RC25238 (W1111-130) November 21, 2011
Mathematics

IBM Research Report

Integrated Stochastic Resource Planning of Human Capital Supply Chains

Heng Cao, Jianying Hu, Chen Jiang, Tarun Kumar, Ta-Hsin Li, Yang Liu, Yingdong Lu, Shilpa Mahatma, Aleksandra Mojsilovic, Mayank Sharma, Mark S. Squillante, Yichong Yu

IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Integrated Stochastic Resource Planning of Human Capital Supply Chains

Heng Cao, Jianying Hu, Chen Jiang, Tarun Kumar, Ta-Hsin Li, Yang Liu, Yingdong Lu, Shilpa Mahatma, Aleksandra Mojsilović, Mayank Sharma, Mark S. Squillante, Yichong Yu

Business Analytics and Mathematical Sciences Department

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

{hengcao@us.ibm.com,jyhu@us.ibm.com,chjiang@us.ibm.com,ktarun@us.ibm.com,thl@us.ibm.com,jyliu@us.ibm.com,yingdong@us.ibm.com, mahatma@us.ibm.com,aleksand@us.ibm.com,mxsharma@us.ibm.com,mss@watson.ibm.com,yichong@us.ibm.com}

In this paper, we present a suite of innovative operations research models and methods, called OnTheMark (OTM); this suite supports the effective management of human capital supply chains by addressing distinct features of human talent, which cannot be handled via traditional supply chain management. OTM consists of novel solutions for: (1) statistical forecasting of demand and human capital requirements; (2) risk-based stochastic human-talent capacity planning; (3) stochastic modeling and optimization (control) of human capital supply evolutionary dynamics over time; (4) optimal multiskill supply-demand matching; and (5) stochastic optimization of business decisions and investments to manage human capital shortages and overages. The OTM suite was developed and deployed as an important part of the human capital management and planning process within IBM, providing support for decision making to drive better business performance. This is achieved through important contributions in the areas of stochastic models and optimization (control), and the innovative application and integration of these models and methods in human capital management applications.

Key words: human capital management and planning; demand forecasting; risk-based capacity planning; stochastic supply evolution; multiskill supply-demand matching; stochastic models; stochastic optimization and control.

Introduction

The services sector has grown over the last 50 years to dominate economic activity in most advanced industrialized countries. In fact, recent Bureau of Labor Statistics data reveal that more than 75% of the labor force in the U.S.A. is employed in the services industry, with increasing projections into the future; and that services industry output represents 70% of the total industry output (in terms of U.S. dollars) Woods (2009). A critical driver of success for any service delivery organization is its ability to manage and deploy the skills, knowledge and competencies of its human resources – or less formally, “having the right people, with the right skills, in the right place, at the right time”. Successful organizations realize that investments in their people, including proper support for personal needs and career goals, are a key driver behind growth, profitability and client satisfaction. This is especially true for providers of information technology services like IBM who offer a broad range of service products, each requiring human resources with certain skills, in markets characterized by highly volatile and uncertain client demands. As a result, forward-thinking businesses are beginning to invest in human capital supply chain methodologies as a major competitive differentiator.

Over the past decades, organizations have achieved significant gains in effectiveness and efficiency by developing advanced models of manufacturing and logistics systems to optimize their supply chain operations. Although such traditional supply chain models & methods have a rich history within the operations research (OR) community, they cannot be directly applied to related problems in human capital supply chains. This is because human resources are quite distinct from the inanimate machine parts in traditional supply chains, and human resources must be managed in a manner that properly addresses these distinct features and personal needs. Human resources are more complex to model than traditional supply chain parts, and thus require new models & methods to capture and represent these complexities. Concepts such as hiring, retaining and acquiring new skills, fundamentally influence certain characteristics of the optimization problems relevant to human capital management, making them more difficult to model and solve, but of critical importance to the successful management and planning of human capital supply chains in practice.

A prime example of the importance of effective human capital management is *Integrated Technology Services* (ITS), a business line within IBM Global Services. ITS is the premier integrator of information technologies in the services industry, with annual revenues of approximately $4 billion. Through its ten Service Product Lines, ITS delivers a range of integration and support service products, henceforth called *service engagements* (such as the design and implementation of a data warehouse), for all IBM hardware and software products and beyond. The ITS business has a focus on human resources, with the vast majority of costs for most ITS service engagements consisting of human services as opposed to those related to hardware and software. Unlike machine parts in traditional supply chains, the human resources of the ITS supply chain are not consumed during the service delivery process as they return upon completion of an engagement, and thus represent a long-term capital investment. In addition, human resource productivity and efficiency are critical factors that depend upon workload and utilization, whereas notions of productivity and efficiency for machine parts is less sensitive to workload and utilization. Human resources also can evolve in various complex ways over time, and are capable of deploying more than one skill at the same time and across multiple service delivery engagements. Service delivery is further complicated by uncertainty in resource staffing, simultaneous allocation of multiple resources (as engagements require different skills), and resource sharing (as human resources time-share their skills across different engagements).

The ability to manage and plan human resources more effectively and efficiently, while addressing these fundamental differences between human capital and traditional supply chains, is pivotal to ITS delivery of service engagements. Looking for novel solutions to help optimize its human capital supply chain, in 2007

ITS turned to IBM Research and initiated the development of OnTheMark (OTM), an end-to-end human capital management & planning framework and methodology. OTM is comprised of innovative OR models & methods that address the many challenges and complexities involved in the ITS human capital supply chain. Specifically, we have developed capabilities to forecast demand for service engagements and their resource requirements applying statistical and machine learning methods. We have derived new stochastic modeling and optimization methods to provide a form of risk-based capacity planning that determines the resource capacity levels to maximize business performance given forecasted demand. We have also developed a novel approach to estimate available human resources in the future (namely, existing resources and those obtained/retained via various business decisions and investments) and their skill composition through the stochastic modeling and optimization/control of supply evolution over time. We then determine an optimal matching of the future estimates of multi-skill human resource obtained from our supply evolution models & methods against the resource capacity targets obtained from our risk-based capacity planning models & methods. Finally, we optimize available business decisions and investments (e.g., hiring, training, retaining) to address the resource shortages and overages estimated via this optimal multi-skill supply-demand matching through a combination of our innovative mathematical modeling and optimization solutions.

OTM has been deployed since 2008 and has become an essential part of the ITS human capital management & planning process. OTM has helped ITS improve its management & planning of human resources and skill composition as a basis for driving growth and profitability, and has helped to realize a major competitive differentiation in the marketplace. In fact, based on our successful experience through the past couple of years with quarterly ITS executive reviews worldwide, the OTM solution has consistently provided relative revenue-cost benefits commensurate with 2-4% of ITS quarterly revenue targets over previously employed approaches. Moreover, beyond its internal application on the ITS business, we have developed prototype implementations of several OTM models & methods within IBM software platforms to support future IBM solution offerings (both products and services) in the marketplace. We are also collaborating with IBM product groups on the development of pilot solutions for a broad collection of clients that is based fully on the OTM models & methods. The possibilities for using the OTM framework and methodology in a broader application setting, beyond information technology services, are numerous and actively being pursued.

This paper presents an overview of the OTM human capital management & planning framework and methodology, focusing primarily on the innovative mathematical modeling and optimization contributions of the project. A high-level overview of the OTM solution methodology is provided first, followed by a more detailed discussion of each set of OR models & methods. We conclude with a summary of our successful worldwide deployment of the integrated OTM suite within ITS, including validation studies and business cases based on ITS data, and business impact on ITS and beyond. Throughout, we shall use the term *skill* to generally refer to the full extent of human resource abilities, competencies, knowledge and skills. Examples include database architect, Java programmer and network specialist.

Overview of Human Capital Management and Planning Solution

To address the distinct challenges and complexities of human capital supply chains in general, and particularly those arising within the ITS business, we developed OR modeling and optimization solutions with the overall objective of determining the best resource capacity levels and investment decisions over time to maximize business performance. While various investment decisions are available, in this paper we focus on hiring, training, promotions and retention (incentivizing to reduce attrition) as representative decisions available to ITS organizations. The OR solutions have been implemented as core capabilities & components of the OTM human capital management & planning suite, which supports an end-to-end process for human capital supply chains. OTM consists of: (I) *demand forecasting*; (II) *risk-based capacity planning*; (III) *supply evolution & optimization*; (IV) *multi-skill shortage & overage optimization*; and (V) *skill shortage & overage management*. Figure 1 illustrates these primary OTM elements and their interrelationships.

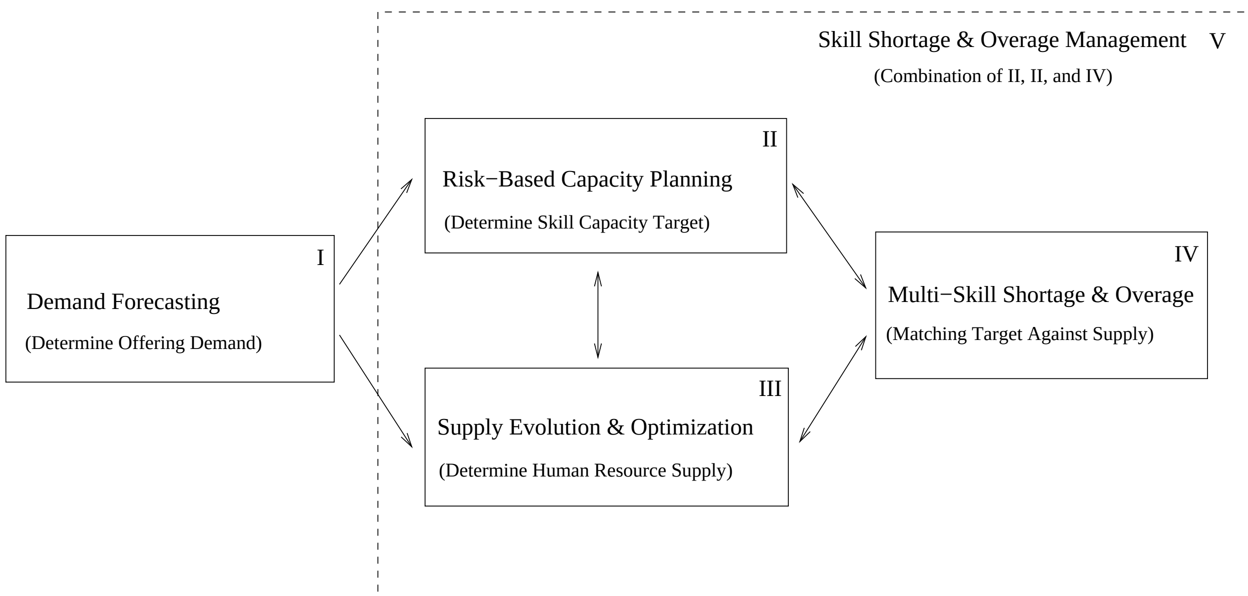


Figure 1 **The flowchart illustrates the capabilities and interrelationships in the OTM suite of OR models and methods.**

Various business and organizational aspects of the ITS human capital supply chain greatly influenced our partitioning of the OTM end-to-end solution among its constituent capabilities & components, depicted in Figure 1 and mapped to their corresponding OR models & methods in Table 1. This highlights another important contribution of the OTM suite, namely the novel integration of this collection of OR methodologies. Moreover, the integrated OTM suite was designed to support interactive sessions that allow users to evaluate different service delivery scenarios in real time, which includes scenario analysis under various assumptions regarding supply, demand, business/economic conditions, and so on. This project requirement in turn created additional methodological challenges to compute solutions of complex stochastic modeling and optimization problems in a very efficient manner to support such interative sessions and scenario analyses.

OTM capabilities	OR models and methods
Demand forecasting	Logistic regression
Risk-based capacity planning	Stochastic loss networks, stochastic optimization
Supply evolution and optimization	Stochastic temporal models, stochastic optimization and control
Multiskill shortages and overages	Mathematical programming
Skill shortage and overage management	Stochastic optimization

Table 1 **The table illustrates a mapping of OTM capabilities to OR models and methods.**

In addition to the use of OTM as an integrated suite, each OTM capability is used independently by different organizations across ITS to support various business processes associated with the ITS human capital supply chain. Given the various perspectives of its different ITS business users, the OTM capabilities & components operate across a diverse set of management & planning horizons with time scales ranging from weeks and months to quarters and even years. An overview of the primary elements comprising OTM is presented below, and then each of these elements is described in more detail in the sections that follow.

Demand Forecasting

A crucial first step in OTM human resource planning is demand forecasting (*I in Fig. 1*), which statistically characterizes the demand for each service engagement over a planning horizon of interest in terms of revenue to the provider, number of service engagements to be delivered, and human resources and skills required to do so. Service engagements are usually described only in terms of revenue, duration and solution type, without linkages to staffing templates (analogous to bills of material or bills of labor in traditional supply chains). To obtain a more accurate view of resource demand, we utilize statistical and machine learning methods to estimate the resource staffing requirements of each service engagement. Statistical forecasting techniques are also utilized to estimate the demand for each service engagement, which is then applied together with our staffing templates to obtain a complete characterization of service engagement demand and resource requirements. As an illustrative example of OTM demand forecasting output, we might respectively have a forecast of ten, five, twelve and eight service engagements over the next four quarters to design and implement data warehouses (engagement demand), each requiring two database architects, four Java programmers, a network specialist, and so on (staffing templates). Such demand forecasting (I) outputs serve as input to each of risk-based capacity planning (II) and supply evolution & optimization (III).

