity, unknown customer demand must be considered.*

Current research on service productivity only partially addresses these challenges in combination. Instead, the focus has fallen on how single antecedents like human resources (Marinova, Ye, and Singh 2008; Singh 2000), technology (De Jong, De Ruyter, and Lemmink 2003; Huang and Rust 2021; Rust and Huang 2012), processes and service design (Melton & Hartline, 2013; Nakata Cheryl & Hwang Jiyoug, 2020), and customers (Frei and Harker 1999; Haumann et al. 2015; Xue and Harker 2002) influence efficiency and/or effectiveness of services and how these antecedents can be managed accordingly. In addition, substituting or interacting effects of resources generally are only partially considered for two resources (e.g., Rust and Huang 2012), whereas service firms usually have to choose from many substitutable resources.

Yet challenges in managing service productivity go beyond which resources to invest. Service managers must balance resource allocations and efforts to meet customers' expectations (e.g., service quality) as well as internal operational objectives (e.g., budget constraints, costs) to get optimal economic results depending on actual demand (Marinova, Ye, and Singh 2008; Rust, Zahorik, and Keiningham 1995; Wirtz and Zeithaml 2018). To fully address the complexity of service productivity, we need a more nuanced understanding and a comprehensive decision model on how to select the right type and quantity of resources from a limited pool of multiple resources (such as human resources, technology, and customer coproduction) under given demand volatility and possible substitutions of resources.

Drawing on decision theory, Rust and Huang's (2012) seminal work discusses service productivity as a strategic decision variable that allows firms to select their optimal level of service productivity and optimize their profits. Yet the practical applicability of this model could be expanded because the particularities of managing service productivity (like specific resource selection, stochastic demand, and substitutions) are only partially addressed. However, literature and managerial practice need more tools and methods to gain the optimal balance of efficiency and effectiveness through strategically optimizing resource selection (Mittal et al. 2005; Rust and Chung 2006). "What is needed is a method to help managers decide where they are likely to get the greatest response for their limited resources" (Rust, Zahorik, and Keiningham 1995, p. 59). With our proposed decision model and its application, we seek to fill this apparent gap in literature.

Decision Problem and Associated Model

We build on the work of Rust and Huang (2012) by developing a strategic decision model that helps managers to optimize service productivity by selecting resources and determining the quantity of substitutable and constrained resources required to deliver services under stochastic demand. We denote the decision problem as Capacitated Service Productivity Model with Substitutions ($CSPM_{SUB}$).

Decision Model and Resource Constraints

As outlined in our first challenge, managers need to select resources j from a set of resources $j \in \mathbb{N}$ and to define associated quantity levels for each resource considering limitations. For example, hospital management needs to define how many nurses with different qualification levels they want to engage in a specific clinical department. A consultancy firm needs to define a strategic level how many senior, mid-level, and junior consultants they want to hire for the upcoming planning cycle to meet their expected demand. In these examples, the seniority levels are the resource types, and the headcount of each type is the quantity level. As services cannot be stored, we draw on a single-period model. That means we have a long-term planning horizon and define the resources for this long-term horizon. Our model optimizes profits by optimizing resource selection and quantity for all items j with $N = \{1, 2, ..., j, ... \mathbb{N}\}$, whereas $\mathbb{N}$ denotes the entire set of resources (e.g., senior, mid-level, and junior workers). Hereby resource selection is expressed by the binary variable x_j, which is set to 1 if resource j is available and 0 otherwise. The second decision variable of quantity q_j determines the quantity level of each resource the firm provides (e.g., number of employees of a certain qualification level). The model is formulated as follows:

$$max\,!\,\Pi(\overline{x}, \overline{q}) = \sum_{j\in\mathbb{N}} \pi_j(q_j) \cdot x_j \tag{1}$$

subject to

$$\sum_{j\in\mathbb{N}} q_j \cdot b_j \leq C \tag{2}$$

$$x_j \in \{0;1\}; q_j \geq 0 and\ integer \forall j \in \mathbb{N} \tag{3}$$

The objective function is quantified in equation (1). Equation (2) limits the available resources to the capacity constraint C. The capacity coefficient is represented by b_j for all items $j\in\mathbb{N}$, that is, the capacity consumption to use one unit of resource j. Equation (3) allows only binary values for x_j as well as positive and integer values for the quantity q_j. The quantity-related profit function $\pi_j(q_j)$ is given by equation (4). Resources are characterized by revenue r, input costs c, salvage value v, and shortage costs s. The profit $\pi_j(q_j)$ per resource depends on the quantity q_j provided for each resource $j\in\mathbb{N}$, and consists of the revenue r_j, which is offset by input costs c_j as well as the trade-off between shortage costs for unsatisfied demand s_j and salvage value v_j for unused resources.

$$\begin{aligned}\pi_j(q_j) &= -q_j c_j + r_j \int_0^{q_j} y f_j^* dy + v_j \int_0^{q_j} (q_j - y) f_j^* dy \\ &+ r_j \int_{q_j}^{\infty} q_j f_j^* dy - s_j \int_{q_j}^{\infty} q_j (y - q_j) f_j^* dy \quad \forall j \in \mathbb{N}\end{aligned} \tag{4}$$

Beyond optimal resource selection, we further consider leftovers: Whenever resources remain unused, there are salvage values. In the best case, those remaining resources can be used for services with lower profitability or, in the worst case, perish fully without being consumed. For example, a senior worker can fulfill the jobs of a junior worker. Furthermore, demand cannot be backlogged, and if demand has been underestimated, shortage costs occur (e.g., costs to retain unsatisfied customers). To factor in stochastic demand, the estimated demand D_j for item j is defined by the probability density function $f_{D_j}(y)$, which specifies the demand function in equation (4) for the model: Please note that any kind of demand distribution can be applied to model the customer demand. The only requirement is non-negative demand values.

Addressing these issues, the first term in equation (4) quantifies the total input costs for the resource j reflecting all processing or purchase costs the firm requires to provide this particular resource. The second term quantifies total revenues generated with the quantity q_j of resource j. The third term denotes the salvage values if resource input has been overestimated and resources remain unused at the end of the period for disposal or discounted sales. In such case, resources need to be disposed at salvage value v_j and the provider incurs a loss of $(c_j - v_j)$ on each resource (e.g., when overqualified workers perform a job below their qualification level). The fourth term defines lost sales costs. If the demand for resource j is greater than its quantity q_j, the excess demand and associated revenues are lost (i.e., a demand cannot be fulfilled as the capacity is not sufficient). Shortage costs s_j occur in such cases representing penalty costs for unsatisfied demand. These costs are expressed in the fifth term.

Demand Model and Substitutions

Beyond the challenge of optimized resource selection, we consider the challenge of substitution effects among resources (e.g., a higher qualified nurse may complete jobs of a lower qualified nurse during a peak period) and the dependence on a stochastic demand (e.g., demand is not per se known when defining the headcount for each qualification level). We consider demand depending on resource availability (e.g., care capacity in a hospital) and the associated quality of a resource (e.g., higher qualified doctors for certain surgeries). For example, consider a less experienced employee who might negatively affect service quality compared to a well-experienced employee and, in turn, negatively impact service demand. Thus, total demand for a selected resource $j, j \in \mathbb{N}^+$ consists of the base demand D_j determined by the quality γ_j of this resource, the gain in demand for substitutions from temporarily unavailable resources, $j \in \mathbb{N}^+$, and of substitutions from the set of unselected resources, $i\in\mathbb{N}^-$, to resource j. Selected and provided resources are denoted by the set $\mathbb{N}^+$ and unselected resources by $\mathbb{N}^-$. Thus, $\mathbb{N}^+, \mathbb{N}^- \subseteq \mathbb{N}, \mathbb{N}^+ \cup \mathbb{N}^- = \mathbb{N}$ and $\mathbb{N}^+ \cap \mathbb{N}^- = \varnothing$. Equation (5) summarizes the three demand components.

$$\begin{aligned}\widehat{D}_j(\overline{x}, \overline{q}) &= D_j(\gamma_j) + \sum_{i\in\mathbb{N}^-, i\neq j} \beta_{ij}^{per} D_i \\ &+ \sum_{i\in\mathbb{N}^+ i\neq j} \beta_{ij}^{temp} [(D_i - q_i) | D_i > q_i] \forall j \in \mathbb{N}^+\end{aligned} \tag{5}$$

The first part of equation (5) denotes the volume that a customer initially prefers of resource j. The base demand is also impacted by the service quality γ_j of a resource. We model a factor for known service quality for each resource, denoted by γ_j with $0 \leq \gamma_j \leq 1$, $j, j \in \mathbb{N}$. This factor expresses a scale parameter of the service quality provided by resource j.

