Title: Stochastic models of resource allocation for services

Author: **Ralph Badinelli**,Professor, Virginia Tech, Department of BIT (0235), Virginia Tech, Blacksburg VA 24061, USA, ralphb@vt.edu Phone : (540) 231-7688, Fax : (540) 231-3752, Web page: http://www.bit.vt.edu/faculty/badinelli.html

Submitted to the EIASM 2009 NAPLES FORUM ON SERVICES

Author bio sketch:
Ralph Badinelli is a professor in the Department of Business Information Technology of the Pamplin College of Business of Virginia Tech. He received Ph.D. and M.S. degrees in Management from the Krannert Graduate School of Business of Purdue University, an M.S. degree in Physics from Purdue University, and a B.S. degree in Mathematics and Physics from Hofstra University.

Prior to pursuing an academic career, Dr. Badinelli worked as Production Planning Supervisor for the optics division of a west-coast laser manufacturer. He has consulted with a number of companies on planning, forecasting, and software development for decision support.

Dr. Badinelli's teaching responsibilities are in the areas of operations management and quantitative methods for business. These responsibilities extend over undergraduate, MBA and doctoral levels. He has taught many executive-development seminars on the subjects of project management, quality improvement, production and inventory control, and business strategy. He has received the R. B. Pamplin College of Business Certificate of Teaching Excellence.

Dr. Badinelli has published refereed articles across a wide range of topics including inventory control policies, revenue management, probability estimation, decision analysis and optimization. His publications have appeared in *Operations Research, Management Science, Decision Sciences, Naval Research Logistics, Computers and Operations Research, International Journal of Production Research, European Journal of Operational Research, and the Journal of the International Academy of Hospitality Research*. He is a member of the Institute for Operations Research and Management Science (INFORMS), Project Management Institute (PMI) and a certified member of the American Production and Inventory Control Society (APICS). He is the Vice-President/President-Elect of the INFORMS Service Science Section and a member of the SRII Leadership Team.

Abstract:

- Purpose: In this paper we develop a resource allocation model with general forms of service production functions, which describe the relationship between inputs and outputs of a process of co-creation of value by a service provider and a service recipient. The model development is directed at providing useful policy prescription for service providers and a foundation for research into the nature of resource allocation policies in service industries.
- Design/methodology/approach: The model development makes use of concepts of probability theory, optimization theory and extant DEA models.
- Findings: A practical optimization for allocating resources to service processes as well as insights into the complexity of service resource management are obtained.
- Research limitations/implications: The model presented in this paper is based on constant returns to scale of the service process.
- Originality/value: To date, service science lacks models for resource management that approach the usefulness of resource-management models for manufacturing enterprises even though the service economy in the industrialized world is larger than the manufacturing economy. This paper initiates a stream of model-building research.

Research Paper

Keywords: resource allocation, dispatching, service process

Stochastic models of resource allocation for services

1. Introduction

This research paper presents the development of a mathematical model for allocating and dispatching resources in a service enterprise. The starting point for this model is the definition of a service process. We define a service process as a coordinated set of activities which transforms a set of tangible and intangible resources (inputs), which include the contributions from the service recipient, into another set of tangible and intangible resources (outputs). Fitzsimmons (1985), Bettencourt et al. (2002), Lance et al. (2002), Sampson (2007) and others have firmly established the defining characteristic of service processes as the co-production of service outputs through the joint effort by the provider and recipient of the service. Tang & Zhou (2009) have refined this concept by emphasizing the coordination of these joint efforts, introducing the word, *taktchronicity*, to indicate the requirement of choreographed effort among multiple participants in the service process. Examples of the kinds of service enterprises that motivate this research are software development, consulting, education and project management.

Resource allocation is the managerial function of making available levels of capacity that can support planned operations. Resource dispatching is the managerial function of assigning resources to particular processes. In the case of service processes, the resources are provided by both the service provider and the service recipient. Furthermore, a typical service process requires several different types of inputs, such as labor, material, information, equipment, and produces several types of output, such as money, information and software. The relationship of the outputs of a process to the inputs is specified by a function that we call the service technology function.