Risk-Based Capacity Planning

The next component in OTM human capital management & planning concerns the need to determine the capacity levels for resource skills that best satisfy the demand for these skills while also maximizing business performance over time. This capability is often generically referred to as capacity planning. Given the distinct features of human resources, the capacity planning requirements of human capital supply chains created a need for new stochastic modeling and optimization solutions to determine the skill capacity targets that maximize business performance given forecasted demand for service engagements. In traditional supply chains, product demand is often converted into machine part demand through bill of materials as direct input to supply-demand matching. However, in human capital supply chains, it is critical to first analyze the financial implications and risks associated with the distinct features of human resource skill capacity levels, including longer term costs and complex dynamics, as a function of service engagement demand.

To address these challenges, we developed a risk-based capacity planning solution of OTM (*II in Fig. 1*) that models the dynamics, tradeoffs and uncertainties associated with allocating human resources and then solves a stochastic optimization problem based on this model to maximize business performance. The risk-based capacity planning models consist of multi-class stochastic loss networks with simultaneous resource allocation, where: losses model the risks of lost demand (and associated revenue) due to insufficient human resources at the time of engagement delivery; the team of resources is required to jointly deliver the service engagement; and the multiple classes represent different types of service engagements and resource skills. Business performance is a function of the revenues for service engagement delivery, discounted by probabilities associated with engagements at risk of being lost, and the costs for maintaining skill capacity levels. Then we solve the corresponding risk-based stochastic optimization problem, with our stochastic loss network as constraints, to determine the skill capacity targets that maximizes expected business performance.

The OTM risk-based capacity planning methodology addresses the revenue-cost dynamics and tradeoffs between the risks of having insufficient resources with appropriate skills to deliver an engagement when needed and the risks of having too many underutilized resources, both resulting in reduced profit to the business. These revenue-cost dynamics and tradeoffs can be evaluated and optimized from various perspectives by using our risk-based capacity planning in different ways. This includes evaluating and determining the ideal resource capacity targets to maximize business performance (independent of existing supply) as well as evaluating and determining the optimal resource capacity levels and business investment decisions to achieve these levels in order to maximize business performance (subject to existing supply constraints).

Our risk-based methodology operates over each interval of the planning horizon that consists of stochastic behaviors invariant in distribution to shifts in time. As an illustrative example of OTM risk-based capacity planning output, expected profit might be maximized over the next quarter with 200 database architects and 500 Java programmers, whereas expected revenue might be maximized over the quarter under a 12% gross profit margin constraint with 250 database architects and 600 Java programmers. The outputs of risk-based capacity planning (II) over each time interval, which include corresponding skill capacity levels, serve as input to each of supply evolution & optimization (III) and multi-skill shortage & overage optimization (IV).

Supply Evolution & Optimization

In parallel, another fundamental OTM component concerns the modeling and analysis of human resources who are expected to be available in the future, taking into account hiring and attrition as well as adjustments in human resource skill composition, and the optimization of additional investment decisions to control the dynamics of the future evolution of such resource capacities and skill compositions. The dynamics of resources are limited in traditional supply chains, and the available decisions are typically restricted to ordering additional resources. Human capital supply chains, however, are often categorized by significant and complex time-varying dynamics on the supply side, with many human resources acquiring skills, gaining efficiencies and changing roles, some resources leaving and new resources being added. This all results in substantial changes in the characteristics and skill composition of human resources over a given planning horizon. The significant flexibility and evolutionary dynamics of human resources created the need for new stochastic temporal models of supply evolution dynamics. These time-varying evolutionary dynamics also created opportunities and challenges for optimization to adjust the future resource characteristics and skill composition through various investment decisions, such as training, promotions, other internal characteristic/skill transitions, and incentives to reduce attrition, in addition to the standard hiring (ordering) options.

To address these challenges, we developed a supply evolution & optimization solution of OTM (*III in Fig. 1*) that first models these evolutionary dynamics of human resources to estimate future human resource skill composition; and then solves a stochastic optimization (control) problem based on these models to maximize business performance over time. The supply evolution models consist of discrete-time, time-varying, multidimensional stochastic processes whose states capture the number of human resources with each combination of skills. The supply evolution optimization/control incorporates these discrete-time human resource evolution processes together with decision variables consisting of hiring, training, promotion and retention as well as the corresponding lead times for each such action. Then we solve this multi-period stochastic optimization/control problem to determine the set of discrete-time actions that adjust future resource characteristics and skill composition to maximize expected business performance over time.

As illustrative examples of OTM supply evolution & optimization output: a quarterly evolution of the human capital supply chain without intervention might suggest shortages in both database architects and Java programmers that result in lost revenue over the next few quarters; whereas expected profit might be maximized over these next few quarters by hiring two database architects and four Java programmers while incentivizing ten Java programmers to not retire. Such business investment decisions over planning horizons comprised of stochastic behaviors that vary in distribution over time, together with the corresponding resource capacity levels and skill composition over the planning horizon, serve as input to each of risk-based capacity planning (II) and multi-skill shortage & overage optimization (IV).

Multi-Skill Shortage & Overage Optimization

The next component in OTM human capital management & planning concerns the analysis of skill shortages and overages as part of matching the multi-skill supply to the demand. In a traditional supply chain, where each resource provides a single type of function, the shortages and overages for skills would then be determined as direct results of demand and supply quantities. On the other hand, the ability of human resources to perform multiple types of functions, and to do so at the same time within one or more engagements on which they are deployed, necessitates advanced forms of matching the multi-skill supply to the demand.

To address these challenges, we developed an OTM multi-skill shortage & overage optimization solution (*IV in Fig. 1*) that models the multiple skills of human resources over each interval of the planning horizon involving time-invariant stochastic behaviors; provides a detailed analysis and optimization in which the expectation of such multi-skill resources are matched to skill capacity targets to minimize a weighted sum of the skill shortages and overages; and then expected shortages or overages for each individual skill are computed under this optimal matching. We formulate the optimization problem for each interval of the planning horizon as a linear program and solve this optimization problem over the multiple intervals comprising the entire planning horizon as a dynamic program. By supporting separate weights for the shortage and overage of each skill, various business policies and priorities can be incorporated in this optimization problem.

As an illustrative example of multi-skill shortage & overage optimization output, among a set of ten human resources with both database architect and Java programmer skills in the next quarter and fourteen such human resources in the quarter after next, six (nine) might be optimally matched to deploy database architect skills with the remaining four (five) optimally matched to deploy Java programmer skills in the next quarter (quarter after next). These multi-skill shortage & overage optimization (IV) outputs, including the optimal matchings of multi-skill supply resources to skill capacity targets and the corresponding skill shortages and overages across time intervals of both time-invariant and time-varying stochastic behaviors, serve as input to each of risk-based capacity planning (II) and supply evolution & optimization (III).

Skill Shortage & Overage Management

The final component of the OTM suite concerns the modeling and optimization of business investment decisions (e.g., hiring, training and retention) to manage the shortages and overages across all skills. To this end, we developed an approach for skill shortage & overage management that combines our solutions for risk-based capacity planning, supply evolution & optimization, and skill shortage & overage optimization through an iterative procedure (*V comprising the dashed box in Fig. 1*). Specifically, our risk-based capacity planning models & optimization (*II in Fig. 1*) are used to determine the resource capacity levels that maximize performance. This is complemented by our supply evolution & optimization (*III in Fig. 1*) that optimizes the evolutionary dynamics of future resource capacity levels and skill compositions together with longer term investment decisions. Our multi-skill shortage & overage optimization (*IV in Fig. 1*) then determines the matching of multi-skill human resources against skill capacity targets to maximize performance.

The final resource planning solution (*V in Fig. 1*) is then obtained through a method of iteration among these three OTM optimization components by exploiting their interactions, dependencies and various time scales until efficiently converging to a fixed-point equilibrium (steady state). This combination of OTM components within an iterative methodology (V) provides a set of investment decisions that manage skill shortages and overages to maximize business performance. As an illustrative example of OTM skill shortage & overage management output, to manage an overage of one database architect and a shortage of six Java programmers, among others, over the next few quarters, it might be best to retrain one of the database architects, hire four Java programmers and incentivize two Java programmers from retiring.

The dynamics and uncertainties of such decision making within human capital supply chains occur at different time scales, as is also reflected in the details of each of the components comprising the OTM integrated suite of OR models & methods. Specifically, our risk-based capacity planning models & optimization operate on stationary intervals (typically on the order of a month or quarter) of the planning horizon; our supply evolution models & optimization operate across the entire non-stationary planning horizon (typically on the order of a few quarters, a year or more); and our multi-skill shortage & overage models & optimization operate across both stationary and non-stationary time intervals. Here, the notion of non-stationarity refers to stochastic behaviors that vary in distribution over time, whereas stationarity refers to stochastic behaviors that are invariant in distribution to shifts in time over a sufficiently long time interval. We generally consider planning horizons that include different forms of dynamics and uncertainties, as well as different forms of business investment decisions. These planning horizons represent non-stationary time intervals that are comprised of multiple stationary subintervals, where the different forms of complex dynamics and uncertainties change from one stationary subinterval to the next according to a general stochastic process.

Demand Forecasting Models

An important first step is the forecasting of service engagement demand over future periods which is comprised of demand from three sources: (a) *ongoing engagements* – those already being delivered; (b) *opportunities* – potential deals at different stages in the sales pipeline; and (c) *expected deals* – various types of deals expected based on market research and experience, but not concrete enough for entry into the sales pipeline. Service engagements typically require a collection of different skills, where the staffing templates that link engagement types and expected revenue to skill requirements in human capital supply chains are considerably more complex than the corresponding bills of material/labor in traditional supply chains.

Statistical forecasting techniques are applied to the latest delivery information to compute probabilistic characterizations of the completion of *ongoing engagements* in future periods and the roll-off dates for deployed resources. For *pipeline opportunities*, we develop statistical models based on logistic regression techniques to predict the probability of each pipeline deal being won (i.e., a service delivery contract is signed) based on attributes such as lapse time and recent movement in the pipeline, client information, deal size, and so on. These win probabilities are then used to probabilistically characterize the number of engagements of each type from the pipeline over future periods. (Note that our approach exploits logistic regression as a preferred method in statistics for estimating binary outcomes and has no connection with the 2009 INFORMS Franz Edelman Award Finalist work by IBM, Lawrence et al., which uses quantile estimation of a single variable that couples revenue and win probability, not the win probability in isolation. This is because quantile estimation is more suited for estimating piecewise linear functions.) For *expected deals*, the quarterly revenue targets and typical deal sizes are used to probabilistically characterize the number of engagements of each type that are expected to be signed in order to make up the difference between revenue targets and expected revenues from ongoing engagements and pipeline opportunities.

Given the lack of direct linkage between service engagements/opportunities and required resources, we develop models to estimate the staffing requirements for each type of engagement. Following the statistical clustering methodology in Hu et al. (2007), we generate groups of similarly staffed engagements using historical information on service engagement delivery. Such clusters for each type of engagement are validated and refined through discussions with subject matter experts. To address the evolution of these staffing templates and the need for some degree of customization, we also develop an automated methodology for generation and adjustment of staffing templates by exploiting the semi-supervised clustering and machine learning framework studied in Hu et al. (2008a,b). The resulting dynamic taxonomy of staffing templates provides critical input to ensure accurate forecasting of resource requirements from all three major engagement demand sources.

Risk-Based Capacity Planning

The estimates of demand and resource requirements for each service engagement from the demand forecasting models are provided as input to our risk-based capacity planning solution. Our goal is to determine the skill capacity levels that best satisfy demand while addressing the complex trade-offs among revenue, cost, and associated risks. To highlight some of the difficulties involved, consider a scenario demand forecast for 10 service engagements over the next quarter. The best skill capacity levels to deliver this set of engagements depend on the degree of overlap among the corresponding service delivery time processes. At one extreme, if the delivery of all engagements overlaps, then 10 times the required staffing levels of one engagement is required; at the other extreme, if the delivery of these engagements is disjoint, then only the staffing levels of one engagement are required. Because the reality is typically somewhere between these extremes, a model of these complex dynamics and stochastic behaviors is needed to appropriately determine the skill capacity levels that maximize business performance subject to service engagement demand forecasts.

We therefore model the capacity planning problem as a stochastic loss network in which the risk of losing service engagement demand (and associated revenue) due to insufficient capacity in one or more required skills at the time the service engagement must be delivered is represented by the stationary loss probabilities of the stochastic network. Here, the stochastic processes modeling service engagement “arrivals” are used to represent the time epochs at which these different service engagements must be delivered. We note that the loss of service engagement demand due to insufficient resources with required skills was a critically important issue at the start of the OTM project and this, together with the cost-revenue tradeoff of having too many underutilized resources or too few required skills, greatly influenced our choice to use a stochastic loss network to model and optimize the capacity planning problem. The overall planning horizon (on the order of several quarters) consists of a sequence of coarse subintervals, each on the order of a month or quarter and involving a stationary loss network under a fixed set of parameters that changes from one subinterval to the next according to a general stochastic modulation process. We then formulate and solve a capacity planning optimization problem to determine the skill resource capacity levels for our stochastic loss network that maximize business performance over the planning horizon.