Given that demand is highly volatile, however, it might exceed the capacity of a service firm. Either demand cannot be satisfied, or it may be more profitable to force customers to switch to more profitable substitutes (e.g., if a lower qualified hairdresser is not available, the customer may choose hair cutting from the more qualified hairdresser). Hence, substitution effects are part of the decision-making process. We assume that if a resource is permanently unavailable (e.g., not provided at all due to changing constraints) or temporarily unavailable (perhaps already consumed by other customers during peak demand), it can potentially be replaced by another resource. The probability of replacing one resource with another is expressed as the substitution rate. Substitution effects are integrated into the demand function by combining the base demand for resource j with the substituted demand for all unavailable resources. We generally follow exogenous demand (ED) models. All customers choose their favorite resource i from set $\mathbb{N}$. If their requested resource i is not available for some reason, the substitution rate β_{ij} expresses that they will choose their second favorite j, $j \in \mathbb{N}$, $j \neq i$. Those customers willing to compromise their initial choice are expressed in the probability of δ_i, with $(1 - \delta_i)$ as the share of customers unwilling to substitute their initial choice (i.e., $\delta_i = \sum_{j \in \mathbb{N}, j \neq i} \beta_{ij}, i \in \mathbb{N}$). Furthermore, we differentiate between permanent and temporary substitution rates, that is, β_{ij}^{per} and β_{ij}^{temp}, respectively. The substitution rate β_{ij}^{per} expresses to what extent customers are willing to substitute permanently unavailable resources i, $i \in \mathbb{N}^-$, with resource j. The substitution rate for resources temporarily not available i, $i \in \mathbb{N}^+$ is denoted with the substitution rate β_{ij}^{temp}. Generally, the substitution rate refers to the ability to replace one resource with another. For example, it may not be possible that a junior employee may replace a senior employee for certain tasks. In this case, the substitution rate can be set to zero.

The demand function of each resource j depends on the vector of the decision variables, $\bar{x} = (x_j, j \in \mathbb{N})$ (i.e., type of resources selected) and $\bar{q} = (q_j, j \in \mathbb{N}^+)$ (i.e., the quantity for each resource provided). The substituted demand is reflected in the second and third parts of equation (5). The second part quantifies the substitution volume of permanently unavailable resources. The third part of equation (5) quantifies substitution demand for those resources temporarily not available. Further, following assumptions in ED models, one round of substitution is allowed. If customers want to substitute their first choice for a resource that is not available, sales are lost. There is no attempt to model individual customer decisions. Instead, an ED model capable of capturing aggregated customer demand is applied.

We assume that the probability density function f_{D_j} of the base demand for resource $j, j \in \mathbb{N}$ is exogenously known and does not include substitutions. The generic demand functions for the CSPM_{SUB} are expressed as $f_j^* = f_{D_j}(\overline{x, q}, y), j \in \mathbb{N}$. If the demand function f_j^* in equation (4) is set to the probability density function of the base demand, that is, $f_j^* = f_{D_j}, j \in \mathbb{N}$, the resulting base model ignores substitution effects. In the numerical results section, we investigate the different demand effects step by step and use these different model formulations. The CSPM_{SUB} is the generic model accounting for all effects. The CSPM is a special case of the CSPM_{SUB} with no substitution, that is, $\beta_{ij}^{temp} = 0$, $i, j \in \mathbb{N}^+$, and the demand model can cope with any type of demand distribution. It only needs to allow a non-negative demand. We refer to the Web Appendix and the demand model presented in Equations (9) and (10). These make sure that only any non-negative demand is considered.

To summarize, the total demand for each selected resource $j, j \in \mathbb{N}^+$ depends on the base demand for resource j, its service quality γ_j, the set of not selected resources $i \in \mathbb{N}^-$, and the associated substitution rate to available resources β_{ij}^{per} as well as the availability of selected resources $j \in \mathbb{N}^+$ and the associated substitution rate of resources that are temporarily out of stock, β_{ij}^{temp}. Please note that the selection variable x_j is an auxiliary variable, as we can also model the selection problem in the following way: if $q_j = 0$, then $x_j = 0$; if $q_j \geq 1$, then $x_j = 1$.

The underlying optimization problem is NP-hard and integer knapsack problem with a nonlinear and non-separable objective function. Solving the integer problem with a full enumeration or commercial solver (like CPLEX or Gurobi) is only possible for toy problems. We, therefore, develop a specialized heuristic built on a Lagrange derivation and a bi-section algorithm to determine optimal quantities under constraints and a rounding algorithm to find optimal integer quantities. Details of the solution approach, its detailed computation steps, and its algorithmical approach can be found in the Web Appendix.

Case Studies and Numerical Analysis

To test our model and to obtain managerial insights, we considered several case studies referring to real data from firms offering distinct services and applying different resources (see Table 1). In addition to reviewing annual demand figures and financials, we conducted workshops with firm-intern experts to obtain data that explains to what degree resources can be substituted. In each workshop, up to six experts from the companies were engaged. Real-life data on service and resource performance from various companies has been obtained. However, due to confidentiality and competitive reasons, we are only able to report on ranges rather than specific values. To demonstrate the decision problems faced by service companies, we have carefully selected case studies covering a wide range of productivity issues. These case studies include the internal and external substitution of employees, customer involvement in coproduction, and the impact of technology on productivity. Table 1 provides an overview of the companies involved. The results obtained are reported as aggregated results and relative numbers compared to the status quo. Next, we describe the settings of our five case studies regarding the main resource while illustrating the data collection process and explaining the decisions required to optimize productivity. Finally, we generalize our results with simulated data in a subsequent analysis.

Case Study Setting and Data Description

Internal and External Substitution of Employees. We conducted three case studies to analyze how employees in different settings impact service productivity from an internal and external substitution perspective.

Internal Substitution. For internal employee substitution, we consider substitutions among employees with varying levels of qualification in distinct departments and at distinct subsidiaries. We worked on the internal substitution with two different companies (see Table 1). First, we applied the model in a large maximum-care hospital with different clinical departments and nursing staff as critical resources. Nurses have different qualification levels (e.g., head of the nursing unit, senior nurse, junior nurse, assistant nurse, or transportation and assistance nurses) and specialize in one department. Yet nurses can partially execute jobs for other departments so that they can be substituted. Furthermore, nurses substitute themselves within a department; whereas all can execute standard treatments, for example, more complicated procedures are performed only by senior nurses. The substitution possibility is expressed in the substitution rate μ_{ji} and is assumed to be 75 percent within the group of nurses of one department and 25 percent among identically qualified nurses of other departments. The demand per period (i.e., nursing days per qualification level) is assumed to be a normal distribution with $\mu_j \epsilon$ [50; 780] and $\sigma_j \epsilon$ [1; 230] based on historical demand data over 36 months. As in all cases studies and applications of our model, we made sure that no negative demand values are considered (see equation (9) in the Web Appendix).

Unit costs c_j of the five differently qualified nurses with $j = 1, 2, \ldots, N$ are obtained from their monthly salaries and revenues r_j assumed to be units costs c_j plus an overall margin of the hospital. This reflects that hospitals may charge patients differently for the treatment depending on the qualification level of the caregiver. If nurses are not engaged (e.g., due to low occupancy), they execute different jobs in the clinical departments; however, the revenues obtained are lower and assumed to be 66 percent only of the original job because they are either executing jobs that require a lower qualification or are engaged in other departments they are not qualified for. We include this fact with the salvage value v_j. If service requirements are not met due to the non-availability of nurses, hospitals suffer shortage costs. These shortage costs s_j are expressed for overtime at maximum-care hospitals or additional management costs required to obtain additional capacities. The model thus addresses the strategic problem of obtaining productive nursing capacities in the hospital context.