Unlike manufacturing processes, service processes are generally not well understood by either the service provider or service recipient. In particular, high value-adding services, such as those that motivate this research, are rarely understood in enough detail to allow the publication of a process sheet that describes detailed steps in the process along with accurate resource requirements and cycle times for each step as one would find in a manufacturing environment. There are several sources of uncertainty for a typical service process. Among these are,

- Uncertainty of client commitment
- Uncertainty of client quality
- Uncertainty of the knowledge of the service process – estimation and specification of usage and yield rates
- Uncertainty of recognized uncontrollable factors which may cause changes to usage and yield rates

The thrust of this research is the development of mathematical decision models for resource management in service enterprises. The long-term goal of this stream of research, which is consistent with the exhortations in Chase and. Garvin (1989), Fitzsimmons and Fitzsimmons (2004), Spohrer et al. (2007) and Machuca et al. (2007), is the accomplishment of model-based decision support for service supply chains that is at least as sophisticated as that which is available to manufacturing enterprise.

The contribution of the research presented herein is three-fold:

1. A modeling framework for service processes that can serve as a foundation for further model development
2. A useful optimization model for resource allocation and dispatch
3. Some basic guidelines for optimal resource allocation/dispatching, for client involvement and adaptation of resource management to process learning

Resource-management modeling in service industries is in its infancy. Most of the effort in this field has been devoted to data envelopment analysis (DEA). DEA provides a typically macroscopic view of service processes and focuses on estimating and comparing the economic efficiency of services. See Charnes et. al. (1994), Fare and Grosskopf (2000) and Golany et. al.(2006) for perspective and overview of DEA. For the resource manager, a microscopic view is needed with the aim of determining the optimal assignment of resources to processes given their existing efficiencies. Few references are available for “shop-floor” resource management models. Korhonen and Syrjanen (2004) utilize a DEA model of service efficiency in order to determine directions for changing the resource allocations to the service in pursuit of multiple objectives. Their model is oriented towards a macro re-distribution of resources to an ongoing service enterprise. Gaimon (1997) takes a more process-level approach to setting workforce levels overtime for IT and knowledge workers. White and Badinelli (2009) extend this work to a workforce planning model in which client involvement at its effect on quality and efficiency are explicitly represented. All of these resource-planning models are deterministic.

In the current paper we model a single-stage service process that requires an arbitrary number of types of inputs and produces an arbitrary set of outputs. We achieve a higher level of realism than the contributions mentioned above because of the incorporation of uncertainty into the process model. Our approach is more specific than that provided by Bordoloi and Matsuo (2001), Carillo and Gaimon (2004), Napoleon and Gaimon (2004) and Dietrich (2006), and focuses on the scheduling problem of assigning resources to particular service processes as opposed to aggregate resource allocation.

We begin with a model of a technology function of a single-stage service process proposed by Athanossopoulus (1998), which was offered for a DEA study. A technology function for a service encounter is a function that effectively maps inputs to outputs according to the capabilities of the service participants to transform inputs into outputs. We construct this functional relationship by considering the inputs and outputs of a process to be functions of the volume, or number of service “cycles”, of the process which are simultaneously executed.

2. Technology functions

Different technology functions types can produce a given set of inputs and outputs. Efficiency is determined by a technology function. In service processes the combination of inputs which produces a specified set of outputs can vary. Hence, there may be more than one technology function type available to a service process. We posit the following principles for any technology function that can represent realistically a service process.

1. The service process has inputs from two sets – one controlled by the provider and one controlled by the client.

2. There are technology constraints in the form of standard usage and yield limits expressed in terms of quality units. The qualities of inputs and outputs are independent of the perceived utility of the inputs and output.

3. The technology constraints may not be fully known to the provider or the client, which makes the model stochastic.

4. The technology function is defined as the optimal solution to a game between the service provider and the service recipient. Different assumptions about knowledge sharing and power can produce different game equilibria.

5. There are constraints in terms of resource allocations that determine the resource units that are actually provided and the resource units that are actually yielded from a process.

6. Awareness – the client may not have full knowledge of the provider's resource commitments and the quality of the provider's resource commitments. Similarly, the providers may not have full knowledge of the client's resource commitments and the quality of the client's resource commitments.