Our derivations and theoretical results for the risk-based capacity planning models are based in part on the slice methods developed in Jung et al. (2008) when restricted to the case of Poisson arrival processes (which model the service delivery epochs from the demand), where we exploit some of these preliminary results in this project for the management of human capital supply chains in practice. In addition to the innovative application of these aspects of the slice methods as part of our risk-based capacity planning models under Poisson arrivals, we also extend these risk-based capacity planning model results and methods as part of this project to support more general arrival processes. This together with our new results and methods for stochastic risk-based optimization supports the accurate modeling of the uncertainty and dynamics that characterize human capital supply chains as stochastic loss networks and our optimization of business performance tradeoffs in the capacity planning of the ITS human capital supply chain. Furthermore, our risk-based capacity planning models and optimization have been calibrated and validated against available ITS data over time. All of these results (and additional results presented below) provide theoretical support for our approach, which has proven quite effective across the full range of model parameters and instances from ITS human capital supply chain data and business cases as part of this project.

Stochastic Risk-Based Models

We model each stationary period of the capacity planning problem as a stochastic loss network consisting of a set of skills $\mathcal{I}$ and a set of service engagements $\mathcal{K}$, where the delivery of service engagements involves collections of skill capacities. In particular, the delivery of engagement $k \in \mathcal{K}$ requires $A_{i,k} \geq 0$ units of capacity from resources with skill $i \in \mathcal{I}$, where each skill i has $C_i \geq 0$ units of capacity overall. Instances of engagement k (demand) need to be delivered according to an independent stochastic process. Such a delivery opportunity is (at risk of being) lost if the available capacity for any skill i is less than $A_{i,k}$, and otherwise the engagement is delivered by reserving capacity $A_{i,k}$ for each skill i throughout the duration of the service delivery. The engagement delivery duration times are independent and identically distributed random variables following a general distribution with unit mean (without loss of generality). Engagement delivery epochs and duration times are mutually independent.

Let us define

$$
\begin{aligned}
E_i &:= \text{the stationary blocking event probability for skill } i,\\
L_k &:= \text{the stationary loss risk probability for engagement } k,\\
n_k &:= \text{the number of active type-}k\text{ service delivery engagements in the network in equilibrium (steady state)},\\
\boldsymbol{n} &:= (n_1, \ldots, n_{|\mathcal{K}|}) \in \mathbb{Z}_+^{|\mathcal{K}|}, \ \text{ the active service delivery engagements vector in equilibrium},
\end{aligned}
$$

where $i \in \mathcal{I}$, $k \in \mathcal{K}$, and $\mathbb{Z}_+$ denotes the set of non-negative integers. Here, E_i represents the probability in equilibrium (steady state) that service delivery arrivals encounter insufficient capacity for skill i (and therefore create a risk of losses of the corresponding engagement demands, i.e., blocking events at the skill level create a risk of corresponding loss events at the level of engagement demands), whereas L_k represents the probability in equilibrium that service delivery arrivals (at the times these service engagements need to be delivered) find insufficient capacity for one or more skills required for engagement k (and therefore the corresponding engagement demands are at risk of being lost).

Recall that

$$
\begin{aligned}
A_{i,k} &:= \text{the amount of capacity from resources with skill } i \text{ required to deliver engagement } k,\\
C_i &:= \text{the amount of overall capacity from resources with skill } i,
\end{aligned}
$$

where $\mathbf{A} := [A_{i,k}]$ and $\mathbf{C} := (C_i)$. Note that the kth column of matrix $\mathbf{A}$ corresponds to the skill capacity staffing template for service engagement k from our demand forecasting models. Then, by definition, we have

$$\boldsymbol{n} \in \mathcal{S}(\mathbf{C}) = \{\boldsymbol{n} \in \mathbb{Z}_+^{|\mathcal{K}|} : \mathbf{A}\boldsymbol{n} \leq \mathbf{C}\}.$$

When the engagement arrivals are properly modeled as Poisson processes with rates ν_k, (i.e., the time epochs when type-k service engagements need to be delivered as part of satisfying demand follow a Poisson point process with rate ν_k), $k \in \mathcal{K}$, then the above model is equivalent to the famous *Erlang loss model* which has been studied for over a century since the seminal work of Erlang Erlang (1917). In this model instance, it is well known that the stationary distribution $\boldsymbol{\pi}$ of $\boldsymbol{n}$ is unique and exhibits a product form given by

$$\boldsymbol{\pi}(\boldsymbol{n}) = G(\mathbf{C})^{-1} \prod_{k=1}^{|\mathcal{K}|} \frac{\nu_k^{n_k}}{n_k!}, \qquad G(\mathbf{C}) = \sum_{\boldsymbol{n} \in \mathcal{S}(\mathbf{C})} \prod_{k=1}^{|\mathcal{K}|} \frac{\nu_k^{n_k}}{n_k!}.$$

Hence, the stationary loss risk probability for type-k engagements can be expressed as

$$L_k = 1 - G(\mathbf{C})^{-1} G(\mathbf{C} - \mathbf{A}\mathbf{e}_k),$$

where $\mathbf{e}_k$ is the unit vector whose k-th element is one and all others are zero. The loss risk probability vector $\mathbf{L} := (L_1, \ldots, L_{|\mathcal{K}|})$ plays an important role in our stochastic model and optimization as a measure of the service engagements at risk of being lost, together with their estimated lost revenue.

However, due to the computational complexity of calculating the normalizing constant $G(\mathbf{C})$, which is known to be $\sharp P$-complete (a higher complexity class than NP-complete) in the size of the network Louth et al. (1994), an Erlang fixed-point approximation has been long used as a more efficient alternative to the exact Erlang loss formula. The Erlang fixed-point approximation is based on approximating the blocking event probabilities of the individual skills, E_i, by a set of fixed-point equations, and then approximating the loss risk probabilities for type-k service engagements, L_k, in terms of these skill blocking event probabilities. Specifically,

$$L_k \approx 1 - \prod_{i=1}^{|\mathcal{I}|} (1 - E_i)^{A_{i,k}}, \qquad \rho_i = (1 - E_i)^{-1} \sum_{k=1}^{|\mathcal{K}|} A_{i,k} \nu_k \prod_{i'=1}^{|\mathcal{I}|} (1 - E_{i'})^{A_{i',k}}, \tag{1}$$

$$E_i \;\approx\; \mathcal{E}\left(\rho_i, C_i\right), \qquad \mathcal{E}(\rho, C) \;=\; \frac{\rho^C}{C!}\left(\sum_{n=0}^{C}\frac{\rho^n}{n!}\right)^{-1}, \tag{2}$$

where the last expression is the (exact) Erlang formula for the loss probability of an isolated skill with capacity C under arrival rate ρ. We refer the interested reader to Kelly (1991) for additional details on the Erlang fixed-point approximation.

Even though the Erlang fixed-point approximation resolves the prohibitive computational costs of the exact formula for the problem sizes of interest in this OTM project, it is also well known that the Erlang fixed-point approximation can provide relatively poor estimates for the engagement loss risk probabilities L_k in various model instances, which we certainly found to be the case when applied to our human capital supply chain demand forecasts. To address this set of challenges, we first observe that by definition the mode $\boldsymbol{n}^*$ of the stationary distribution $\boldsymbol{\pi}(\cdot)$ corresponds to a solution of the optimization problem

$$\max_{\boldsymbol{n}} \;\sum_{k=1}^{|\mathcal{K}|} n_k \log \nu_k - \log n_k! \qquad \text{subject to} \qquad \boldsymbol{n} \in \mathcal{S}(\mathbf{C}). \tag{3}$$

Next we define a natural continuous relaxation of the state space $\boldsymbol{n} \in \mathcal{S}(\mathbf{C})$ and subsets of this relaxation

$$\bar{\mathcal{S}}(\mathbf{C}) \;:=\; \{\mathbf{x} \in \mathbb{R}_+^{|\mathcal{K}|} : \mathbf{A}\mathbf{x} \le \mathbf{C}\} \qquad \text{and} \qquad \bar{\mathcal{S}}_{\ell,k}(\mathbf{C}) \;:=\; \bar{\mathcal{S}}(\mathbf{C}) \cap \{\mathbf{x} : x_k = \ell\},$$

respectively, for which we first obtain the following optimization problem corresponding to (3):

$$\max_{\mathbf{x}} \;\sum_{k=1}^{|\mathcal{K}|} x_k \log \nu_k - \log \Gamma(x_k + 1) \qquad \text{subject to} \qquad \mathbf{x} \in \bar{\mathcal{S}}(\mathbf{C}), \tag{4}$$

where $\mathbf{x} = (x_k)$ is the corresponding continuous relaxation of $\boldsymbol{n}$ and $\Gamma(\cdot)$ denotes the gamma function.

Then, for each service engagement k, we derive the following convex relaxation for (4) by exploiting Stirling's approximation (i.e., $\log\Gamma(x_k+1) = x_k \log x_k - x_k + O(\log x_k)$), ignoring the $O(\log x_k)$ term and restricting to each slice defined by $n_k = \ell$, for $\ell \in \{n_k : \mathbf{n} \in \mathcal{S}(\mathbf{C})\}$:

$$\max_{\mathbf{x}} \;\sum_{k=1}^{|\mathcal{K}|} x_k \log \nu_k + x_k - x_k \log x_k \qquad \text{subject to} \qquad \mathbf{x} \in \bar{\mathcal{S}}_{\ell,k}(\mathbf{C}).$$

We solve this optimization problem to obtain the mode $\mathbf{x}^*(\ell,k)$ of the distribution for each slice $n_k = \ell$, which is used to obtain our approximation of $\Pr[n_k = \ell]$ as a function of $\mathbf{x}^*(\ell,k)$ for each slice $n_k = \ell$. Namely, $\Pr[n_k = \ell] \approx \exp(q(\mathbf{x}^*(\ell,k)))$ where $q(\mathbf{x}) = \sum_k x_k \log \nu_k + x_k - x_k \log x_k$. This then yields an approximation for L_k in terms of $\mathbb{E}[n_k]$ by exploiting the derived relationship $L_k = 1 - \mathbb{E}[n_k]/\nu_k$ and the definition $\mathbb{E}[n_k] = \sum_{\ell=0}^{\infty} \ell \Pr[n_k = \ell]$. Specifically, upon restricting the range of ℓ to the polytope, we have

$$\mathbb{E}[n_k] \;\approx\; \frac{\sum_\ell \ell \exp(q(\mathbf{x}^*(\ell,k)))}{\sum_\ell \exp(q(\mathbf{x}^*(\ell,k)))}.$$

To reduce the computational complexity of this approach to that of the Erlang fixed-point approximation, we also develop a 3-point slice method where the above convex relaxation is solved for $\ell = 0$, for the maximum value of ℓ and for the mode $\mathbf{x}^*$ of the overall distribution, and then $\mathbf{x}^*(\ell, k)$ is approximated for all other values of ℓ by interpolation between successive pairs of the 3 computed modes.

It has been shown in Jung et al. (2008) that the above solutions are asymptotically exact in the following limiting regime. Consider a scaled version of the stochastic loss network defined by the scaled capacities $\mathbf{C}(N) = N\mathbf{C} = (NC_1, \ldots, NC_{|\mathcal{I}|})$ and the scaled arrival rates $\boldsymbol{\nu}(N) = N\boldsymbol{\nu} = (N\nu_1, \ldots, N\nu_{|\mathcal{K}|})$, where $N \in \{1, 2, \ldots\}$ is the system scaling parameter. Then the loss risk probabilities from the above slice methods converge to the exact loss risk probabilities as the scaling parameter N tends to infinity.

In addition, as part of this OTM project, we have developed accurate solutions for stochastic loss networks under general renewal arrival processes with rates ν_k and (interarrival) variances σ_k^2 (i.e., the time epochs when type-k service engagements need to be delivered as part of satisfying demand follow a renewal point process with parameters ν_k and σ_k^2). Specifically, we derive a Gaussian fixed-point approximation based on the corresponding underlying multidimensional Gaussian process in which the loss probabilities L_k for type-k engagements are given by (1) together with

$$E_j \;=\; \mathcal{G}\left(\rho_j, \sigma_j, C_j\right), \qquad \mathcal{G}(\rho, \sigma, C) \;=\; 1 - \mathcal{N}\left(\frac{C-\rho}{\sigma}\right), \tag{5}$$

where the last expression is the one-dimensional Gaussian process for an isolated skill with capacity C under interarrivals with mean ρ^{-1} and variance σ^2, and $\mathcal{N}(\cdot)$ is the standard normal distribution function. We have established asymptotic exactness of our Gaussian fixed-point approximation under the above large network scaling. This asymptotic exactness follows from an instance of the central limit theorem for conditional random variables resulting from the capacities and arrival rates growing together. We have also proven strict dominance of the sliced-based methods in terms of accuracy over previous Erlang approximations by deriving and comparing large deviations results for the exact Erlang formula, the Erlang fixed-point approximation and our slice based methods. Throughout our ITS deployment over the past couple of years, we have found our risk-based capacity planning models to yield accurate solutions.