Second, we apply our model to a manufacturing firm offering maintenance services to industrial customers. The main resources are engineers with distinct qualification levels (e.g., basic mechanics, advanced mechanics, and engineers with a university degree). The firm operates subsidiaries in different locations with various qualification levels and wages (i.e., costs for the resource) by location. The firm charges different revenue rates r_j for each qualification level and location, with different internal unit costs c_j. Engineers can be partially substituted as long as customers accept the different qualification levels. The substitution can happen either by an employee with a different qualification level at the same location or by an employee with the same qualification level from another site. The substitution rates μ_{ij} among engineers are estimated at 50 percent when substituted by an identically qualified engineer from another

Table 1. Overview of Case Studies.

Case Study	Service Setting	Critical Resources	Substitutions
Internal and external substitution of employees	Nursing	Employees at five different qualification levels	Qualification level (high and low)
		Employees at several departments	Departments (departments A and B)
	Aftersales	Employees at six different qualification levels	Qualification level (high and low)
		Employees at several subsidiaries	Location (locations A and B)
	Store replenishment	Employees at two different qualification levels	Qualification level (high and low)
		External service provider (refiller)	Employee or service provider
Customer coproduction	Consulting	Employees at three different qualification levels	Qualification level (high and low)
		Customer	Customer coproduction (high and low)
Technology	Maintenance	Employee (on-site service) Technology (remote service)	Employees and technology

The exact sales and financial data are subject to confidentiality agreements with the companies.

location, 30 percent for the higher qualified level at the same location, and 20 percent for the lower qualified level. Only for the highest (or lowest) qualified staff, there is no upward (or downward) substitution. Cross-location replacement increases unit costs c_j for a service (e.g., due to shipments of material or travel to other countries). Engineers not engaged in customer service during idle times can complete internal jobs such as product development and training. This generates revenues for the alternative jobs at lower rates (on average between 25 percent and 33 percent of the original job) and can be described as salvage value v_j. Further, suppose the firm does not meet customers' expectations toward service quality with respect to timing or quantity of engineers available. In that case, customers may charge penalty costs s_j, which amount to 10-20 percent of the sales price. The demand per period for the services (i.e., an engineer's days per qualification level) is assumed to be normally distributed with $\mu_j \epsilon [12.5; 250]$ and $\sigma_j \epsilon [1.5; 5]$ based on historical demand data over the last 24 months in two countries.

External Substitution. To go beyond firm-internal substitution among employees, we conducted a case study considering the substitution of employees by external service providers. In a retail context, in-store logistics represent an expensive part of the supply chain, and retailers often engage external and dedicated shelf refillers to replenish the shelves. This regular shelf replenishment is scheduled, and the frequency is determined based on the weekly pattern of deliveries. Thus, refillers come on days with warehouse deliveries to execute shelf replenishment. In between the two warehouse deliveries, the regular sales staff restocks if the shelf inventory of a product is too low. Varying sales and delivery sizes result in varying demand for shelf replenishment jobs.

In our case study, the retailer had to determine the appropriate working hours of external refillers ($j = 1$), sales staff ($j = 2$), and store/category managers ($j = 3$) for shelf replenishment. All resources have unit costs c_j per hour and can be fully substituted if higher qualified employees cover the task ($\sum_{j \in N} \mu_{ji} = 1.0 \forall j > i$) but only partially if lower qualified employees cover for higher qualified employees ($\sum_{j \in N} \mu_{ji} < 0.25 \forall j > i$). Salary increases with staff qualification level. Demand for replenishment can be expressed as working hours required to replenish shelves during a period entirely. Demand (with $\mu_j \epsilon [16; 180]$) and its deviations (with $\sigma_j \epsilon [3; 25]$) for replenishment jobs are given by historical engagement plans. If refillers are oversupplied, the retailer engages them with different jobs (e.g., backroom cleaning). The value of alternative jobs is expressed with v_j, but obtains a lower value than primary jobs. We assume $v_j = c_{j-1} \forall j \geq 2$ and $v_1 = 0.5c_1$.

Revenue per resource r_j is calculated based on unit costs for each resource plus the average store margin. This represents the revenues that can be generated by each resource. Shortages of refillers result in overtime costs or, in the worst case, in lost sales due to empty shelves. This is expressed by s_j and assumed with $s_j = 0.1r_j$. In our case study, the share of external refillers responsible for replenishment is 85 percent.

Customer Coproduction. To analyze the distinct resource of customer coproduction, we obtained data from a consulting firm focused on engineering and planning services for logistics systems. Joint planning teams of employees and customers coproduce the design of warehouse systems. Results are achieved in customer coproduction by partially or fully taking over the planning, operations, and maintenance of the logistics system. Each planner thus generates revenues r_j and costs c_j per day, depending on three different qualification levels with $j = 1, 2, 3$ (senior engineers, junior engineers, and trained workers). When a resource is unavailable, jobs can be completed (=substituted) by the nearest two qualification levels of the firm and a member of the customer's team with an identical qualification level. The substitution rates are 40 percent to the next qualification level and 20 percent to the next plus one level within the firm, and the customer team can substitute 40 percent. For the middle qualification level ($j = 2$), 40 percent is substituted upward and 20 percent downward.

If planners engaged exceed requirements, they take over partially internal responsibilities or work part-time for other customers. This generates salvage values v_j, assumed to be only 33%–40 percent of unit costs. If capacities are underestimated, planners must work overtime for the firm to find additional resources or to violate quality objectives. This results in additional costs, represented as shortage costs s_j assumed to be 15%–20 percent of unit costs c_j. The demand per period (i.e., an engineer's days per qualification level) is assumed to be a normal distribution with $\mu_j \epsilon$ [25; 100] and $\sigma_j \epsilon$ [5; 20] based on historical demand data over the last 12 months. About 20 percent of the team are customers' engineers. With an identical ratio across all qualification levels, we assume 20 percent of demand and its variation for customers' engineers. Unit costs of internal engineers that are replaced by customers' engineers represent the revenues of these engineers. Further cost ratios are applied for the internal resources.

Technology. Finally, we focus on technology as a lever of service productivity within a healthcare firm providing hospital maintenance services. The firm can either send employees to the customer's site or use technology to deliver the service remotely. The decision problem is to optimally select the required service level (i.e., the resources x_j) and the appropriate quantity q_j for each service type. For remote service, the firm can log in to a customer's system and maintain or complete the servicing of medical devices from the firm's site. The onsite service ($j = 1$) is more costly and time intensive, whereas the remote service ($j = 2$) requires more intensive technology, which is much less cost intensive. The firm charges different revenue rates r_j for their services according to internal unit costs c_j plus a certain margin for each service, whereas the remote service is more profitable. 95 percent of the services at the customer site can be completed remotely, and a technician can complete 100 percent of the remote jobs. The execution possibility of remote services instead of onsite services depends on the customer's acceptance. The substitution rates μ_{ij} are estimated at 90 percent and 100 percent.

If technicians are not engaged in any customer service during idle times, they can complete other jobs resulting in revenues for alternative jobs, which are described as salvage value v_j. Remote service operators can directly take over other services; thus, alternative revenues nearly compensate for their costs (i.e., $v_2 = 0.9c_2$), whereas onsite service operators can only take over onsite jobs with lower revenue (on average 10 percent of the original job, i.e. $v_1 = 0.1c_1$). There is no penalty on remote services ($s_2 = 0$), and it is assumed that the shortage costs for onsite services are equal to the costs of the remote operator ($s_1 = c_2$). The demand per period for the services is assumed to be of a normal distribution with an equal average demand for both services ($\mu_j \epsilon [50; 50]$) but a different variation ($\sigma_j \epsilon [5; 20]$). Total capacity is represented by the current sum of all resources.