7. The objective function that determines the optimal process inputs and outputs is the maximization of utility through a of the service participants.

Define,

$x_p = (x_{p1}, x_{p2}, ..., x_{pm})$ = a vector of quantities of the m input types required by process p

$y_p = (y_{p1}, y_{p2}, ..., y_{pn})$ = a vector of quantities of the n output types required by process p

In general, a technology function for any process, p, is a vector function that expresses the vector of outputs as a function of the inputs.

$$T_p\left(t, \{x_{pi}\}_{i \in S_{Ip}}\right) = \{y_{pj}\}_{j \in S_{Op}}$$

A linear, variable-returns-to-scale (VRS) technology function can be written,

$$T_p \; x_p + b = y_p$$

Where T_p is a matrix of constants.

A linear, constant-returns-to-scale (CRS) technology function can be written,

$$T_p x_p = y_p$$

The form of the technology function that we use for our model is a special case of linear, CRS function that can be found in Athanassopoulis (1998). We will refer to this type of function as a "recipe" function because it describes the relationship of inputs to outputs in terms of usage rates and yield rates of a process cycle. For a recipe-type technology function, the $x_i's$ must be procured according to usage rates of a process cycle and $y_j's$ are generated according to yield rates of a process cycle. The "recipe" for inputs and outputs pre process cycle forces all inputs to be in fixed proportions with respect to one another.

<u>The recipe technology function form</u>

Note: The word "benchmark" specifies an ideal process that is 100% efficient

$\mu_{pi} = \frac{1}{\beta_{pi}}$ = benchmark usage of resource i per cycle of process p

$\gamma_{pj} = \frac{1}{\alpha_{pj}} =$ benchmark generation (yield) of resource j per cycle of process p

$\beta_{pi} =$ benchmark technological coefficient of input i of process p (number of process cycles per unit of resource)

$\alpha_{pj} =$ benchmark technological coefficient of output j of process p (number of process cycles per unit of resource)

$v_p =$ volume of the process p execution. Can be thought of as the number of cycles of process p that are executed

$$v_p \mu_{pi} = x_{pi} \;, i = 1,...,m \tag{1}$$

$$v_p \gamma_{pj} = y_{pj} \;, j = 1,...,n \tag{2}$$

$$\Rightarrow v_p = \frac{y_{pj}}{\gamma_{pj}} = \frac{x_{pi}}{\mu_{pi}},$$

$$y_{pj} = \frac{\beta_{pi}}{\alpha_{pj}} x_{pi} = \frac{\gamma_{pj}}{\mu_{pi}} x_{pi} = \gamma_{pj} \beta_{pi} x_{pi} = \frac{x_{pi}}{\alpha_{pj} \mu_{pi}}, \tag{3}$$

Fixed proportions of inputs and outputs are inherent in these formulas for the elements of the technology function. For example,

$$\frac{x_{pi}}{x_{pk}} = \frac{\mu_{pi} v_p}{\mu_{pk} v_p} = \frac{\mu_{pi}}{\mu_{pk}} = \frac{\beta_{pk}}{\beta_{pi}}$$
$$\frac{y_{pj}}{y_{pk}} = \frac{\gamma_{pj} v_p}{\gamma_{pk} v_p} = \frac{\gamma_{pj}}{\gamma_{pk}} = \frac{\alpha_{pk}}{\alpha_{pj}}$$
$$\frac{y_{pj}}{x_{pi}} = \frac{\gamma_{pj} v_p}{\mu_{pi} v_p} = \frac{\gamma_{pj}}{\mu_{pi}} = \frac{\beta_{pi}}{\alpha_{pj}}$$

<u>A matrix form of the technology function</u>
The technology function, which expresses outputs as functions of inputs, can be written as a matrix multiplication as follows:

$$y_p = \tau_p x_p$$

where,

$$\tau_p = \frac{\gamma_p \beta_p^T}{m}$$

$$\tau_{pji} = \gamma_{pj} \beta_{pi} = \frac{\gamma_{pj}}{\mu_{pi}} = \frac{\beta_{pi}}{\alpha_{pj}} = \frac{1}{\alpha_{pj} \mu_{pi}}$$

The columns of the technology matrix are linearly dependent, each column being a multiple of a single, base column of technology coefficients. Similarly, the rows are linearly dependent. Note that this version of the matrix, is more constrained than that of the general linear CRS case. In the case of the process model the matrix is a dyad. The dyad can be built from the usage and yield parameters or from the technology coefficients.