Stochastic Risk-Based Optimization

Although stochastic loss networks model the dynamics and complexities of risk-based capacity planning, our objective is to determine the skill capacity levels that maximize business performance over the planning horizon. Specifically, we formulate and solve a stochastic capacity planning optimization problem to determine the skill resource capacity levels (namely, the capacity vector $\mathbf{C}^*$) for a stochastic loss network (with capacity requirement matrix $\mathbf{A}$, arrival rate vector $\boldsymbol{\nu} := (\nu_1, \ldots, \nu_{|\mathcal{K}|})$ and, in the case of general renewal arrivals, interarrival variance vector $\boldsymbol{\sigma} := (\sigma_1, \ldots, \sigma_{|\mathcal{K}|})$) that maximize business performance over a long-run planning horizon. We shall assume that the length of each subinterval is sufficiently long for the multidimensional stochastic process modeling the loss network to reach stationarity, where the multiple time scales involved in human capital supply chains provide both theoretical and practical support for our stationary stochastic approach. Focusing on a single stationary subinterval, in order to simplify the presentation, our objective function is based on rewards (revenues) gained for delivering engagements that can be serviced at the time of their required delivery epoch (arrival) and on penalties (costs) incurred as the result of deploying resource skill capacity levels. More precisely, our objective function is given by

$$\max_{\mathbf{C}} \; \sum_{k=1}^{|\mathcal{K}|} u_k \, (1 - L_k) \, \nu_k \; - \; \sum_{i=1}^{|\mathcal{I}|} v_i \, C_i, \tag{6}$$

where u_k is the base revenue rate for service engagement k and v_i is the base cost rate for skill i resource capacity. The constraints of this optimization problem include our foregoing approximations for the loss risk probabilities L_k for each engagement of the stochastic loss network. Constraints of the form $L_k \le \beta_k$, for $\beta_k \in (0, 1]$, can also be included in our formulation to guarantee a desired serviceability level or market share. We also considered in this project related optimization formulations which include maximizing revenue or minimizing cost subject to constraints on the gross profit margin or on cost/revenue targets.

The above formulation yields a nonlinear program and our solution exploits its properties based on our results for the underlying stochastic loss network with either Poisson or general renewal arrival processes. This includes our establishing that, for a stochastic loss network under the preceding large system scaling by parameter N, the solution of the above stochastic optimization problem converges asymptotically, in the limit as N tends to infinity, to the optimal solution under the exact loss probabilities. In practice, our experience has found the sizes of the risk-based capacity planning optimization to be sufficiently large so as to obtain accurate solutions.

Another set of properties we exploit includes determining a region of capacities $\mathbf{C}$ that ensures the arrival rate vector $\boldsymbol{\nu}$ (and the interarrival variance vector $\boldsymbol{\sigma}$, when arrivals come from a general renewal process) will be served with loss risk probabilities of at most $\mathbf{L}$. This region characterizes fundamental properties between the skill supply capacity and engagement demand loss-probability vectors that can be exploited to efficiently search the feasible region in our stochastic capacity planning optimization problem. In particular, we establish that, for a loss network scaled by parameter N, there exists a constant $\delta(N)$ such that for any given feasible loss risk probabilities $L_k(N)$ and any small positive number $\epsilon \ll 1$, the capacity vectors $\mathbf{C}(N)$ that achieve these loss risk probabilities fall within the region defined by the following system of polynomial equations and inequalities, for each skill i and engagement k:

$$\log(1 - L_k(N) - \delta(N) N^{-1/2+\epsilon}) \le -\sum_{i'=1}^{|\mathcal{I}|} A_{i',k} E_{i'}(N), \tag{7}$$

$$\rho_i(N) = \sum_{k=1}^{|\mathcal{K}|} \nu_k(N) A_{i,k} \prod_{i'=1, i' \neq i}^{|\mathcal{I}|} (1 - E_{i'}(N))^{A_{i',k}}, \tag{8}$$

$$\rho_i(N)(1 - E_i(N)) < C_i(N) < \rho_i(N)(1 - E_i(N)) + 1/E_i(N). \tag{9}$$

These polynomial equations and inequalities can be shown to hold with respect to our loss risk probability approximations under both Poisson and general renewal arrival processes.

The above polynomial equations and inequalities characterizing the feasible region of our risk-based capacity planning optimization problem are instrumental in significantly improving the efficiency of its solution. Specifically, by incorporating the polynomial equations and inequalities (7) – (9) into our optimization problem, adding the corresponding constraints on L_r and exploiting the methodologies developed in Lasserre (2001), we obtain a near-optimal solution with polynomial computational complexity and probabilistic accuracy guarantees. To this end, we convert such a polynomial optimization problem into a positive semidefinite program that can approximate this optimization problem as closely as desired. We first combine the polynomial equations and inequalities (7) – (9) with cost and revenue functions of $\mathbf{L}$ and $\mathbf{C}$, along with the introduction of Lagrangian multipliers. Then we formulate the problem as a polynomial optimization problem over a compact set $\mathcal{C}$ defined by these polynomial equations and inequalities, where the objective function $g(x)$ is an m-degree multivariate polynomial having representation $\sum_\alpha g_\alpha x^\alpha$ such that $x^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ forms the basis of the space of the m-degree polynomials with (g_α) the coefficient vector of $g(x)$ and $\sum_i \alpha_i \leq m$. We then exploit the theoretical results established in Lasserre (2001) to compute solutions of our capacity planning optimization problem with provable probabilistic accuracy guarantees in polynomial time.

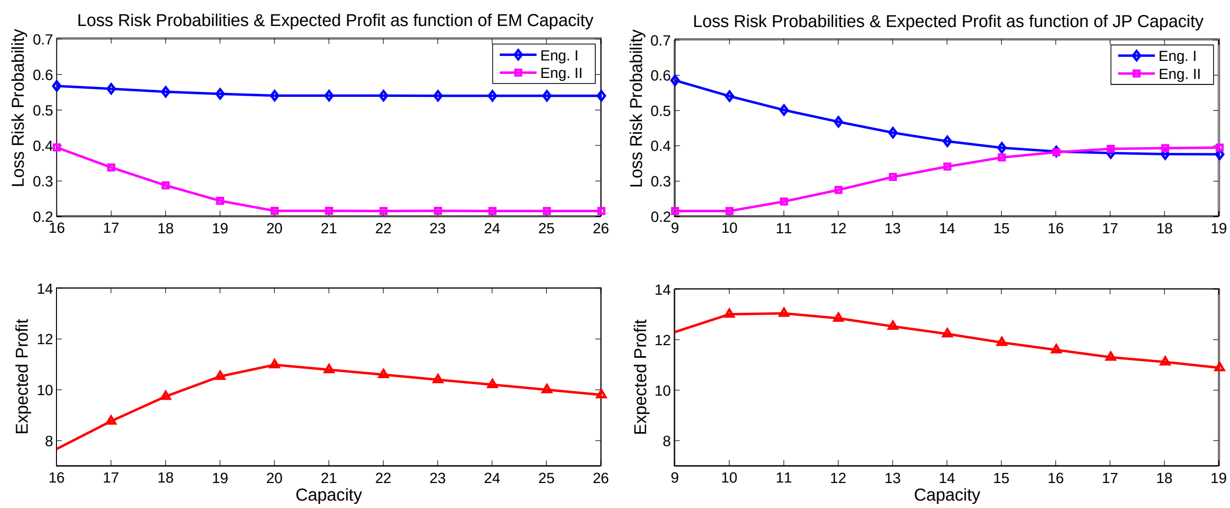
The preceding formulation considers only the costs of deploying resource capacity levels without addressing the costs of adjusting capacity levels and skill composition through business decisions and investments. This approach was taken both to elucidate the exposition and because the first-phase of risk-based capacity planning as part of our end-to-end solution is used to determine the ideal skill resource capacity levels from a business/financial perspective. Our formulation and solution, however, is quite general and easily extends to incorporate the costs of decisions/investments for realizing resource capacity levels. We then use this in combination with the supply evolution & optimization and skill shortage & overage optimization components of OTM to determine the optimal skill resource capacity levels and to optimally address skill shortages and overages. Additional details along these lines are discussed in subsequent sections.

Finally, we note that the previous study in Bhadra et al. (2007) considers the structural properties of a multi-period stochastic loss network model under Poisson arrivals whose solution is obtained via the Erlang fixed-point approximation. The resulting properties do not generally apply to our multi-period stochastic loss network model under renewal arrival processes nor to our solution of the stochastic network using either slice method.

Simple Illustrative Example

To illustrate the complexities of risk-based capacity planning and the benefits of our approach, we briefly present a stochastic loss network example consisting of three skills S_1, S_2, S_3 serving two service engagement types T_I, T_{II}. While the capacity for S_2 is shared by both engagements, the capacity for S_1 and S_3 independently support T_I and T_{II}, respectively, where T_I (T_{II}) consume one unit of S_1 (S_3) and S_2. The arrival epochs for when engagements of type I and II need to be delivered follow independent Poisson processes with rates 20 and 10, respectively. Let C_1, C_2, C_3 denote the capacity of S_1, S_2, S_3, respectively.

Generally speaking, the higher the capacities, the lower the loss risk probabilities. However, resource sharing among the service engagements (S_2 in this example) can create complex interactions among the resource capacities. In this specific example, which sheds light on the complex dynamics in the general problem, we observe that the simple operation of increasing the shared skill capacity by one unit can result in dramatically different loss risk probabilities for a given demand depending on the different parameter regions in which the system operates. Figure 2 displays the corresponding results for this simple example.



(a) Case 1 (b) Case 2

Figure 2 **The graphs show an example of risk-based capacity planning.**

In Case 1, we fix $C_1 = 10$ and $C_3 = 10$, and increase C_2 from 16 to 26. We observe that, while both loss risk probabilities decrease monotonically, the rates of decrease for the two engagement types are quite different. We further see that for the same engagement type, the rate can vary quite significantly for different values of C_2. In Case 2, the capacity $C_2 = 20$ and $C_3 = 10$ are fixed, and C_1 varies from 9 to 19. The example indicates that increasing one unit of C_1 can have a very different impact on the loss risk probabilities of different engagement demands. Specifically, increasing one unit of C_1 results in a decrease in the loss risk probability of T_I, but increases the loss risk probability of T_{II}. Intuitively, this is because under the current parameter settings, S_1 is a bottleneck for T_I, i.e., when a type-I engagement is lost, it is more likely due to a lack of S_1 capacity. By increasing S_1 capacity, we allow more type-I engagements to be accepted. But this, in turn, results in more of the common resource S_2 being utilized by T_I, and hence the loss risk probability of T_{II} will increase. Furthermore, these complexities significantly impact the question of optimal capacity levels, as can be observed from the expected profit curves for both cases in Figure 2.

More generally, in larger and more realistic instances of our risk-based capacity planning model of human capital supply chains, it is not possible to distinguish the degrees of sharing among the skills, given the complex interactions and dependencies among the service engagements and their skill requirements. Thus, the effects of changing skill supply capacity on engagement demand loss risk probabilities is often very complicated, which is further complicated in turn by the complex dynamics and interactions underlying the optimization of capacity levels, and hence both require advanced OR models & methods.

Finally, we consider a representative application of risk-based capacity planning using ITS data. Figure 3 plots expected profits, revenues, and costs on the y-axis as a function of risk tolerance constraints along the x-axis. The leftmost set of results represents the solution that maximizes expected profit. The risk tolerance constraints become more tight as we move to the right along the x-axis, where the rightmost set of results represent when most of the demand being satisfied. We observe that although expected revenues increase as more demand is satisfied, expected costs also increase and tend to do so at a faster rate. This application illustrates how executives can use risk-based capacity planning to determine the best way to operate their business with respect to expected revenues, costs, and profits, revenue loss-risk tolerances, and other business and economic concerns.

Supply Evolution and Dynamics

The management of future resource and skill capacity levels over a long-run planning horizon is another fundamental aspect of human capital supply chains. Our goal is to understand the evolution of human-talent levels over time and determine the best actions to adjust these evolutionary dynamics. To highlight some difficulties involved, consider a common scenario in the military in which a long period is required for one to reach the highest ranks (e.g., General). Steps must be taken at all levels of this human capital supply chain (e.g., recruiting, training, experience, and promotions) to ensure that enough talent is available at all levels, with lower-level ranks feeding higher-level ranks over time. A model of these complex dynamics and stochastic behaviors is needed to determine appropriate talent levels that maximize expected performance over time, subject to supply, demand, and organizational constraints. This challenge involves stochastic modeling and analysis of the evolutionary dynamics and flexibility of human capital, including

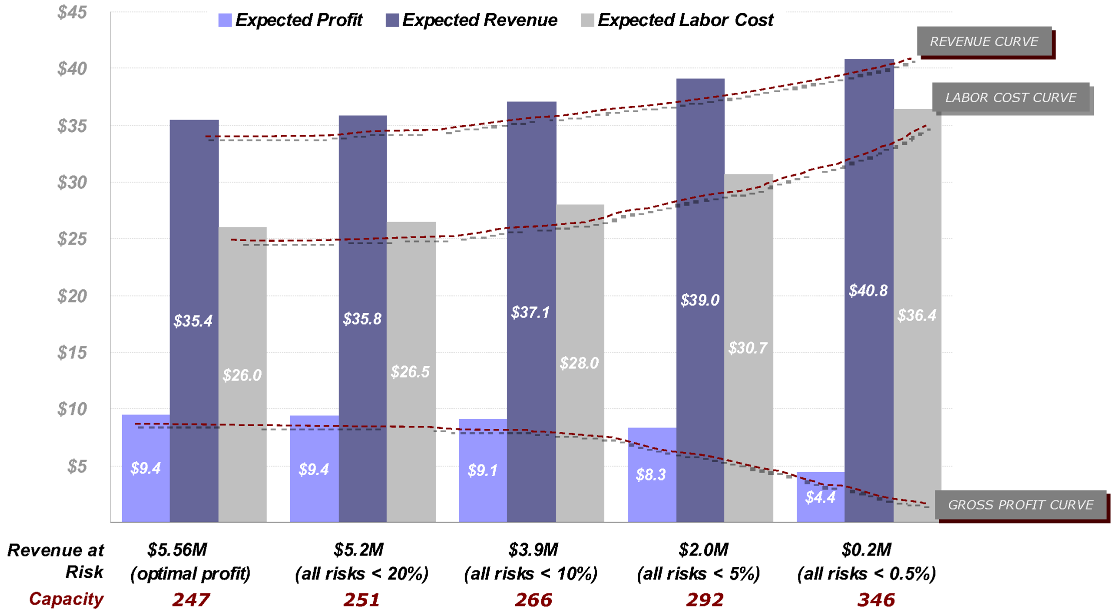


Figure 3 **The graph shows an example business application of risk-based capacity planning.**

future internal transitions (e.g., promotions, certification, and training) and future external transitions (e.g., hiring, attrition, and acquisitions). The time scale of these dynamics is typically on the order of days, weeks, or months, whereas the overall planning horizon is often on the order of months, quarters, or years.