Summary of the Case Study Results. Based on the data provided by the firms, we can apply our model to different service settings and test substitutions among different resources. In so doing, we show that applying our model optimizes service productivity in each case study by determining the optimal level of input resources. This optimization generates total savings of between 4.8 percent and 15.5 percent (see Table 2). We find that the total savings are extraordinarily high in case studies with substitution among employees of different subsidiaries and departments. The same is true for the substitution of employees by technology. The lowest savings are generated in the case study focusing on third-party service providers. Generally, savings generated in our case studies can be explained by the effect of substitutions and by correctly accounting for demand volatility, thus lowering oversupply and shortage costs.

The insights gained in the five case studies reveal that the challenges of service productivity represent important issues. Firms must select resources from a mid- to long-term perspective and define the appropriate quantity of service when meeting varying customer demands. Choices are subject to capacity constraints. Each service setting analyzed faces a specific decision problem that can be solved using our proposed model. We further demonstrate the seminal potential of substitution effects and the importance of correctly accounting for demand to optimize service productivity. Although results confirm the appropriateness and functionality of our model in managerial practice to optimize service productivity, we cannot draw general conclusions or guarantee more general applicability. We apply our approach to simulated data and derive general rules to generalize our insights further and increase external validity. We generate several test cases and use different numerical analyses to prove the model's general applicability and generalize the findings.

Generalization of Findings with Simulated Data

Approach, Test Problems, and Data Applied. Each test instance consists of 1,000 examples with randomly generated parameters. The simulation is informed by the data constellation that we obtained from the case studies with the industry partners. The examples adhere to the following rules: Each resource $j, j \in \mathbb{N}$, has a positive profit ($r_j > c_j$) as well as a salvage value v_j and shortage costs s_j. The parameters for all items $j, j \in \mathbb{N}$, follow these boundaries: Revenues of $10 \geq r_j \geq 6$; unit costs of $r_j - 1 \geq c_j \geq 4$; positive salvage value at the end of period: $c_j - 0.1 \geq v_j \geq 3$; and positive shortage costs: $v_j \geq s_j \geq 1$. We assume unit costs are equivalent to the capacity coefficient b_j, representing a budget constraint. We apply normally distributed demand parameters with $D_j \sim N(\mu_j; \sigma_j)$; with $\sum_{j \in \mathbb{N}} \mu_j = N \bullet 10$; and with $\sum_{j \in \mathbb{N}} \sigma_j = N \bullet 5$; and identical positive substitution rate for all unavailable resources $i, i \in N^-$: $\delta_i = \delta = 0.6$. If the first

Table 2. Summary of Results Obtained From Case Studies (in % of Current Profit).

Productivity Issue Addressed	Service Setting	Total Savings, %	Savings Explained by: Impact Through Substitution Effects,[a] %	Impact of Correctly Accounting for Stochastic Demand,[b] %	Impact of Better Capacity Management,[c] %
Internal and external substitution of employees	Nursing	12.1	5.6	10.3	−0.8
	Aftersales	15.5	13.1	14.5	0.2
	Store replenishment	4.8	1.5	16.6	−22.7
Customer coproduction	Consulting	8.8	0.6	1.5	−20.0
Technology	Maintenance	13.4	1.2	12.7	−20.3

Please note that the explanations of the savings do not add up to the total savings as these are different ex-post analyses and the effects may mix up, balance, or even amplify each other.
[a]Ex-post comparison of model application with substitution ($\mu_{ji} > 0$) and without substitution ($\mu_{ji} = 0$) effects.
[b]Ex-post comparison of model application with variance ($\sigma^2 \geq 0$) and without variance ($\sigma^2 = 0$).
[c]Difference between underage and overage costs.

alternatives for i are also not selected (i.e., $j \in N^-$) or if the alternatives are selected but temporarily not available, the share of lost demand is accordingly higher than $1 - \delta_i$. Without any loss of generality, we assume that the fraction of consumers who are willing to substitute item i, δ_i is equally distributed to all other resources j: $\beta_{ij} = \delta_i/(N-1)$, $i \in N, j \in N, i \neq j$ and $\beta_{ii} = 0, i \in N$, in both cases of substitution. That also means $\beta_{ij}^{per} = \beta_{ij}^{temp}$.

The numerical tests investigate the effects of volatile demand (Test 1), substitution (Test 2), perishability of resources (Test 3), capacity constraints (Test 4), and joint resource selection and quantity (Test 5) on optimizing service productivity.

Test 1: Effect of Stochastic Demand. Figure 1 (Panel A) shows the impact of increasing stochastic demand (i.e., coefficients of variation, CV) on the solution structure considering different levels of substitution (δ). In the base scenario, demand is only slightly stochastic (CV = 1 percent). If substitution is low (δ = 20 percent or δ = 40 percent) and CV changes to 20 percent, we already find significant changes in the solution structure. At δ = 20 percent, each second resource, on average, has different optimal quantities. The more CV increases, the higher the share of resources with quantity changes. Considering higher substitution rates (δ = 60 percent or δ = 80 percent), we still find increasing CV to impact the solution structure significantly. For δ = 80 percent, an increase in CV to 20 percent changes the quantities of the resources on average by 59 percent. Again, the more CV increases, the higher the share of resources that have changes in quantities.

In summary, between 20 percent and 70 percent of the resources reveal different quantities when stochastic demand is considered correctly. These differences in optimal solutions always increase as CV increases for all δ-values. It also increases as δ increases when comparing the same CV values, except for very high CV. The latter effect is because the demand for one resource and its optimal quantities is determined more by the (high) substitution and less by its variation.

Figure 1 (Panel B) shows the impact on profits for the same setting. Here, a firm receives a profit advantage if it accounts for demand variation are considered correctly. Quantifying the profit advantage, we calculate (Profit of CSPM$_{SUB}$/Profit of CSPM*$_{SUB}$)-1, whereas CSPM*$_{SUB}$ denotes the solution obtained assuming CV = 1 percent, but where the profit is evaluated with the actual CV. Thus, it identifies the error firms make when not correctly accounting for the actual CV. Intuitively, we find that the more stochastic demand, the higher the advantage. For example, at δ = 20 percent, a firm's average profit advantage, if correctly accounted for CV = 20 percent instead of CV = 1 percent, is 2 percent. This advantage increases to 5 percent, if CV = 40 percent. Furthermore, we see profit advantages increase with a lower magnitude when substitution rates increase: In the case of δ = 80 percent and CV = 20 percent, the average advantage is 1.8 percent; if we further increase CV = 40 percent, the advantage is up to 3.8 percent. This can be explained by a buffering function of substitutions. Thus, substitutions work as a buffer if firms make suboptimal decisions on resource investments due to not correctly accounting for demand.

Our generalized data reveal that correctly considering stochastic demand results in 0.5 percent to 6.0 percent higher profits. If stochastic demand is considered correctly, profit differences increase with higher demand variation and lower substitution levels.

Test 2: Effect of Substitutions. To analyze the effect of substitution, we CSPM$_{SUB}$ considering substitution effects that are not reflected in CSPM. We add a posteriori substitution to the results of CSPM to ensure comparability in the event of substitutions. This is denoted by CSPM*, which is compared with CSPM$_{SUB}$. Figure 2 illustrates the change in required resources (left) and profit (right) when substitution effects are taken into account. This leads to significant profit improvements and lower resource requirements. In all cases, profit rises while the required resources fall for higher δ_i. Since resources with a higher profit margin benefit from the substitution demand of unselected resources, profit increases to 35 percent, and up to 35 percent fewer resources are assigned to the firm's portfolio (see Figure 2).