From DEA we have the familiar general form of the relationship between inputs and outputs of an efficient CRS technology function, $u^T y_p = v^T x_p$. One may wonder about the correspondence of the recipe technology function to this relationship. The following theorem establishes the generality of the recipe model.

<u>Theorem 1:</u> A technology function implies the relationship, $u^T y_p = v^T x_p$, if and only if the technology function is a process-model function.

<u>Proof:</u>
Suppose $y_p = \tau_p x_p$

Multiplying both sides by the vector, α_p^T, we obtain

$$\alpha_p^T y_p = n \beta_p^T x_p$$

Now suppose, $u^T y_p = v^T x_p$, where $v = n\beta_p, u = \alpha_p$

Using the same definitions given in the development above, we can derive the result,

$$y_p = \alpha_p \gamma_p^T x_p$$ ||

Efficiency within the recipe model

We define the actual technology parameters $(b_p, u_p, a_p g_p)$ for a given instance of a process from the ideal parameters $(\beta_p, \mu_p, \alpha_p, \gamma_p)$ as follows:

$$u_p \geq \mu_p, g_p \leq \gamma_p$$
$$b_p \leq \beta_p, a_p \geq \alpha_p,$$

For the reader who has a background in DEA, we make one final connection of the current model to this large body of knowledge. DEA studies are directed at measuring the overall efficiency of a process. For the recipe model we can evaluate the difference between volume supported by inputs to volume required by outputs to measure inefficiency. Suppose, for a particular execution of a process,

$$\sum_{i=1}^{m} \beta_{pi} x_{pi} \geq \sum_{j=1}^{n} \alpha_{pj} y_{pj} \Rightarrow \sum_{i=1}^{m} \beta_{pi} x_{pi} - s = \sum_{j=1}^{n} \alpha_{pj} y_{pj}$$

The surplus variable s represents the overall level of inefficiency of the process. (See Athanassopoulos, 1998).

$s_p = (1 - e_p) \sum_{i=1}^{m} \beta_{pi} x_{pi}$, which effectively measures the amount of wasted input resources.The surplus of weighted inputs over weighted outputs represents the *overall* inefficiency of the process. If the efficiency is 100%, then $s_p = 0$.

$$e_p = 1 - \frac{s_p}{\sum_i \beta_{pi} x_{pi}}$$

For resource allocation and dispatching, we must consider efficiency as a multi-dimensional quantity. That is, the effectiveness of each input resource on a process can be different and the productivity of the process in terms of each output resource can be different. We view inefficiency in terms of deviations from the benchmark recipe. Hence, efficiency has a multi-dimensional foundation in the technology function. For resource planning purposes, we need to measure the components of the efficiency in order to know which input resources should be adjusted.

Randomness and inefficiency in the recipe model

Inefficiency can be due to systemic shortcomings in the DMU as well as random variations in performance. There is a parallel to the notions of common vs. specific causes of variation in modeling of manufacturing processes.

We can represent the inherent inefficiency of a particular service encounter as well as the inherent uncertainty in the performance of the service process by defining non-negative random variables to represent the deviation of the process parameters from their ideal, deterministic values.

$$u_p = \mu_p + \delta_{pu}$$

$$g_p = \gamma_p - \delta_{pg}$$

$$b_p = \beta_p - \delta_{bp}$$

$$a_p = \alpha_p + \delta_{pa}$$

where each delta random variable has non-negative support. These expressions define the process parameters, $(b_p, u_p, a_p g_p)$, as random variables.

We assume that the inputs to the service process are controllable and inputs are dispatched prior to the realization of random variations of process parameters. Outputs, on the other hand, are produced through the realization of process parameters. Hence, the outputs, $y_{pj}, j = 1,...,n$, are also random variables.