To capture the various sources of uncertainty in these complex supply dynamics, we model the future temporal evolution of human resources and skill composition as a discrete-time multidimensional stochastic process in which the dynamics and uncertainties vary over time. For this purpose, we leverage data on current supply, historical data and information on human resource dynamics and business/economic conditions, as well as input from subject matter experts. In addition to probabilistically estimating future resource and skill capacity levels, the dynamic evolution over time of human resource characteristics and skill composition can be influenced in desired directions through various business investments/policies and in response to uncertain and changing business/economic conditions. We therefore formulate and solve a multi-period stochastic optimization (control) problem based on our supply evolution model to determine the future evolution of human resources and skill composition through available investment decisions in order to maximize business performance over time. Such available investment decisions in each period include hiring, training, promotions, other internal characteristic/skill transitions, and incentivizing to reduce attrition, together with the lead times associated with each of these actions.

To model and optimize the human capital supply chain evolutionary dynamics at temporal granularities dictated by business operations, we use a multiperiod discrete-time stochastic process defined over a planning horizon comprised of $T+1$ periods. Although our OTM approach is completely general, the time granularity most often used is monthly and quarterly periods.

Stochastic Evolution Models

Our stochastic supply evolution models are based on aggregating people into human capital groups according to attributes of interest. Examples include all talent and skills relevant to the business, and other factors such as levels of competency, productivity, proficiency, and certification. Because each person comprising the supply side is capable of attaining and employing a collection of attributes, our supply evolution models are more precisely based on aggregating people into human capital groups according to various combinations of attributes. Within each period of the planning horizon, individuals make dynamic transitions between these human capital groups. Such dynamic transitions include both internal and external transitions that lead to a complex stochastic network topology, both within each period and from one period to the next across the planning horizon.

Consider a model comprised of five human capital groups: C programmers, Java programmers, SQL programmers, programmers proficient in C and Java, and programmers proficient in Java and SQL. Now consider the evolution of individuals among these human capital groups over a single period. People are hired into each group or attritted out of each group with certain probabilities, such as the probability that some number of C programmers are hired (attritted) within the current period. Similarly, probabilities are associated with internal transitions among human capital groups, such as the probability with which some number of C programmers become equally proficient in Java programming. The corresponding sets of probabilities depend upon the period in which the dynamic transitions occur and the number of C programmers at the start of this period. We exploit properties of this complex network topology to obtain scalable solutions for our stochastic supply evolution models; one example is sparse matrix methods, given that transitions tend to be localized.

Our stochastic supply evolution models consist of a discrete-time, multidimensional stochastic process that records the number of people in each human capital group. Now consider a portion of this stochastic process corresponding to the population of a single human capital group (e.g., Java programmers). At the start of the planning horizon, we will know with certainty that the organization has a specific number of Java programmers, thus yielding a single state for each human capital group at time 0. At the next period, the state space of the stochastic process then needs to allow for all possible populations within the human capital group of Java programmers, governed by the corresponding set of transition probabilities. For example, assuming a current count of Java programmers at time 0, the number of Java programmers will increase or decrease to a specific number of Java programmers at time 1 with a corresponding set of probabilities for hires and internal transitions into this human capital group and for attrition and internal transitions out from this human capital group. Upon considering this portion of the stochastic process for all possible populations and expanding this view to the entire stochastic process across the large number of human capital groups, we clearly have a prohibitive curse of dimensionality problem as the state space blows up.

To address this curse of dimensionality, we consider a decomposition of the discrete-time stochastic process that records the expected number of people in each human capital group. This renders a single state for each human capital group and an analogous set of probabilities that governs transitions from one state to another within each period and across periods that comprise the planning horizon. More formally, recall that $\mathcal{I}$ denotes the set of skills labeled by i. Because each resource comprising the supply is capable of employing a subset of different skills, let us define

$$\mathcal{J} := \text{the family of subsets of the set of skills } \mathcal{I} \text{ that are possessed by human resources,}$$

in which case the types of resources comprising the supply consist of elements of $\mathcal{J}$, labeled by j. Note that these resource types j can include factors such as productivity and efficiency levels. Now we define

$$\begin{aligned}
y_j(t) &:= \text{the expected number of resources of type } j \text{ at time } t, \\
h_j(t) &:= \text{the expected number of hires of type } j \text{ over the time interval } [t, t+1), \\
a_j(t) &:= \text{the expected amount of attrition of type } j \text{ over the time interval } [t, t+1),
\end{aligned}$$

where $\mathbf{y}(t) := (y_1(t), \ldots, y_{|\mathcal{J}|}(t))$, $\mathbf{h}(t) := (h_1(t), \ldots, h_{|\mathcal{J}|}(t))$ and $\mathbf{a}(t) := (a_1(t), \ldots, a_{|\mathcal{J}|}(t))$ represent the corresponding human capital state, hiring and attrition vectors, respectively, for $j \in \mathcal{J}$ and $t = 0, \ldots, T$.

The complex dynamics of evolving human capital supply chain resources and skills over time need to be properly captured in our stochastic temporal models of human resource evolution. To this end, we define

$$\begin{aligned}
p_{j,j'}(t) := \;& \text{the stationary probability that a type-}j \text{ human resource transitions to become a type-}j' \text{ human} \\
& \text{resource over the time interval } [t, t+1), \text{ where } \textstyle\sum_{j' \in \mathcal{J}} p_{j,j'}(t) \leq 1.
\end{aligned}$$

In other words, $p_{j,j'}(t)$ represents the probability of transitions from state j to state j' over the interval $[t, t+1)$. When type-j attrition is strictly positive ($a_j(t) > 0$), the above inequality is strict ($\sum_{j' \in \mathcal{J}} p_{j,j'}(t) < 1$) and $1 - \sum_{j' \in \mathcal{J}} p_{j,j'}(t)$ represents the stationary probability that a type-j resource leaves the human capital supply chain through attrition. The corresponding human capital evolution one-step transition probability matrix is given by $\mathbf{P}(t) \equiv [p_{j,j'}(t)]_{j,j' \in \mathcal{J}}$. Note that by making the transition probability matrices and model vectors functions of the time interval t, our stochastic models support time-varying behaviors of various forms (including seasonal effects) for the evolution of human capital supply chain dynamics.

A wide variety of approaches to set the parameters of our stochastic temporal models were investigated as part of this project. Based on a considerable amount of historical data and information, we found that an approach which extracts the base model parameters, such as one-step transition probabilities, from this historical human capital supply chain data and information provided accurate predictions. In particular, let

$$\mathcal{N}_{j,j'}(t) := \text{the number of transitions from state } j \text{ to state } j' \text{ over the time interval } [t, t+1),$$

for all $j, j' \in \mathcal{J}$. Then the elements of $\boldsymbol{P}(t)$ can be calculated as

$$p_{j,j'}(t) \;=\; \frac{\mathcal{N}_{j,j'}(t)}{\sum_{i\in\mathcal{J}} \mathcal{N}_{j,i}(t) + a_j(t)}, \qquad \text{ensuring that} \qquad 1 - \sum_{j'\in\mathcal{J}} p_{j,j'}(t) \;=\; \frac{a_j(t)}{\sum_{i\in\mathcal{J}} \mathcal{N}_{j,i}(t) + a_j(t)}.$$

Other model parameters can be calculated from historical data/information in an analogous manner. Our experience with this approach for predicting future resource and skill capacity levels has been in excellent agreement with realized capacity levels over time during the past couple of years.

It is important to note that, in a few cases, we needed to slightly adjust our use of the data in the above approach to maintain a high degree of accuracy in the human resource capacity estimates. This was essentially due to two causes: events that were not representative of actual business trends (e.g., one-time acquisition events) or a lack of statistical confidence in the number of samples from the data for some states. In the case of one-time events, we worked with subject matter experts to filter out the impact of these events from the data so that $\mathcal{N}_{j,j'}(t)$ properly represented the number of transitions with respect to future business trends. An analogous approach was taken for other events that might bias the statistics to deviate from actual business trends, including the use of business rules to factor out, as a representative example, errors resulting from human data entry mistakes. In the case of small sample sizes, we leverage standard statistical tests on each $\mathcal{N}_{j,j'}(t)$ to first verify a sufficiently large sample to guarantee a desired margin of error. Whenever this statistical test fails, we aggregate similar skills into a combined type $\hat{j}$ to obtain $\mathcal{N}_{\hat{j},j'}(t)$ whose sample size is within the desired margin of error and then, upon solving the resulting stochastic temporal model, we probabilistically distribute the estimated human resource capacities for type $\hat{j}$ into its constituent skills j.

Recall that $h_j(t)$ represents the expected flow of human resources into state j due to hiring over $[t, t+1)$ and that $y_j(t)$ represents the expected number of resources in state j at the beginning of this time interval. Then the net dynamics for human resources of type-j over each time interval $[t, t+1)$ are given by

$$\sum_{j'\neq j} y_{j'}(t) p_{j',j}(t) \;:\; \text{the transitions into state } j \text{ from other human resource types over the time interval;}$$

$$y_j(t) \sum_{j'\neq j} p_{j,j'}(t) \;:\; \text{the transitions to other human resource types from state } j \text{ over the time interval;}$$

$y_j(t)(1-\sum_{j'} p_{j,j'}(t))$: the outflow of human resources from state j due to attrition over the time interval.

More precisely, for $t=0,1,\dots,T-1$, we have

$$\begin{aligned} y_j(t+1) &= y_j(t) + h_j(t) + \sum_{j' \in \mathcal{J}: j' \neq j} y_{j'}(t) p_{j',j}(t) - y_j(t) \sum_{j' \in \mathcal{J}: j' \neq j} p_{j,j'}(t) - y_j(t)(1 - \sum_{j' \in \mathcal{J}} p_{j,j'}(t)), \\ &= h_j(t) + \sum_{j' \in \mathcal{J}} y_{j'}(t) p_{j',j}(t), \end{aligned}$$

or in matrix form (using column vector notation)

$$\mathbf{y}(t+1) = \mathbf{h}(t) + \mathbf{P}(t)\mathbf{y}(t). \tag{10}$$

Upon iterating this equation for every $s=1,\dots,T$, it follows that

$$\mathbf{y}(s) = \sum_{t=0}^{s-1} \left(\prod_{t'=t+1}^{s-1} \mathbf{P}(t') \right) \mathbf{h}(t) + \left(\prod_{t=0}^{s-1} \mathbf{P}(t) \right) \mathbf{y}(0),$$

from which we obtain the terminal human capital supply chain state vector

$$\mathbf{y}(T) = \sum_{t=0}^{T-1} \left(\prod_{t'=t+1}^{T-1} \mathbf{P}(t') \right) \mathbf{h}(t) + \left(\prod_{t=0}^{T-1} \mathbf{P}(t) \right) \mathbf{y}(0).$$

Stochastic Evolution Optimization

Although discrete-time, multidimensional stochastic processes model the dynamics and complexities of human capital evolution, our goal is to determine the investment decisions and policies that influence human capital supply chain dynamics in desired directions over the planning horizon. The driving objective is to maximize expected profit across all periods, where revenues depend upon demand and costs include maintaining and achieving the skill capacity levels, noting that maximizing expected revenue or minimizing expected cost subject to constraints on gross profit margin or on cost and revenue targets is included within our solution framework and methodology.

To this end, we associate costs and rewards with each state of our stochastic supply evolution models. Namely, the costs and rewards for each state are captured as functionals of the number of people in the state and the amount of time they spend in the state. For example, the per-period costs for C programmers at a specific experience and proficiency level include the corresponding salary and benefits (e.g., medical, pension) as a function of the state population, with the analogous per-period rewards including revenues driven through service delivery as a function of the population and demand for this state. Cost and reward functions are also associated with transitions between states, where the costs and rewards for transitioning from one state to another is a function of the states involved and the number of people making the transition. For example, the per-period costs and rewards for a set of C programmers becoming equally proficient in Java programming include training costs and related service-delivery revenues. Our stochastic supply evolution optimization models associate lead times for any available actions taken with respect to each state transition (e.g., hiring). These lead times capture the delays between the time the action is initiated (e.g., hiring starts) and the time the result of the action is actually realized (e.g., new employee comes on board). Note that adjusting transition probabilities between states within the corresponding stochastic decision process represents investments, policies, and actions such as training and promotion.