Test 3: Effect of Perishability of Resources. Service resources perish over time. To express the associated costs, we model shortage costs s and salvage values v. Precisely estimating those costs, in reality, is difficult, however, as shortage costs are opportunity costs for unsatisfied demand. Sales could be lost in future periods due to shortages, but it is almost impossible to calculate those effects exactly. Further, a firm's reputation is impaired by shortages, influencing demand. Salvage values v are also difficult to set as these are estimates for prices at the end of a period, for example, consumer willingness to pay in an alternative sales channel, at later periods or for different resources. To solve these estimation issues, ratios between over- and underestimation costs can be analyzed. Table 3 shows that the relationship of s_j to v_j for all $j \in \mathbb{N}$ has a moderate impact on profits and solution structure. The reference case is $s_j = v_j$. We find that decreasing s_j in relationship to v_j leads to increasing profits, while an increase in s_j compared to v_j leads to decreasing profits. Resource structure varies between 5 percent and -3 percent for changes of $\pm$ 30 percent of s_j against v_j. This variance results from the decision criteria applied: the higher the shortage costs, the more the underestimation is penalized; the quantities will increase, and the overall number of different resources will decrease. The situation is reversed for higher salvage values.

If either s_j or v_j, $j \in \mathbb{N}$ is set to zero, the profit and solution structures change significantly, as illustrated in Table 4. The analyses show a moderate impact of the relationship of s_j to v_j, $j \in \mathbb{N}$, whereas integrating one of these cost parameters significantly influences decisions and profit levels. If s = 0, lost profit for undershooting demand is lower, resulting in lower quantities and enabling a higher amount of different resources. If v = 0, overshooting of demand is less penalized, resulting in lower quantities and a higher amount of different resources.

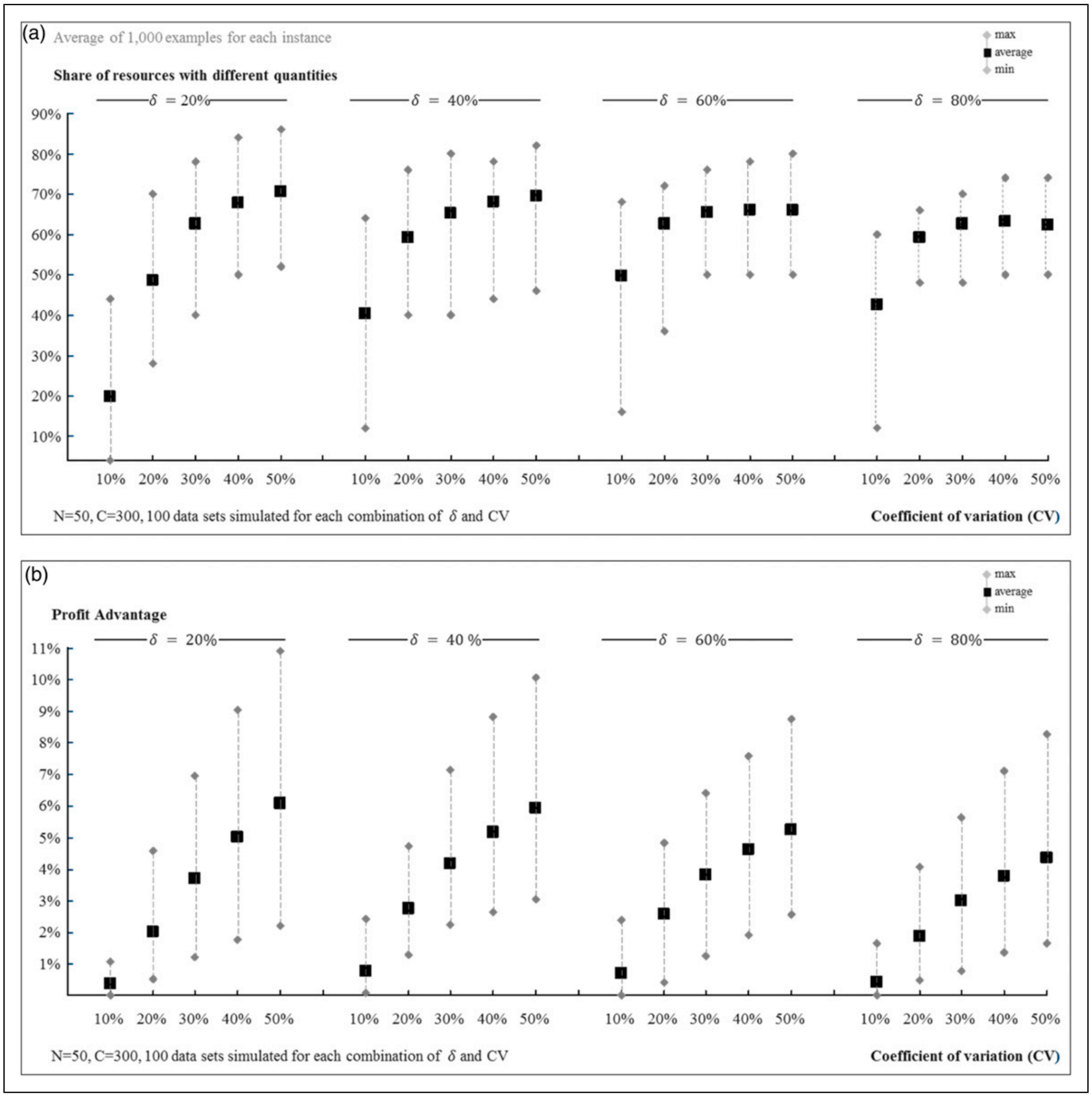

Figure 1. Panel A impact of stochastic demand and substitution level on solution structure. Panel B profit advantage of $CSPM_{SUB}$ over CSPM*SUB (assumed CV = 1 percent).

Test 4: Effect of Capacity Constraints. Total capacity constraint is a managerial decision determined by various overarching parameters. Table 5 shows that decisions regarding C significantly impact profit and solution structures. Reducing capacity leads to fewer resources and quantities per resource and a significant profit decrease. For example, an increase in capacity of 30 percent leads to a profit increase of 17 percent and an increase in different resources by 28 percent in our examples. However, the changes in capacities have a limited impact on the average quantity per resource. A capacity reduction of 30 percent only leads to an average reduction in quantities of less than 1 percent. On the contrary, a capacity increase of 30 percent leads to an increase in quantities of 1.2 percent. Therefore, changes in C influence the number of resources more than the quantities. This is driven by our decision calculus, which weighs over- and underestimation of demand. That means q_j is only increased until overestimation costs force a limit.

Test 5: Effect of Integrated Resource and Quantity Planning. $CSPM_{SUB}$ can be treated as an integrated capacity management problem with resource selection and quantity determination. Therefore, we evaluate the effect of $CSPM_{SUB}$ over a sequential planning (SP) approach whereby the resources are selected, and then the quantities are determined. Comparing $CSPM_{SUB}$ and SP, the integrated planning results in up to 13 percent higher profits on average and up to 23 percent fewer resources (see Figure 3). The left part of Figure 3 shows that resources required in $CSPM_{SUB}$ are smaller compared to SP for $\delta_i \geq 0.6$. With an increasing substitution level $\delta_i, i \in N^-$, the number of resources in the $CSPM_{SUB}$ decreases as more and

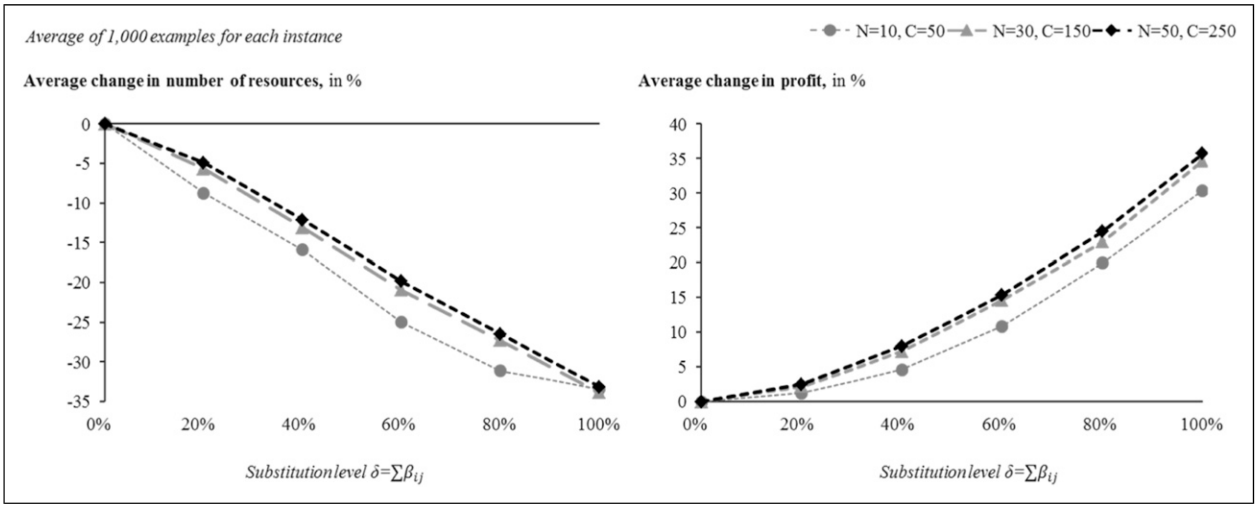

Figure 2. Change in required resources and profit when substitution effects are regarded.