Since,

$$b_{pi} x_{pi} = (\beta_{pi} - \delta_{pb}) x_{pi} = v_p \text{ for } i = 1,2,...,m.$$

the random variables, $b_{p1}, b_{p2}, ..., b_{pm}$, must be perfectly correlated. That is, after the process inputs are dispatched, the process usage rates mutually adjust to values that support a certain volume and which are consistent with the inefficiencies and random variations of the usages.

Define, $\nu_p = \beta_{pi} x_{pi}$ and $\varepsilon_p = \dfrac{v_p}{\nu_p} \le 1$

Then,

$$b_{pi} x_{pi} = \varepsilon_p \nu_p \tag{4}$$

Each random variable, $b_{pi} x_{pi}$, can be replaced by the random variable ν_p Hence, the m decision variables, $x_{p1}, ..., x_{pm}$ can be replaced with a single decision variable, ν_p.

3. Optimization

The criteria for the resource dispatching decision can be stated as follows:

<u>Problem P1</u>

$$\min_{\{x_p\}} \sum_p \sum_j w_{pj} \int_0^{\hat{y}_{pj}} (\hat{y}_{pj} - y) f_{y_{pj}}(y)\,dy + c^T \sum_p x_p$$

subject to:

$$r - \sum_p x_p \ge 0$$

$x_p \geq 0$ for all p

$\hat{y}_p =$ a vector of target outputs for process p

$f_{y_{pj}} =$ the distribution of y_{pj}, which is a function of the resource allocations, x_p

$r =$ vector of capacities of available resources

The Loss Function

The portion of the objective function that captures the effects of failing to produce enough outputs to reach desired targets is called the loss function.

$$\sum_p \sum_j w_{pj} \int_0^{\hat{y}_{pj}} (\hat{y}_{pj} - y) f_{y_{pj}}(y) dy \qquad (5)$$

The loss function in the objective function makes the optimization a rather complex NLP.

Lemma 1: The loss function increases with inefficiency

Proof: The proof is established most easily from the form of the loss function, (5). Inefficiency is reflected in the probability distribution for y_{pj}.

If we consider any two service processes such that the second process has less efficiency than the first, then the probability distribution for y_{2j} stochastically dominates the probability distribution for y_{1j}.

$$\int_0^{\hat{y}_{1j}} (\hat{y}_{1j} - y) f_{y_{1j}}(y) dy \geq \int_0^{\hat{y}_{2j}} (\hat{y}_{2j} - y) f_{y_{2j}}(y) dy \qquad \|$$

Lemma 2: Loss is increasing in the targets, $\hat{y}_{pj}$

Proof: Obvious from original form of the loss function, (5).

The Loss Function in terms of technology parameters

By expressing x_p, y_p in terms of their dependence on the volume through $x_p = \mu_p \nu_p$, $y_p = g_p \varepsilon_p \nu_p$, we can reduce the set of decision variables from $\{x_{p1}, x_{p2}, ..., x_{pm}\}$ to the single variable, ν_p. Furthermore, we introduce the probability distribution of the random parameters, $g_p \varepsilon_p$, which allows us to re-state the problem more directly in terms of the sources of randomness. Define,

$z_{pj} = g_{pj}\varepsilon_p$

Then, $F_{y_{pj}}(y) = F_{z_{pj}}\left(\frac{y}{\nu_p}\right)$ and $f_{y_{pj}}(y) = \frac{1}{\nu_p} f_{z_{pj}}\left(\frac{y}{\nu_p}\right)$

$$\int_0^{\hat{y}_{pj}} \left(\hat{y}_{pj} - y\right) f_{y_{pj}}(y)dy = \int_0^{\hat{y}_{pj}} \left(\frac{\hat{y}_{pj} - y}{\nu_p}\right) f_{g_{pj}\varepsilon_p}\left(\frac{y}{\nu_p}\right)dy = \nu_p \int_0^{\hat{z}_{pj}} \left(\hat{z}_{pj} - z\right) f_{z_{pj}}(z)dz$$