Then, in addition to skill composition trajectory, our stochastic optimization models characterize the evolution of expected cost, revenue, and related financial metrics of the human capital supply chain over time as functionals of the discrete-time stochastic process. Specifically, let us define

$$\begin{aligned} c_j(t) &:= \text{the expected cost (e.g., salaries, benefits) of human resources of type } j \text{ at time } t, \\ \mathbf{c}(t) &:= (c_1(t), \dots, c_{|\mathcal{J}|}(t)), \text{ the human capital supply chain cost vector at time } t, \end{aligned}$$

where $j \in \mathcal{J}$ and $t = 0, \dots, T$, and denote by $K(t)$ the expected total cumulative costs of all human resources over the time interval $[0, t)$, $t = 1, 2, \dots, T$. We then have $K(T) = \sum_{t=1}^{T} [\mathbf{c}(t) \cdot \mathbf{y}(t)]$, where $[\mathbf{c}(t) \cdot \mathbf{y}(t)] = c_1(t)y_1(t) + \ldots + c_{|\mathcal{J}|}(t)y_{|\mathcal{J}|}(t)$ under appropriate independence assumptions. The expected cumulative revenues are obtained in an analogous manner where the expected revenue for each period t is a function of $\mathbf{y}(t)$, subject to the componentwise demand constraints. Expected profit over the planning horizon is computed as a function of these revenue and cost metrics, including factors such as productivity and efficiency levels that depend upon human capital type, workload, and utilization.

We therefore have also developed solutions for a corresponding stochastic optimization problem based on our stochastic temporal models to determine the best evolution of the human capital supply chain over time in order to maximize business performance. To this end, we rewrite the system dynamics equation (10) into the following discrete-time linear dynamical system:

$$\mathbf{y}(t+1) = \mathbf{y}(t) + \mathbf{B}(t)\mathbf{u}(t), \quad \text{where} \quad \mathbf{B}(t) = \begin{bmatrix} \tilde{\mathbf{B}}(t) \ \mathbf{I}_n \ -\mathbf{I}_n \end{bmatrix}, \ \mathbf{u}(t) = \begin{bmatrix} \mathbf{P}^{num}(t) \\ \mathbf{h}(t) \\ \mathbf{a}(t) \end{bmatrix}. \tag{11}$$

Here, the $n \times n^2$ matrix $\tilde{\mathbf{B}}(t)$ captures the sparsity patterns of from the one-step transition probability matrix $\mathbf{P}(t)$, and the $n^2 \times 1$ vector $\mathbf{P}^{num}(t)$ is the vector of decision variables for transitions between states. We refer to $\mathbf{y}(t)$ ($\mathbf{u}(t)$) as the state (decision) vectors of the dynamical system at time t. Our goal is to choose the decision vectors $\mathbf{u}(0), \dots, \mathbf{u}(T-1)$ that maximize business performance over the planning horizon T.

We consider a general objective function of the form $\sum_{t=0}^{T-1} [\mathbf{c}(t+1) \cdot \mathbf{y}(t+1) + \mathbf{d}(t) \cdot \mathbf{u}(t)]$, where $\mathbf{c}(t)$ and $\mathbf{d}(t)$ are respectively the state and decision costs at time t. Because this objective is linear in the state and decision vectors for the model instances herein, we formulate the minimization problem as a linear program with decision variable vector $\mathbf{z}$ and weight vector $\mathbf{w}$ given by

$$\mathbf{z} = \begin{bmatrix} \mathbf{y}(1) \\ \vdots \\ \mathbf{y}(T) \\ \mathbf{u}(0) \\ \vdots \\ \mathbf{u}(T-1) \end{bmatrix}; \qquad \mathbf{w} = \begin{bmatrix} \mathbf{c}(1) \\ \vdots \\ \mathbf{c}(T) \\ \mathbf{d}(0) \\ \vdots \\ \mathbf{d}(T-1) \end{bmatrix}.$$

Our business objective is to maximize expected profit subject to forecasted demand for resource skills over the planning horizon. We also considered in this project related optimization formulations that include maximizing revenue or minimizing cost subject to constraints on the gross profit margin or on cost/revenue targets. Hence, the vector $\mathbf{c}(t)$ corresponds to the negative profit contributed by each resource of type j present in the system at time t, denoted by $-\mathcal{P}_j(t)$. It is important to note that these profit functions include factors such as productivity and efficiency levels that depend upon resource type, workload and utilization.

The linear program has two different forms of constraints, one on the state vectors due to system dynamics and another on the control vectors due to physical considerations. We assume the initial state vector $\mathbf{y}(0)$, weight vector $\mathbf{w}$ and set of demand forecast vectors $\mathbf{dem}(t)$ to be given. The system dynamics constraints dictate that equation (11) holds, for each $t = 0, \ldots, T-1$, which we compactly write as

$$\mathbf{M}\mathbf{z} = \tilde{\mathbf{y}}(0)$$

where

$$\mathbf{M} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & -\mathbf{B}(0) & \mathbf{0} & \ldots & \mathbf{0} \\ -\mathbf{I}_n & \mathbf{I}_n & \ldots & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{B}(1) & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & -\mathbf{I}_n & \mathbf{I}_n & \mathbf{0} & \mathbf{0} & \ldots & -\mathbf{B}(T-1) \end{bmatrix}, \qquad \tilde{\mathbf{y}}(0) = \begin{bmatrix} \mathbf{y}(0) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}.$$

For each resource type, the total internal flow of resources within the system should be balanced, i.e., sum to zero. More precisely, for each time t and appropriately defined matrix $\mathbf{Q}(t)$, the vector $\mathbf{u}(t)$ should satisfy

$$\mathbf{Q}(t)\mathbf{u}(t) = \mathbf{0}_{T\times 1}, \qquad \text{rewritten as} \qquad \underbrace{\begin{bmatrix} \mathbf{Q}(0) & \cdots & \mathbf{0}_{n\times(n^2+2n)} \\ \mathbf{0}_{n\times(n^2+2n)} & \ddots & \mathbf{0}_{n\times(n^2+2n)} \\ \mathbf{0}_{n\times(n^2+2n)} & \cdots & \mathbf{Q}(T-1) \end{bmatrix}}_{\tilde{\mathbf{Q}}} \begin{bmatrix} \mathbf{u}(0) \\ \vdots \\ \mathbf{u}(T-1) \end{bmatrix} = \mathbf{0}_{nT\times 1},$$

or in standard form

$$\left[\mathbf{0}_{nT\times nT}\ \tilde{\mathbf{Q}}\right] \mathbf{z} = \mathbf{0}_{nT\times 1}.$$

To simplify the presentation, let us initially suppose that the resource staffing levels at time t are limited to not exceed the demand forecast $\mathbf{dem}(t)$. We then have some inequality constraints that arise from physical considerations, which clearly require $\mathbf{z} \geq \mathbf{0}$. The total outflow of resources of type j also has to be less than the current number of type-j resources, which we express by an inequality of the form $\tilde{\mathbf{R}}\mathbf{z} \leq \tilde{\mathbf{y}}(0)$.

We now have the complete specification of our linear program, which can be expressed as

$$\begin{array}{cc} \min\limits_{\mathbf{z}} & \mathbf{w} \cdot \mathbf{z} \\ \text{subject to} & \mathbf{M}\mathbf{z} = \tilde{\mathbf{y}}(0); \quad [\mathbf{0} \quad \tilde{\mathbf{Q}}]\,\mathbf{z} = \mathbf{0}; \quad \mathbf{z} \geq \mathbf{0}; \\ & \tilde{\mathbf{R}}\mathbf{z} \leq \tilde{\mathbf{y}}(0); \qquad [\mathbf{I}_{nT} \quad \mathbf{0}_{nT \times (n^2+2n)T}]\,\mathbf{z} \leq \begin{bmatrix} \mathbf{dem}(1) \\ \vdots \\ \mathbf{dem}(T) \end{bmatrix}. \end{array}$$

The output of this linear program provides an optimal vector $\mathbf{z}^*$ from which we can cull the relevant information regarding optimal state and decision vectors $\mathbf{y}^*(1), \ldots, \mathbf{y}^*(T)$ and $\mathbf{u}^*(1), \ldots, \mathbf{u}^*(T)$, respectively.

Turning now to our general formulation, various additional constraints can be incorporated by appending or modifying the inequality constraints of the linear program. This includes using historical data on attrition to lower bound the outflow from a resource type. An important generalization used in this project is to allow the retention of surplus human capital resources in the optimization state space. Initially, to elucidate the exposition, our stated objective drove staffing levels to a profit-maximizing optimal point which is upper bounded by the demand forecast. However, given the costs of staffing decisions (e.g., hiring and training), the time-varying system dynamics over the planning horizon and the nature of statistical demand forecasts, it is often preferable to retain resources in excess of demand estimates. Without provisions for the retention of surplus human capital, the optimal decision for a scenario with temporary dips in demand lasting one quarter might very well be to allow attrition during that quarter only to rehire resources of the same type in the very next quarter. This is clearly an undesirable solution from both a cost and human resource perspective.

In our general optimization formulation supporting human capital retention, we introduce an additional variable, $s_j(t)$, for the surplus of resource-type j. Hence, the state vector $\mathbf{y}(t)$ is comprised of two parts: the utilized portion of human resources $\mathbf{y}(t) - \mathbf{s}(t)$ and the surplus portion $\mathbf{s}(t)$. Then the preceding formulation of the linear program is modified, first by expanding the decision variable $\mathbf{z}$ to include the surplus vectors $\mathbf{s}(1), \ldots, \mathbf{s}(T)$. The weight vector will correspondingly be appended with the revenue vectors $\mathbf{rev}(1), \ldots, \mathbf{rev}(T)$, where $\mathbf{rev}(t)$ is the vector of expected revenues earned by the different resource types in period t. Finally, instead of capping the total number of resources of each type by the demand forecast, we only upper bound the number of utilized resources, namely $\mathbf{y}(t) - \mathbf{s}(t) \leq \mathbf{dem}(t), t = 1, \ldots, T$.

Illustrative Examples

To illustrate the complexities of supply evolution and optimization and the benefits of our approach, consider a simple application scenario starting with the evolution of a single skill. Using historical data, we predict a significant increase in attrition in November and December because of baby-boomer retirements.

Furthermore, the demand forecast for this skill is relatively steady for the subsequent months (see the black curve in Figure 4). This suggests that there will be a significant shortage in this skill relative to demand under these retirement predictions. To mitigate this problem, we investigate the possibility of incentivizing some of this attrition to postpone retirement until enough people with this skill can be brought on board. We model the behavioral responses of baby-boomer retirees as a concave function of increasing incentives where the higher the incentives, the larger the fraction of people who are willing to postpone retirement, with diminishing returns. For this representative scenario, our supply evolution optimization solution determines the set of investments to maximize profit relative to forecasted demand (see the red curve in Figure 4).

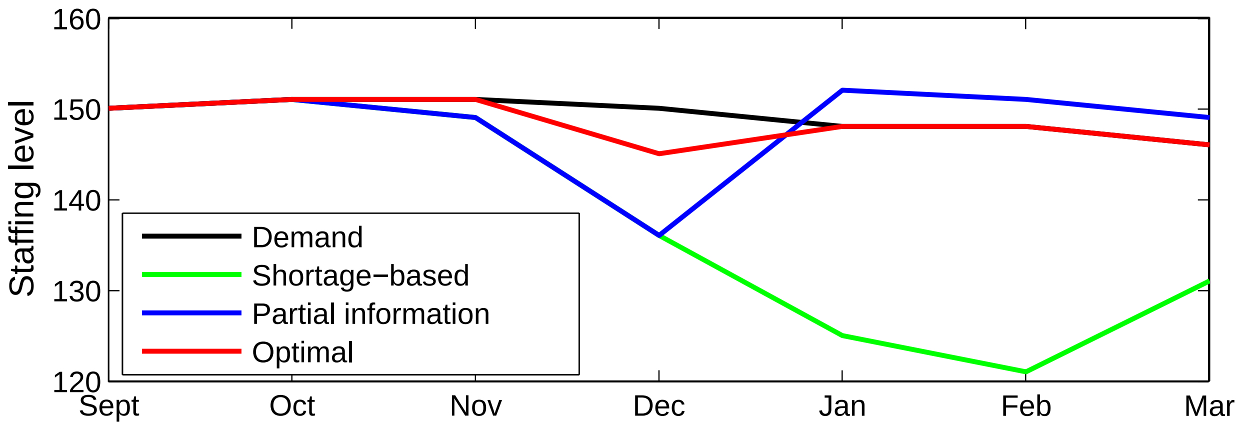


Figure 4 **The graph shows an example of supply evolution and optimization.**

In contrast, many organizations would never realize the upcoming shortage because of future retirements under this scenario until it is too late; thus, they will lose considerable revenue while trying to catch up over the hiring lead time, as the green curve in Figure 4 shows. Somewhat more enlightened organizations might realize this upcoming shortage in advance. However, they will hire enough people to fill this shortage as they believe to be the case in September without understanding that transitions into this skill will naturally occur over the next few months; thus, they will hire too many (too few) people and create a future skill overage (shortage) problem (see the blue curve in Figure 4).

Skill Shortage & Overage Models and Optimization

We have now obtained the skill capacity targets on the demand side from our risk-based capacity planning optimization and the multi-skill human resources on the supply side from our supply evolution & optimization. We next take these targets and expected resources as input and determine the optimal matching of the latter to the former. The overall planning horizon (on the order of a year) consists of a sequence of coarse subintervals, each on the order of a month or quarter and involving a shortage & overage model under a fixed set of parameters that changes from one subinterval to the next according to a general stochastic modulation process. We then formulate and solve across these stationary periods the skill shortage & overage optimization problem to determine the matching of skill capacity targets and multi-skill human resources that minimizes a weighted sum of the expected shortages and overages for each period. Any existing or pre-determined assignments of skill resource capacities can be incorporated as constraints in the shortage & overage optimization problem and then our solution optimally matches the remaining multi-skill resources to the remaining demand, both over the entire planning horizon. The solution of our shortage & overage optimization also computes the skill shortages and overages under the optimal supply-demand matching.