Table 3. Change in Profits and Solution Structure due to Variations for $\mathbf{s_j}$ in Relationship to $\mathbf{v_j}$.

Average Changes	$\mathbf{s_j = 0.7v_j}$	$\mathbf{s_j = 0.8v_j}$	$\mathbf{s_j = 0.9v_j}$	$\mathbf{s_j = 1.1v_j}$	$\mathbf{s_j = 1.2v_j}$	$\mathbf{s_j = 1.3v_j}$
In profits	+3.1%	+1.8%	+1.0%	−1.2%	−2.1%	−3.0%
In number of resources	+4.7%	+1.9%	+1.1%	−1.1%	−2.1%	−3.1%

Note. Average of examples with N = 10, N = 30, and N = 50; reference $s_j = v_j$.

Table 4. Change in Profits, Quantity, and Number of Resources by Integrating $\mathbf{s_j = s}$ and $\mathbf{v_j = v}$, $\mathbf{j}\in\mathbb{N}$.

Average Change vs. v, s ≠ 0, v ≥ s	s = 0	v = 0
In profits	+12.0%	−21.0%
In quantity	−25.0%	−15.0%
In average number of resources	+33.0%	+19.0%

Note. Average of 1,000 examples with N = 10, N = 30, and N = 50.

more demand from low-profit resources is transferred to high-profit resources. This is because the higher the substitution level, the lower the lost sales. The right part of Figure 3 illustrates the average profit increase via integrated planning depending on $\delta_i = \delta, j \in N^-$. First, it is not surprising that the integrated CSPM_{SUB} achieves higher profits than SP. However, there are two contrary effects. On the one hand, the higher the lost sales (i.e., the lower the substitution rate), the more a firm may be penalized for suboptimal resource configurations. On the other hand, substitutions also work as a buffer for suboptimal decisions, as suboptimal resource and quantity decisions are still partially absorbed by substitutions. Thus, the profit delta between the CSPM_{SUB} and SP is the greatest when substitution rates are lower. This buffer diminishes with lower substitution rates; profit then deltas between CSPM_{SUB} and SP increase as substitution rates fall. As substitution effects further increase, additional profit from other resources continues to grow, which results in a convex coherence of profit increase and substitution level. This also means that the CSPM_{SUB} is more beneficial in settings with either very low or high substitution rates.

Discussion

Theoretical Implications

Managing productivity is a major challenge for service managers, as it requires an optimal selection from a limited and substitutable set of resources to meet productivity and profitability objectives. Technological advancements and shortage of labor call for an efficient engagement of resources to match the unknown customer demand, prompting calls from academia and managerial practice for comprehensive support systems (Mittal et al. 2005; Rust and Huang 2012; Rust, Zahorik, and Keiningham 1995; Wirtz and Zeithaml 2018). Building on the seminal work of Rust and Huang (2012), we develop such a decision support system and extend knowledge on service productivity. We demonstrate the functionality and applicability of the proposed decision model in multiple, distinct service settings and industries using five case studies. Moreover, to increase external validity, we draw on data simulation and generalize our insights through numerical examples.

First, we focus on optimizing the long-term resource selection and quantity determination as levers of service productivity management. The current literature examines how different resources—employees, customers, processes,

Table 5. Impact of Capacity Variations.

Average Change	70% C	80% C	90% C	100% C	110% C	120% C
In profits	−22.5%	−14.4%	−6.7%	+6.0%	+11.9%	+17.0%
In quantity	−29.6%	+19.2%	−9.8%	+9.6%	+19.2%	+28.4%
In average number of resources	−0.6%	−0.9%	−0.3%	+0.4%	+0.6%	+1.2%

Note. Average of 1,000 examples with N = 10, N = 30, and N = 50.

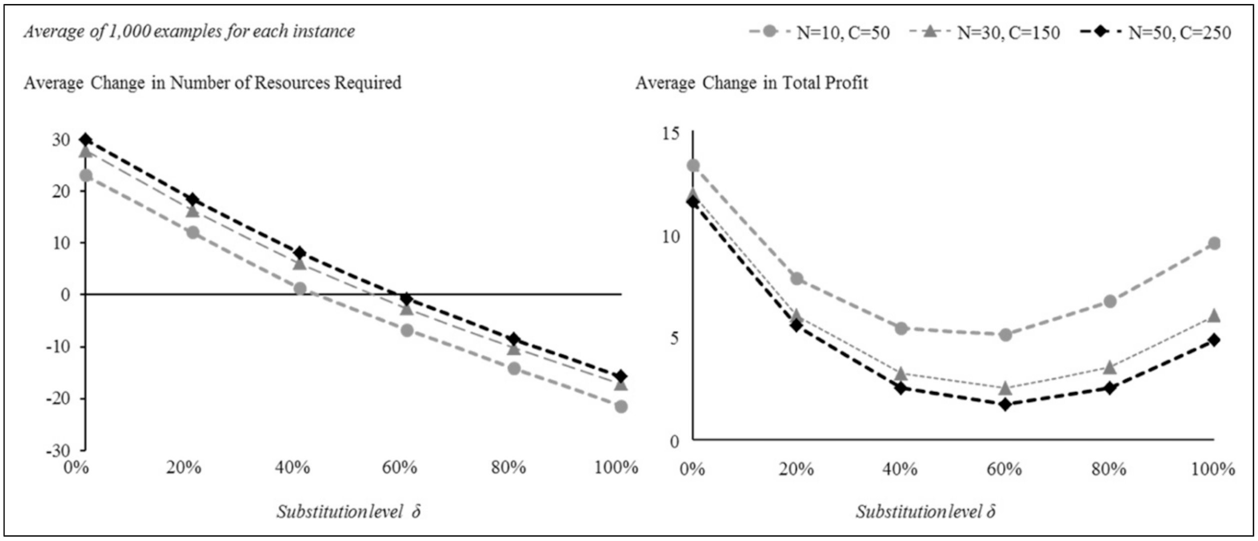

Figure 3. Change in required resources and profit when resources and quantities are integrated planned ($CSPM_{SUB}$ vs. SP).

and technology—influence how well services are delivered in terms of service productivity (e.g., Hofmeister, Kanbach, and Hogreve 2023a, 2023b; De Jong, De Ruyter, and Lemmink 2003; Marinova, Ye, and Singh 2008; Xue and Harker 2002). Rust and Huang (2012) show how resource investment decisions about either technology or labor as the major resource influence optimal service productivity. Yet, in managerial practice, firms usually apply multiple resources. Thus, we consider multiple resources within our proposed decision model, each generating revenues and costs. In so doing, we provide a decision support system that determines the optimal resource combination to invest in terms of type and quantity of resource available to optimize service productivity.

Our five case studies focus on internal and external substitution of employees, customer coproduction, and technology as major resources, revealing savings of up to 15.5 percent. Our generalized results demonstrate the importance of integrated resource and quantity planning compared to sequential planning (i.e., the selection of type comes before the determination of quantity). Our comprehensive approach results in up to 13 percent higher productivity and up to 23 percent lower number of resources needed. In addition, we reveal how applying our decision model makes resource planning more accurate and reduces the costs of over- and undersupply of resources. Furthermore, information gathered by a sensitivity analysis of the capacity constraint can be leveraged for hierarchical planning and be used to inform the overarching assignment of overall budgets per business unit that determine the overall capacities. Furthermore, our model can also be applied for an overarching network planning or for planning resources on a regional or governmental level.