Where $\hat{z}_{pj} = {}^{\hat{y}_{pj}}\!\big/_{\nu_p}$

<u>Problem P2</u>

The resource dispatching problem becomes,

$$\min_{\{\nu_p\}} \sum_p \sum_j w_{pj}\nu_p \int_0^{\hat{z}_{pj}} \left(\hat{z}_{pj} - z\right) f_{z_{pj}}(z)dz + c^T \sum_p \mu_p \nu_p$$

subject to:

$$r - \sum_p \mu_p \nu_p \geq 0$$

$$\nu_p \geq 0 \text{ for all } p$$

The loss function now can be manipulated into a form that is more convenient for computation and further analysis.

$$\int_0^{\hat{z}_{pj}} \left(\hat{z}_{pj} - z\right) f_{z_{pj}}(z)dz = \hat{z}_{pj} F_{z_{pj}}(\hat{z}_{pj}) - \left(zF_{z_{pj}}(z)\right)\Big|_0^{\hat{z}_{pj}} + \int_0^{\hat{z}_{pj}} F_{z_{pj}}(z)dz$$

$$= \int_0^{\hat{z}_{pj}} F_{z_{pj}}(z)dz$$

Define,

$$G_{z_{pj}}(z) = 1 - F_{z_{pj}}(z) = \int_z^{\infty} f_{z_{pj}}(s)ds$$

$$A_{z_{pj}}(z) = \int_z^{\infty} G_{z_{pj}}(s)ds$$

$$B_{z_{pj}}(z) = \int_z^{\infty} A_{z_{pj}}(s)ds$$

<u>Lemma 3:</u> $A_{z_{pj}}(0) = E\left\lfloor z_{pj} \right\rfloor$

<u>Proof:</u> See Hadley-Whitin (1963)

Denote $E\left\lfloor z_{pj} \right\rfloor = \bar{z}_{pj}$. The loss function can be written,

$$\sum_p \sum_j w_{pj} v_p \left(\hat{z}_{pj} - \bar{z}_{pj} + A_{z_{pj}}(\hat{z}_{pj}) \right) \tag{6}$$

<u>Lemma 4:</u> Loss is decreasing and convex in volume
<u>Proof:</u>
Recall that $\hat{z}_{pj} = \hat{y}_{pj} / v_p$

$$\frac{d}{dv_p}\left(w_{pj} v_p \left(\hat{z}_{pj} - A_{z_{pj}}(0) + A_{z_{pj}}(\hat{z}_{pj}) \right) \right)$$

$$= w_{pj}\left(-A_{z_{pj}}(0) + A_{z_{pj}}(\hat{z}_{pj}) \right) + w_{pj} \hat{z}_{pj} G_{z_{pj}}(\hat{z}_{pj})$$

Since $A_{z_{pj}}$ is decreasing and convex, and $\dfrac{dA_{z_{pj}}}{dz} = -G_{z_{pj}}$,

$$-A_{z_{pj}}(0) + A_{z_{pj}}(\hat{z}_{pj}) < -\hat{z}_{pj} G_{pj}(\hat{z}_{pj})$$

Therefore the loss function is decreasing in volume.

$$\frac{d^2}{dv_p^2}\left(w_{pj} v_p \left(\hat{z}_{pj} - A_{z_{pj}}(0) + A_{z_{pj}}(\hat{z}_{pj}) \right) \right)$$

$$= \frac{d}{dv_p}\left(w_{pj}\left(-A_{z_{pj}}(0) + A_{z_{pj}}(\hat{z}_{pj}) \right) + w_{pj} \hat{z}_{pj} G_{z_{pj}}(\hat{z}_{pj}) \right)$$

$$w_{pj} \frac{\hat{z}_{pj}}{v_p} G_{z_{pj}}(\hat{z}_{pj}) - w_{pj} \frac{\hat{z}_{pj}}{v_p} G_{z_{pj}}(\hat{z}_{pj}) + w_{pj} \frac{\hat{z}^2}{v_p} f_{z_{pj}}(\hat{z}_{pj}) > 0$$