We capture the fundamental aspects of this optimization problem for each stationary period using a linear programming formulation. Recall that $\mathcal{I}$ denotes the set of skills, labeled by i, and that $\mathcal{J}$ denotes the family of subsets of the set of skills $\mathcal{I}$ possessed by human resources, labeled by j. Next we define

$$\begin{aligned}
\mathcal{T}_i &:= \text{the capacity target for skill } i \in \mathcal{I}, \\
\mathcal{A}_j &:= \text{the expected number of available human resources with skill subset } j \in \mathcal{J}, \\
\mathcal{S}_i &:= \text{the expected shortage for skill } i \in \mathcal{I}, \\
\mathcal{O}_i &:= \text{the expected overage for skill } i \in \mathcal{I}, \\
\mathcal{M}_{i,j} &:= \text{the amount of human resources capable of employing skill subset } j \text{ that are matched to employ skill } i.
\end{aligned}$$

Then our objective is to determine the matchings $\mathcal{M}_{i,j}$ of multi-skill human resources $\mathcal{A}_j$ to skill capacity targets $\mathcal{T}_i$ that minimize a weighted sum of the skill shortages $\mathcal{S}_i$ and overages $\mathcal{O}_i$, $i \in \mathcal{I}$, $j \in \mathcal{J}$. We cannot allow the matching of human resources to employ a skill that they do not possess, and therefore we enforce the constraint $\mathcal{M}_{i,j} = 0$ for all $i \in \mathcal{I}$ which are not an element of the subset $j \in \mathcal{J}$, with $\mathcal{M}_{i,j} \geq 0$ otherwise. All human resources need to be matched, thus implying $\sum_{i=1}^{|\mathcal{I}|} \mathcal{M}_{i,j} = \mathcal{A}_j$ for all $j \in \mathcal{J}$. The shortages and overages must satisfy a balance equation in which the sum of skill matchings and expected shortage equate with the sum of corresponding skill targets and expected overage, or more precisely

$$\sum_{j \in \mathcal{J}} \mathcal{M}_{i,j} + \mathcal{S}_i = \mathcal{T}_i + \mathcal{O}_i, \qquad \forall i \in \mathcal{I}.$$

Our use of weights for each skill shortage $\mathcal{S}_i$ and overage $\mathcal{O}_i$ in the objective function helps to facilitate skill priorities and importance factors when matching multi-skill human resources to the skill capacity targets. For example, it might be desirable to push the optimal matching toward “hot” skills (to reduce shortages of such skills), all else being equal, and similarly to push the optimal matching away from “commodity” skills (to reduce overages of such skills). In addition, it can be desirable to include preferences among the subsets of skills in multi-skill supply-demand matching. To this end, we introduce an additional family of variables $\mathcal{Z}_{i,j}$ that represent the remaining amount of human resources capable of employing skill subset $j \in \mathcal{J}$ that are not matched to employ skill $i \in \mathcal{I}$. This causes the above set of matching constraints to become $\sum_{i=1}^{|\mathcal{I}|} \mathcal{M}_{i,j} + \mathcal{Z}_{i,j} = \mathcal{A}_j$, $\forall j \in \mathcal{J}$. We also need to enforce the constraint $\mathcal{Z}_{i,j} = 0$ for all $i \in \mathcal{I}$ that are not an element of the subset $j \in \mathcal{J}$, with $\mathcal{Z}_{i,j} \geq 0$ otherwise. Then, whenever subset j is preferred over subset j' for skill i, we add $M\mathcal{Z}_{i,j} + \mathcal{Z}_{i,j'}$ in the objective function where M is a large real number according to the big-M method for solving linear programs; see, e.g., Bertsimas and Tsitsiklis (1997). More precisely, our shortage & overage optimization with preferences for each stationary subinterval consist of solving the linear program

$$\min_{\mathcal{M}_{i,j}} \sum_{i=1}^{|\mathcal{I}|} (w_i^{\mathcal{S}} \mathcal{S}_i + w_i^{\mathcal{O}} \mathcal{O}_i) + \sum_{i,j,j'} (M_{i,j,j'} \mathcal{Z}_{i,j} + \mathcal{Z}_{i,j'}), \tag{12}$$

subject to the foregoing constraints, where $w_i^{\mathcal{S}}$ and $w_i^{\mathcal{O}}$ are weights for the expected shortages and overages associated with skill $i \in \mathcal{I}$, respectively. Of course, setting $\mathcal{Z}_{i,j}$ equal to 0 for all i, j yields our shortage & overage model and optimization without preferences (but with weights). The solution to our overall multi-skill shortage & overage optimization problem across multiple stationary subintervals under a general stochastic modulation process is then obtained by solving the corresponding dynamic program.

Skill Shortage & Overage Management

Once all expected skill shortages and overages have been determined by solving the multi-period version of the linear program (12), we lastly consider the set of business investment decisions to manage these expected shortages and overages. Our approach is based on a combination of each of the optimization capabilities of OTM, namely risk-based capacity planning, supply evolution & optimization, and skill shortage & overage optimization. More specifically, an initial first-phase application through our end-to-end solution process results in: (a) risk-based capacity planning providing the ideal resource capacity targets based on business/economic considerations and the dynamics of the human capacity supply chain within each quarter; (b) supply evolution & optimization providing the estimated human resource capacities and required investments across future quarters to realize these capacity levels; and (c) multi-skill shortage & overage optimization providing the optimal matching of expected multi-skill human resources to skill capacity targets.

Note that the intra-quarter dynamics are captured by each instance of our risk-based optimization and the inter-quarter dynamics are directly captured by our supply evolution & optimization, while both the intra-quarter and inter-quarter dynamics are respectively captured by single- and multiple-period instances of our shortage & overage optimization. This partitioning of the overall OTM solution is based on various business and organizational aspects of the ITS human capital supply chain. Given these dependencies and interrelationships among the different problems and solutions, we develop a subsequent-phase iterative procedure that combines all three optimization components. First, the costs of hiring, training and retaining are additionally incorporated into the risk-based capacity planning optimization. Then its skill capacity output and the skill composition output of the supply evolution & optimization are optimally matched via the skill shortage & overage optimization. This skill matching output is then fed in turn to both the risk-based capacity planning optimization and the supply evolution & optimization for updates on the business investment decisions to act upon. The entire process is then repeated until convergence to equilibrium (steady state).

The general formulation of our risk-based capacity planning optimization incorporates the full set of resource capacity costs which include the costs of adjusting human resource capacity levels as well as the costs of maintaining the desired levels of resource capacity. In this general formulation, the objective (6) is augmented to include the costs for hiring additional capacity possessing skill i, training existing capacity to acquire skill i, and retaining existing capacity with skill i, where existing capacity takes into account estimation of attrition and internal transitions from our supply evolution & optimization. The capacity vector $\mathbf{C}$ used in the objective function and in the stationary loss network is also augmented to reflect the sum of existing and retained capacity for skill i and new capacity for skill i based on hiring and training. More precisely, the objective for our general risk-based capacity planning formulation is given by

$$\max_{\mathbf{C}} \ \sum_{r=1}^{|\mathcal{K}|} u_k \, (1 - L_k) \, \nu_k \ - \ \sum_{i=1}^{|\mathcal{I}|} v_i \, C_i \ - \ \sum_{i=1}^{|\mathcal{I}|} v_i^H \, C_i^H \ - \ \sum_{i=1}^{|\mathcal{I}|} v_i^T \, C_i^T \ - \ \sum_{i=1}^{|\mathcal{I}|} v_i^R \, C_i^R,$$

where $C_i = C_i^E + C_i^H + C_i^T + C_i^R$, v_i^H is the cost rate for hiring skill i capacity, v_i^T is the cost rate for training skill i capacity, v_i^R is the cost rate for retaining skill i capacity, and C_i^E, C_i^H, C_i^T and C_i^R is the amount of capacity existing, hired, trained and retained for skill i, respectively. In addition to maximizing profits, our risk-based capacity planning optimization can be used to maximize revenue or minimize cost subject to constraints on the gross profit margin or on cost/revenue targets, as previously noted. The constraints of the optimization problem can also include a budget for hiring costs, training costs, and retaining costs, either jointly or separately, which has proven especially useful in practice where various business investments are managed by different organizations each having their own budget.

OTM Business Benefits

The OTM integrated suite of OR models & methods has been implemented and successfully deployed around the world to support the management & planning of the ITS human capital supply chain. In this section we discuss various aspects of the business benefits realized through the deployment of OTM over the past couple of years. We first summarize the vast set of validation studies over this period which demonstrates excellent agreement between the OR models & methods of OTM and real-world ITS business outcomes. We then present a representative collection of business cases based on recent ITS data that highlight the significant benefits of our OTM solution methodology over previous approaches used by ITS, which includes business cases where the final end-to-end OTM solution provides a 200% improvement in expected profit. We conclude with reviews of the OTM impact on the ITS business over the past couple of years, which consistently show relative quarterly financial benefits commensurate with 2-4% of the ITS quarterly revenue targets, as well as efforts to support future IBM solution offerings (both products and services) in the marketplace.

Validation Studies

Our successful experience with the worldwide deployment of the OTM suite to manage and plan the ITS human capital supply chain includes a detailed analysis of the accuracy of each capability & component comprising the OTM solution methodology over the past couple of years. A feedback loop was instituted from the start to validate OTM predictions and results against actual business outcomes and available data from ITS. In addition, quarterly reviews have been conducted on a regular basis with ITS executives to track and evaluate the accuracy of the OTM end-to-end solution. Such validation of our OR models & methods throughout the OTM deployment has consistently demonstrated excellent agreement between the outputs of OTM modeling and optimization solutions and the corresponding business outcomes realized in practice.

Specifically, we have observed a typical accuracy range of 85-90% for our pipeline revenue forecasts, which represents an error reduction of more than 100% over forecasts computed using the win probabilities estimated by sales representatives. The accuracy of the overall resource demand in terms of total hours has been observed to be generally higher than 90% relative to quarterly revenue targets. Our experience demonstrates that the loss risk probability estimates predicted by the OTM risk-based capacity planning models have been within a few percentage points of ITS data on real-world engagement losses. In other words, we found that the business was in fact suffering engagement losses that matched the predictions from our risk-based capacity planning capability. We have observed the accuracy of the OTM supply evolution modeling capability to consistently be within a few percentage points in estimating the skill resource composition of organizations on the order of thousands or larger over planning horizons of up to one year. The accuracy of such estimates for smaller organizations on the order of hundreds were consistently found to be within 12% of the real-world evolution of skill resource composition over planning horizons of up to one year.

Business Cases

The OTM suite supports the ITS end-to-end human capital management & planning process through both the individual and collective real-world application of its innovative OR solution methodologies. This includes service engagement delivery, resource capacity planning, service engagement portfolio management, sales-delivery interlock, and higher level strategic planning. These business processes and related resource investment decisions are now based on the insights and information obtained via the advanced analysis and decision support capabilities of the OTM suite. Since its original deployment in 2008, OTM has helped ITS in making profitable and effective business decisions at different time scales. To highlight the benefits and successful use of OTM capabilities within ITS, in this subsection we briefly present a few business cases based on recent ITS data to which the various OTM capabilities have been applied. These include representative comparisons that demonstrate the benefits of our models & methods over the approaches used by ITS prior to the deployment of OTM.

Risk-Based Capacity Planning The OTM risk-based capacity planning capability, designed as a decision making aid for delivery executives, supports the detailed analysis and optimization of the complex financial/business risks and tradeoffs associated with different service delivery models and strategies, as well as the investigation of different policies and investments of *Human Resources* (HR) organizations to drive revenues while maintaining acceptable levels of costs and risks. This includes the stochastic analysis and optimization of business investments and capacity profiles that best balance the tradeoffs between the costs of maintaining and achieving resource capacity levels and the revenues of satisfying and growing service engagement demand. In practice, risk-based capacity planning can be applied in three different ways, thereby providing business users with the greatest insight and flexibility in determining the most efficient and effective manner in which to manage and operate their business. First, one can fix revenue targets in some way and then determine the optimal skill capacity levels that minimize cost while achieving the desired revenue targets (or gross profit margins). Second, one can fix capacity costs and then determine the optimal skill capacity levels that maximize revenue while achieving desired cost targets (or gross profit margins). The final application is to determine the optimal skill capacity levels that maximize profit (as originally described).

As a representative business case of the benefits obtained by applying this OTM capability within the ITS capacity planning process, for a quarterly planning of 132 skill types and 216 service engagement types, we observe that the optimal OTM risk-based capacity planning solution provides significant improvements over the deterministic solution approach previously used by ITS based on linear projections of the demand at the skill level (as described in more detail in subsequent sections). Specifically, in the cost-minimizing application, since both solutions guarantee the same revenue targets, the aggregate loss risk probability across all engagement types will also reach roughly the same levels. In stark contrast, the portfolios of skill capacity levels under the two solutions are quite different, with risk-based optimization providing a superior portfolio through lower loss risk probabilities for engagements with relatively high profit margins and lower capacity levels for engagements with relatively low profit margins. While the individual loss risk probabilities for different engagement types obtained under the two approaches vary by more than 15% in this business case, the risk-based optimization yields a solution with a 42% reduction in expected capacity costs. In the revenue-maximizing application, risk-based optimization tends to increase the capacity levels for those skills that are critical to engagements with relatively high profit margins or relatively high usage across a wide range of engagement types (with relatively good profit margins). Hence, the optimal risk-based solution renders reductions in both the aggregate loss risk probability level across all engagement types and the individual loss risk probabilities of most engagement types. The aggregate loss risk probability is 5% under risk-based optimization, as compared to 23% under the previous linear projection approach. This in turn yields a 35% increase in expected revenue. In the profit-maximizing application, risk-based optimization determines the ideal balance of the tradeoffs between engagement revenues and capacity costs. We observe that this instance of risk-based optimization provides a 100% increase in expected profit over that obtained under the previous ITS approach. While our risk-based solution yields lower expected revenue for this business case, it also renders significantly lower capacity levels than the previous linear projection approach, both for individual skill capacity levels and in the aggregate. These cost reductions in turn yield a significant increase in expected profit.