Second, we are the first to consider substitution effects among resources available in managing service productivity. The literature and managerial practice prove that resources might be temporarily or permanently unavailable. To overcome resource shortages and/or demand peaks, substitutions among resources might work as a buffer; customers might settle for an alternative resource instead of switching the service provider. Customers who are partial employees in coproduction processes (Bowen 1986; Mills et al. 1983) or remote services using technology are ripe for substitutions (Rust and Huang 2012). Therefore, to get a finer-grained understanding of optimal resource allocation and, in turn, to optimize service productivity by either buffering exceeded demand or compensating exceeded capacities, we figure in permanent and temporary substitution.

Our case studies analyze substitution among employees of different qualification levels, distinct departments, and subsidiaries. We further consider substituting the internal and external workforce, the input of employees and customers (i.e., coproduction), and human workforce and technology. We find extraordinarily high savings for substitution among employees. The case studies and simulated data reveal the significance of substitution as a lever of service productivity. Considering substitution in capacity management results in lower resource requirements (up to 35 percent) and productivity improvements (up to 35 percent). Consequently, resources with higher profit margins benefit from substitution demand of unselected resources. We also examined how substitution among employees increases service productivity and discovered that considering substitution is especially relevant given stochastic demand. Substitutions are a buffer for non-optimal decisions (e.g., without a decision support system).

Third, we further enhance the deterministic decision model of Rust and Huang (2012) by considering stochastic demand that is more representative of actual customer behavior, especially for services. Demand and capacity cannot be backlogged, so capacity management must account for volatility (e.g., Armistead and Clark 1994; Armistead, Johnston, and Slack 1993; Chase and Apte 2007). Considering the volatility of demand, the applicability and usability of our model in managerial practice increases as costs for over- and undersupply of resources can be significantly reduced.

Our case studies and the generalization with simulated data further stress the importance of considering stochastic demand in service productivity management. Correctly accounting for demand volatility lowers the costs of oversupply and shortages and generates significant savings. Our data simulation further reveals that correctly considering demand increases profits by up to 6 percent. However, the solution structure significantly changes if demand volatility is not considered. In fact, the higher the demand volatility, the more it needs to be factored in to manage service productivity. Otherwise, up to 70 percent of resources may receive non-optimal levels.

Managerial Implications

Our research provides several insights on how to optimize service productivity successfully. We delineate and prove the relevance of three concrete challenges for service managers to consider while achieving cost efficiency and quality effectiveness. As outlined in further research, productivity in a service context goes beyond a ratio of input and output. Service productivity must be handled as a strategic decision variable enabling firms to choose their optimal service productivity level and optimize their profits. Service managers need to select the optimal type and quantity of multiple yet constraint resources while simultaneously considering the substitutability of resources. This decision problem is further increased as customer demand is highly volatile.

The insights gained in our studies allow for deriving several managerial implications in managing these three challenges. First, we showed that the optimal resource selection of multiple resources available in terms of type and quantity serves as a lever of success for service productivity. More concretely, we show that focusing on the right resource type before considering quantity aspects has a higher impact on service productivity. A comprehensive planning framework and hierarchy with feedback loops would thus facilitate decision-making accuracy. Consequently, we encourage managers to develop a comprehensive planning framework and hierarchy with iterative information flows. Doing so presupposes a clear understanding of the tasks to be done and customer expectations.

The second challenge formulated focuses on substituting effects among resources. We outline the outstanding importance of substituting resources in managing service productivity as it helps in overcoming resource shortages as well as it buffers not correctly accounted demand. Hereby, especially substituting among employees within a firm and substituting employees by technology shows high gains in service productivity. A further opportunity is to substitute resources within a network and coordinate the resource engagement across different entities (e.g., coordinating resources for health services of a region).

From a managerial perspective, these insights underscore the importance of enabling the substitution of these major resources. We assume that certain services or customer segments will accept substitution, so firms must carefully monitor their substituting activities and corresponding development of demand. We recommend that service managers invest in transparency and clarity in job requirements and customer expectations within a firm to enhance options for substitution among employees and other company entities. Furthermore, enabling job rotation, empowering employees in terms of self-organization, and investing in internal communication activities and knowledge transfer will further enhance the substitutability of employees. Doing so, our support system enables managers to optimize their resource investments and enhance flexibility. Regarding human resource management, we provide support for planning personnel deployment (within a department, across departments and locations), especially during peak times of demand or given personnel absence due to vacations or disease.

The third challenge signposts the importance of correctly accounting for customer demand, which is not always known beforehand. As outlined before, substituting effects among resources buffer wrong estimated demand and thus further outline the importance for service managers to consider our suggestions formulated above. Moreover, based on our findings, we strongly recommend that service managers consider stochastic demand to improve capacity management. Furthermore, investing in market research and monitoring demand will make capacity planning and decision-making more accurate.

Limitations and Directions for Further Research

Our theoretical decision support models significantly add to the research stream of service productivity. However, to provide a balanced discussion, we must recognize several limitations. Discussing these limitations raises avenues for further research. First, our insights are based on five case studies, their specific context (e.g., firm size, budget constraints, and employee qualification levels), and their current situation. We used simulated data to further generalizations the findings. Although we chose our case studies in an economic context that suffers from service productivity management like high coproduction, heterogeneity of the service outcome, and personnel-intense service environments (see Anderson, Fornell, and Rust 1997), service productivity might operate differently in other industries. Rust and Huang (2012) indicate that technology tremendously impacts optimal service productivity. Therefore, economic environments characterized by highly dynamic technological development might perform differently when it comes to service productivity. Also, no long-term effects of resources being unavailable are considered. We assume a constant service productivity among the workers and over time. Our model does not include incentives or other system changes (e.g., team structure, further technological support). We apply historical productivities and hence the average productivity of resources. We apply examples where substitutions among the team members are possible and necessary. We encourage researchers to conduct studies that take into account the longitudinal effects of resource optimization on service productivity. Doing so would require a multi-period model. For example, there might be long-term effects of substituting resources on demand and, in turn, on profitability. An extension could also include an analysis of seasonal effects or other variable demand patterns. Furthermore, as some of the data are based on expert estimations (e.g., substitution rates), objectivity might suffer. To further validate our approach, a more extensive dataset is needed.

Second, in formulating our substitution effects, we follow assumptions in ED models. The resulting model is cruder but has the advantage of being much easier to analyze and requires less data. However, more knowledge about substituting effects and their effects on productivity is needed.

Third, we develop a demand function wherein the demand depends on resources. We are aware, however, that there are further demand sources. For example, having higher quantities of one particular resource may also increase demand, for example, a certain type of service is offered in high quantity, thereby increasing demand due to higher visibility. Furthermore, the perceived quality of identical resource types might differ. We, therefore, encourage further research to explore additional opportunities to represent service quality and to show how service quality impacts optimal resource selection and service performance. We also assume that demand follows a normal distribution. Our modeling and solution approach is capable of coping with different demand distributions that include any positive demand. Further research is needed into how this impacts decisions and profits.

Finally, interdisciplinary research on service productivity is at an early stage. As discussed, using quantitative and modeling methods for service productivity-related problems can be beneficial. We would welcome more empirical research addressing the managerial challenges of service productivity from an interdisciplinary perspective.

Acknowledgements

Jens Hogreve and Mirjam Dobmeier thank the German Federal Ministry of Education and Research (grant 01FL12002) for the financial support.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Bundesministerium für Bildung und Forschung (grant 01FL12002).

ORCID iD

Jens Hogreve https://orcid.org/0000-0003-0997-7806

Supplemental Material

Supplemental material for this article is available online.

References

Anderson, E. W., Claes Fornell, and Roland T. Rust (1997), "Customer Satisfaction, Productivity, and Profitability: Differences between Goods and Services," *Marketing Science*, 13 (2), 129-45.