Therefore the Loss function is convex in volume. ||

<u>KKT conditions</u>
The lagrangian for Problem P2 is,

$$L(v,\lambda) = \sum_p \sum_j w_{pj} v_p \left(\hat{z}_{pj} - \bar{z}_{pj} + A_{z_{pj}}(\hat{z}_{pj}) \right) + c^T \sum_p \mu_p v_p + \lambda^T \left(\sum_p \mu_p v_p - r \right)$$

The convexity of the loss function ensures a unique solution. The first-order KKT conditions for the optimization problem include,

$$\frac{\partial L}{\partial v_p} = 0 = \sum_j w_{pj}\left(-\bar{z}_{pj} + A_{z_{pj}}(\hat{z}_{pj}) \right) + w_{pj} \hat{z}_{pj} G_{z_{pj}}(\hat{z}_{pj}) + \left(c^T + \lambda^T \right) \mu_p \tag{7}$$

The derivative of the loss function in (7) represents the marginal value of one unit of volume to the achievement of target levels of output. We denote this marginal value, $M_p(v_p)$.

$$M_p(v_p) = -\left(\sum_j w_{pj}\left(-\bar{z}_{pj} + A_{z_{pj}}(\hat{z}_{pj}) + \hat{z}_{pj} G_{z_{pj}}(\hat{z}_{pj}) \right) \right)$$

Then (30) states,

$$M_p(\nu_p^*) = \left(c^T + \lambda^T\right)\mu_p \tag{8}$$

The right-hand side of (8) is the marginal resource cost of a unit of volume. Note that, by Lemma 4, $M_p(\nu_p)$ is positive and decreasing ensuring a unique solution to (8). Standard NLP methods such as KNITRO (see Byrd, et al., 2006) can be used to find the optimal volumes of all available processes and, hence, the optimal dispatch of input resources to the processes.

Theorem 2: Processes that have lower usage rates will be allocated higher proportions of available input resources and achieve higher volumes under an optimal policy.

Proof: From the definition of ν_p

$$x_{pi}^* = \beta_{pi}\nu_p^*, \frac{x_{pi}^*}{x_p^*} = \frac{\beta_{pi}}{\sum_i \beta_{pi}} \qquad ||$$

Allocation of resources across processes

The optimality conditions indicate that the optimal allocation of input resources across processes that are different in terms of their efficiencies, uncertainties and/or output targets is quite complex and, in some cases counter-intuitive. The behavior of $M_p(\nu_p)$ as these characteristics of a process change is not uni-directional. Furthermore, the fact that $M_p(\nu_p)$ and the optimality condition (8) are based on vector inner products, there is the possibility that different process can differ in one direction on some dimensions and in another direction on other dimensions. Therefore, simplistic guidelines for dispatching resources across different processes are not to be expected.

Conflict between service providers and service recipients

The resources that are dispatched as inputs to a process are provided by both the service recipient and the service provider. Similarly, the benefits of the outputs of a service encounter are enjoyed by both parties, however, not necessarily in the same way. Specifically, the weights, w_{pj}, and the targets, $\hat{y}_{pj}$ may be different for the two parties. Clearly, such differences will lead each party to arrive at a different optimal policy.

Another opportunity for a conflict in the support of a process provided by the service provider and the service recipient stems from their differences in the probability distribution of the process parameters, z_p. Clearly, any difference in the specification and estimation of these distributions will manifest themselves in differences in $-\bar{z}_{pj} + A_{z_{pj}}(\hat{z}_{pj}) + \hat{z}_{pj}G_{z_{pj}}(\hat{z}_{pj})$, which would imply different solutions for (8).

Solutions to Problem P2 for different parameter values and probability distributions reveal that the magnitude of differences in service-process support and expectations of outcomes are bound to create mis-understandings and conflicts between providers and recipients that are well-known in service enterprises such as consulting and education.

A clear recommendation can be derived from this research. Service providers and service recipients should make every attempt to educate themselves jointly about the nature of a service process before they engage in dispatching resources to it. This joint understanding of a process has the benefits of minimizing the uncertainty about the process and, perhaps more importantly, create a consensus about specification and estimation of the process technology. Furthermore, we can recommend that, having achieved this consensus, the solution to Problem P2 be computed by both parties so that the optimal plan for the service encounter is consistently executed.

4. Conclusion

This paper provides a model for resource dispatching to specific service encounters within a service enterprise. The optimization is straightforward and achievable by standard NLP software. Therefore, the model has practical utility.

The model presented herein is a stochastic model of a service process which has the potential to capture randomness of service processes as well as uncertainty in the specification and estimation of process parameters by the process participants. Consequently, the data collection and parameter estimation that is necessary to apply this model necessarily involves assessing the extent of the knowledge of the process by the service provider and the service recipient. The model allows the investigation of the sensitivity of the policy for resource dispatch to process knowledge and the asymmetry of these policies across participants who differ in their understanding of the service process. This type of investigation holds the promise of bringing service providers and service recipients towards common ground in planning service processes.

Future research will expand this model into a multi-stage framework and the investigation of two-party games as a means to describe the interplay of a service provider and a service recipient in their decisions to dispatch resources to a service process.

References

Athanassopoulos, A. (1998). “Decision support for target-based resource allocation of public services in multiunit and multilevel systems”, Management Science 44(2), 173-187.

Bettencourt, L. A., A. L. Ostrom, et al. (2002). "Client Co-Production in Knowledge-Intensive Business Services." *California Management Review* 44(4): 100-128.

Bordoloi, S. and H. Matsuo (2001). "Human resource planning in knowledge-intensive operations: A model for learning with stochastic turn-over." *European Journal of Operational Research* 130: 169-189.

Bowen, D. E. (1986). "Managing Customers as Human Resources in Service Organizations." *Human Resource Management (1986-1998)* 25(3): 371.

Byrd, R. H., J. Nocedal, et al. (2006). KNITRO: An Integrated Package for Nonlinear Optimization. *Large-Scale Nonlinear Optimization*. D. Pillo, Gianni, Roma and Massimo, Springer-Verlag. 83**:** 35-59.

Carrillo, J. E. and C. Gaimon (2004). "Managing Knowledge-Based Resource Capabilites Under Uncertainty." *Management Science* 50(11).

Charnes, A., Cooper, W., Lewin, A., Seiford, L. (1994). Data envelopment analysis – theory, methodology and applications, Kluwer.

Chase, R. B. and D. A. Garvin (1989). "The Service Factory." *Harvard Business Review* 61-66.

Dietrich, B. (2006). "Resource Planning for Business Services." *Communication of the ACM* 49(7): 62-64.

Fare, R., Grosskopf, S. (2000). “Network DEA”, Socio-economic planning sciences 34, 35-49.

Fitzsimmons, J. A. (1985). "Consumer Participation and Productivity in Service Operations." *Interfaces* 15(3): 60.

Fitzsimmons, J. A. and M. J. Fitzsimmons (2004). Service Management. New York, McGraw-Hill.

Gaimon, C. (1997). "Planning Information Technology-Knowledge Worker Systems." *Management Science* 43(9).

Golany, B., Hackman, S., Passy, U. (2006). “An efficiency measurement framework for multistage production systems”, *Annals of Operations Research*, 145, 51-68.

Korhonen, P., Syrjanen, M. (2004). “Resource allocation based on efficiency analysis”, *Management Science* 50(8), 1134-1144.

Lance, A. B., L. O. Amy, et al. (2002). "Client co-production in knowledge-intensive business services." *California Management Review* 44(4): 100.

Machuca, J. A. D., M. d. M. González-Zamora, et al. (2007). "Service Operations Management research." *Journal of Operations Management* 25(3): 585.

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

Sampson, S. E. (2007). *A Customer-Supplier Paradigm for Service Science*. 2007 DSI Services Science Miniconference, Pittsburgh.

Spohrer, J., P. Maglio, et al. (2007). “Steps Toward a Science of Service Systems,“ *Computer* **,** 71-77.

Tang, V., Zhou, R. (2009). “First-Principles for services and product-service-systems: an R&D agenda”, International Conference on Engineering Design, ICED'09, 24 - 27 August 2009, Stanford University, Stanford, CA, USA.

White, S., Badinelli, R. (2009). Aggregate planning in complex, high-contact services, Working Paper, Virginia Tech, Blacksburg, VA, USA, June.