Andreassen, Tor W. (2021), "Service Productivity: A Call for Action!," https://www.servsig.org/wordpress/2021/05/service-productivity-a-call-for-action/?utm_source=ServSIG&utm_campaign=459aa234cf-SERVSIG_News_04_17_COPY_02&utm_medium=email&utm_term=0_001adc251a-459aa234cf-208495021. Accessed on March 10th 2023.

Armistead, Colin and Graham Clark (1994), "The 'Coping' Capacity Management Strategy in Services and the Influence on Quality Performance," *International Journal of Service Industry Management*, 5 (2), 5-22.

Armistead, Colin, Robert Johnston, and Nigel Slack (1993), "The Strategic Determinants of Service Productivity," *International Journal of Operations & Production Management*, 8 (3), 95-108.

Armstrong, J. Scott, Vicki G. Morwitz, and V. Kumar (2000), "Sales Forecasts for Existing Consumer Products and Services: Do Purchase Intentions Contribute to Accuracy?" *International Journal of Forecasting*, 16 (3), 383-97.

Auh, Seigyoung, Bulent Menguc, Constantine S. Katsikeas, and Yeon Sung Jung (2019), "When Does Customer Participation Matter? An Empirical Investigation of the Role of Customer Empowerment in the Customer Participation–Performance Link," *Journal of Marketing Research*, 56 (6), 1012-1033.

Bendapudi, Neeli and Robert P. Leone (2003), "Psychological Implications of Customer Participation in Co-Production," *Journal of Marketing*, 67 (1), 14-28.

Bhattarai, Abha and Maggie Penman (2023), "Restaurants Can't Find Workers Because They've Found Better Jobs," *The Washington Post*, https://www.washingtonpost.com/business/2023/02/03/worker-shortage-restaurants-hotels-economy/. Accessed on March 15th, 2023.

Bowen, David E. (1986), "Managing Customers as Human Resources in Service Organizations," *Human Resource Management*, 25 (3), 371-83.

Chase, Richard B. (1978), "Where Does the Customer Fit in a Service Operation," *Harvard Business Review*, 56 (6), 137-42.

Chase, Richard B. (1981), "The Customer Contact Approach to Services: Theoretical Basses and Practical Extensions," *Operations Research*, 29 (4), 698-706.

Chase, Richard B. and Uday M. Apte (2007), "A History of Research in Service Operations: What's the big Idea?" *Journal of Operations Management*, 25 (2), 375-86.

De Jong, Ad, Ko de Ruyter, and Jos Lemmink (2003), "The Adoption of Information Technology by Self-Managing Service Teams," *Journal of Service Research*, 6 (2), 168-79.

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

Dobni, Dawn J. R. (2004), "A Marketing-relevant Framework for Understanding Service Worker Productivity," *Journal of Services Marketing*, 18 (4), 303-17.

Frei, Frances X. and Patrick T. Harker (1999), "Measuring the Efficiency of Service Delivery Processes: An Application to Retail Banking," *Journal of Service Research*, 1 (4), 300-12.

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.

Haumann, Till, Pascal Güntürkün, Laura Marie Schons, and Jan Wieseke (2015), "Engaging Customers in Coproduction Processes: How Value-Enhancing and Intensity-Reducing Communication Strategies Motivate the Negative Effects of Coproduction Intensity," *Journal of Marketing*, 79 (6), 17-33.

Hofmeister, Johannes, Dominik K. Kanbach, and Jens Hogreve (2023a), "Measuring and Managing Service Productivity: A Meta-Analysis," *Review of Managerial Science*, Forthcoming.

Hofmeister, Johannes, Dominik K. Kanbach, and Jens Hogreve (2023b), "Service Productivity: A Systematic Review of a Dispersed Research Area," *Management Review Quarterly*, Forthcoming.

Hogreve, Jens and Andrea Beierlein (2023), "Value Creation and Cost Reduction in Health Care – Outcomes of Online Participation by Health Care Professionals," *Journal of Service Management*, 34 (3), 553-579.

Horowitz, Julia (2021), "Millions of Jobs and a Shortage of Applicants. Welcome to the New Economy," *CNN Business*, https://edition.cnn.com/2021/06/29/economy/global-worker-shortage-pandemic-brexit/index.html, Accessed on March 15th, 2023.

Huang, Ming-Hui and Roland Rust (2018), "Artificial Intelligence in Service," *Journal of Service Research*, 21 (2), 155-72.

Huang, Ming-Hui and Roland Rust (2021), "Engaged to a Robot? The Role of AI in Service," *Journal of Service Research,* 24 (1), 30–41.

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 (2), 28-45.

McLaughlin, Curtis P. and Sydney Coffey (1990), "Measuring Productivity in Services," *International Journal of Service Industry Management*, 1 (1), 46-64.

Melton, Horace L. and Michael D. Hartline (2013), "Employee Collaboration, Learning Orientation, and New Service Development Performance," *Journal of Service Research,* 16 (1), 67-81.

Meyer Goldstein, Susan (2003), "Employee Development: An Examination of Service Strategy in a High-Contact Service Environment," *Production and Operations Management*, 12 (2), 186-203.

Mills, Peter K. and J. H. Morris (1986), "Clients as "Partial" Employees of Service Organizations: Role Development in Client Participation," *Academy of Management Review*, 11 (4), 726-735.

Mills, Peter K., James L. Hall, Joel K. Leidecker, and Newton Margulies (1983), "Flexiform: A Model for Professional Service Organizations," *Academy of Management Review*, 8 (-January), 118-31.

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), 544-55.

Nachum, Lilach (1999), "The Productivity of Intangible Factors of Production: Some Measurement Issues Applied to Swedish Management Consulting Firms," *Journal of Service Research*, 2 (2), 123-37.

Nakata Cheryl and Hwang Jiyoug (2020). Design Thinking for Innovation: Composition, Consequence, and Contingency. *Journal of Business Research, 118*, (September), 117–128.

Rust, Roland T. and Tuck Siong Chung (2006), "Marketing Models of Service and Relationships," *Marketing Science*, 25 (6), 560-80.

Rust, Roland T. and Ming-Hui Huang (2012), "Optimizing Service Productivity," *Journal of Marketing*, 76 (2), 47-66.

Rust, Roland T., Christine Moorman, and Peter R. Dickson (2002), "Getting Return on Quality: Revenue Expansion, Cost Reduction, or Both?" *Journal of Marketing*, 66 (4), 7-24.

Rust, Roland T., Anthony J. Zahorik, and Timothy L. Keiningham (1995), "Return on Quality (ROQ): Making Service Quality Financially Accountable," *Journal of Marketing*, 59 (2), 58-70.

Singh, Jagdip (2000), "Performance Productivity and Quality of Frontline Employees in Service Organizations," *Journal of Marketing*, 64 (2), 15-34.

Wirtz, Jochen and Valarie A. Zeithaml (2018). "Cost-Effective Service Excellence," *Journal of the Academy of Marketing Science*, 46 (1), 59-80.

Xue, Mei and Patrick T. Harker (2002), "Customer Efficiency: Concept and its Impact on e-Business," *Journal of Service Research*, 4 (4), 253-67.

Zeithaml, Valarie A., Mary Jo Bitner, and Dwayne D. Gremler (2017), *Services Marketing: Integrating Customer Focus Across the Firm*, 7th ed. New York: McGraw-Hill.

Author Biographies

Jens Hogreve is Professor of Service Management at the Catholic University of Eichstaett-Ingolstadt, Germany. His research focuses on issues such as service recovery, future of mobility, transformative consumer research, and the consequences of digitalization on customers and employees. His research has been published in leading scholarly service as well as marketing journals.

Alexander Hübner is a Full Professor of Supply Chain Management at the Technical University of Munich, Germany. His research focuses on issues such as service operations management, retail operations, health care management, and logistics. His research has been published in leading service operations, logistics and operations management journals.

Mirjam Dobmeier is former PhD Candidate at the Department of Service Management at the Catholic University of Eichstaett-Ingolstadt, Germany. Her research focused on service productivity and service transformation. Since 2017, she is senior associate at the Department Innovation Management at AUDI AG, Germany.