Supply Evolution and Dynamics The supply evolution & optimization capability provides decision making support for HR organizations, allowing users to analyze historical human capital trends and dynamics, perform predictive modeling of future human capital dynamics, model future scenarios to understand the effect of different investments/policies on human capital trends/dynamics, and optimize strategic decisions relative to business goals. We next describe results comparing the estimated evolution of human capital from OTM capabilities with that provided by previous ITS approaches. Specifically, we consider the staffing levels predicted by two different forms of OTM supply evolution & optimization: (a) supply evolution models (SEM) using historical transition information; and (b) supply evolution & optimization (SEO) using the costs of investments together with SEM and forecasted demand. The staffing levels from these two OTM capabilities are then compared against the corresponding approaches previously used by ITS: (a) linear projections of growth and shrinkage in resource skills based on historical hiring and attrition trends for each skill in previous quarters; (b) myopic policies that adjust skill capacity levels to address a certain fraction of the skill shortages and overages relative to forecasted demand.

We consider a particular aspect of the ITS business and perform quarterly evolutions of the ITS resource supply. As a representative business case, we use an ITS data set comprised of around 500 skills (including different levels of competency in each area). For SEM, we calculated historical hiring, attrition and transition (from one skill type to another) amounts from quarter to quarter and applied them to parameterize the OTM supply evolution models for future quarters. To demonstrate the richer dynamics captured by SEM, we contrast the predicted staffing levels for different skills with that achieved by the deterministic approach previously used by ITS based on historical hiring/attrition trends. Figure 5 presents the staffing levels for three skills looking eight quarters out into the future (with current quarter Q0) under both approaches. We observe, on the one hand, that SEM tends to provide better predictions of future staffing levels by preserving various business cycles and dynamics as is evidenced by the moderate degrees of periodicity and growth in the supply predictions. SEM is able to faithfully reproduce intricate business cycles and staffing patterns because it captures the essential features and corresponding parameters of the dynamics of the human capital supply chain. On the other hand, the previously used ITS approach based on linear projections leads to over- and under-staffing because it ignores important dynamics such as the probabilities associated with transitions between states of the human capital supply chain; refer to, e.g., skill 3 in Q4 and skill 1 in Q4 in Figure 5.

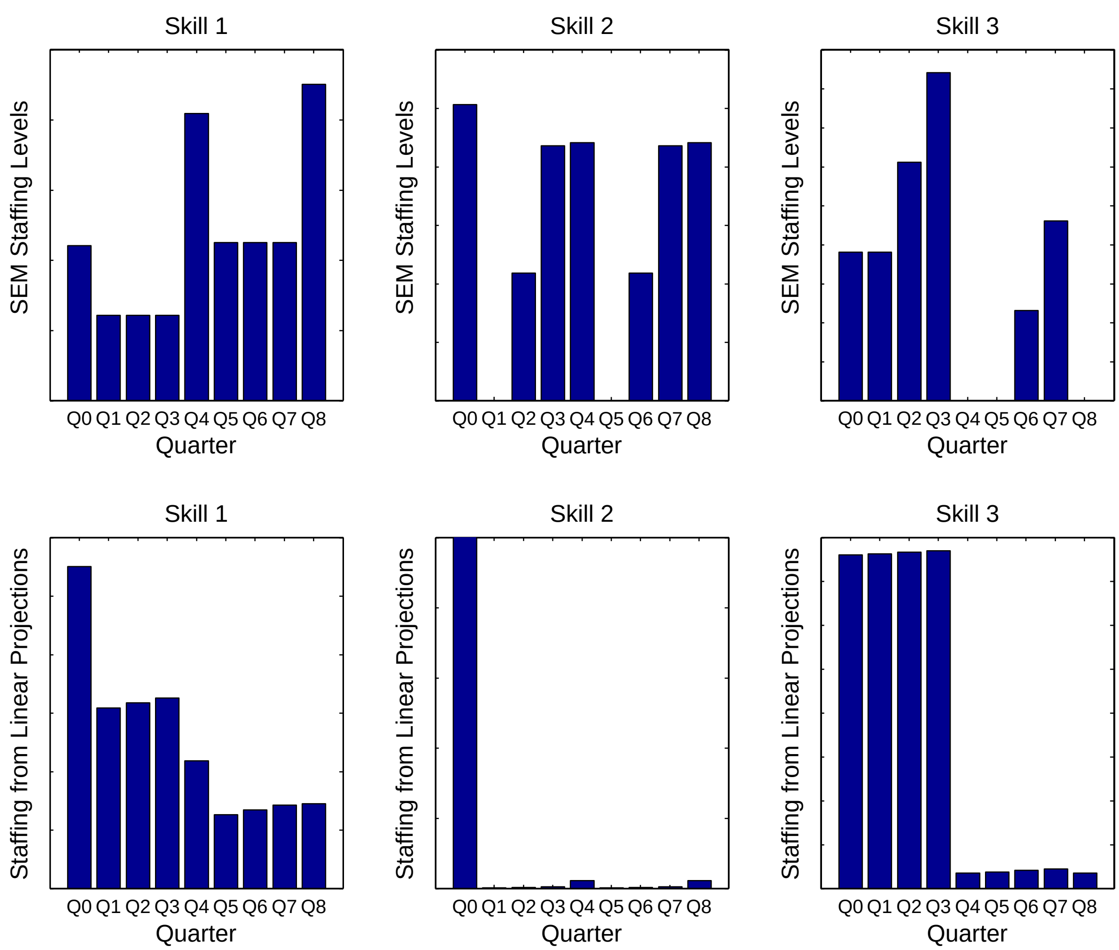
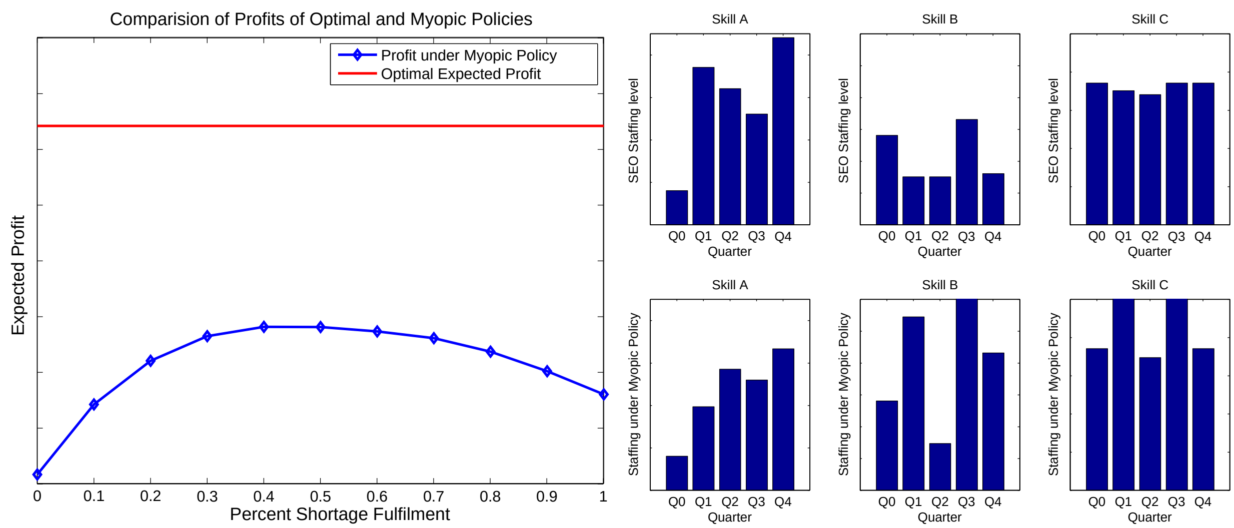


Figure 5 **The graphs compare supply evolution models and linear projections.**

Finally, for a given demand forecast over the next four quarters, we present results that contrast the profit-optimal staffing decisions obtained from SEO with those provided by the previous ITS myopic approach that adjusts capacity levels up to a certain fraction $\alpha = 0, 0.1, \ldots, 0.9, 1$ of the skill shortages and overages. The α-myopic policy exhibits an interesting concave behavior as shown in Figure 6(a). Even under the best setting for $\alpha \approx 0.4$, SEO provides an increase of close to 50% in profit. As explained below, capacity level adjustments to address skill shortages and overages under the previous ITS approach tended to be conservative and consist of relatively small values of α. This instance of the business case then indicates that SEO would provide an even larger increase in expected profit of around 80% in comparison with the corresponding α-myopic policy. These dramatic improvements in expected profit are related to our earlier observations that unnecessary responses to temporary dips in demand can be counterproductive relative to the various cost and revenue structures of the business. For example, Figure 6(b) depicts significant over- and under-staffing under the previous ITS myopic policy; refer to, e.g., skill B in $Q1/Q3$ and skill A in $Q1/Q4$, respectively. This, in turn, entails unnecessary investment costs and reduced profit.



(a) Profit Results (b) Staffing Levels

Figure 6 **The graphs compare supply evolution optimization and an α-myopic policy, $\alpha \in \{0, 0.1, \ldots, 0.9, 1\}$.**

Skill Shortages and Overages OTM multi-skill shortage & overage optimization provides an efficient and accurate solution for determining the best matching of multi-skill supply resources against capacity targets in order to minimize skill shortages and overages, where there are on the order of thousands of different skills (including different competency levels) in ITS. In the end-to-end OTM framework and methodology, once capacity targets for each skill have been determined for each quarter via risk-based capacity planning optimization under forecasted demand, then multi-skill shortage & overage analysis is performed against current and future quarterly supply resources estimated via supply evolution & optimization. Using ITS data as part of a business case, we compare the optimal solution from the OTM multi-skill shortage & overage capability against results from the previously used ITS approach, where multi-skill resources are simply partitioned into individual skills based on historical averages of their skill deployment in the prior quarter (as described in more detail in subsequent sections). A representative business case with 950 employees and 307 skills shows that the OTM shortage & overage optimization yields a 25% reduction in skill shortages, 42% reduction in skill overages and 32% reduction in the sum of skill shortages and overages.

Skill Shortage & Overage Management We have seen that the ideal capacity target output from the risk-based optimization and the resource skill output from the supply evolution & optimization are both input to the shortage & overage optimization. The resulting skill shortage and overage output is then fed back into the risk-based capacity optimization and the supply evolution & optimization. Here, the costs of skill capacities in the risk-based optimization should include not only the costs for maintaining these resource capacity levels, but also the costs of achieving these resource capacity levels (e.g., the costs of hiring and training decisions), which involve the outputs of both the supply evolution & optimization and the shortage & overage optimization. Such interactions have been realized in practice by an iterative methodology based on these three OTM capabilities, in which the costs of hiring, training and retaining are incorporated into the risk-based optimization, then its output and the resources output from the supply evolution & optimization are optimally matched by the shortage & overage optimization, whose output is in turn fed into the supply evolution & optimization for HR investment updates. The process is then repeated until equilibrium (steady state) is reached.

Applying this iterative OTM skills shortage & overage management capability to the above business case, we observe that the overall end-to-end OTM solution identifies the most critical skills in which the business needs to invest, with respect to the impact on both expected revenues through higher capacity levels and expected costs to realize these capacity levels. Each iteration of this process adjusts business investment decisions to move skill capacity levels to their optimum point according to the criticality of the skill. The final OTM solution provides a 200% improvement in expected profit over that obtained via the collection of ITS approaches used prior to the OTM deployment as previously noted. On the one hand, this overall solution yields expected revenues that are only 7% below that obtained as part of the first-phase (idealized) application of risk-based capacity planning optimization in which the costs to realize the corresponding ideal resource capacity targets are ignored. On the other hand, when such costs are factored into the analysis, the idealized risk-based capacity planning solution renders an expected profit improvement of only 177% in comparison with the previous ITS approaches. The increase to relative expected profit improvements of 200% is the result of our iterative OTM skills shortage & overage management methodology.

Business Impact

Our experience with the business impact of the worldwide deployment of the OTM suite to manage and plan the ITS human capital supply chain has been very successful. The OTM suite serves as a central analytics and information base for all organizations and business users within ITS, integrating information from multiple sources and creating reports for multiple business functions. Although the final outputs of OTM are optimal resource matchings, reports on shortages and overages, and business decision and investment recommendations, distinct components of the integrated suite have also been independently used in practice by different ITS business users to support various additional management and planning activities.

To track and evaluate the business impact of the OTM solution methodology, quarterly reviews have been performed on a regular basis with ITS executives. These estimates of business impact are based on comparisons with the previous approaches employed by ITS in which linear projections were used to forecast ongoing demand and pipeline demand (based on win probability inputs from sales representatives), first at the level of service engagement demand and then down to the level of individual skill demand. The previous ITS approach also consisted of partitioning the multi-skill resources among their individual skills based on historical averages of their skill deployment in the previous quarter as being representative of their likely skill deployment in the next quarter; e.g., if a person spent 60% of their time employing skill A and 40% of their time employing skill B last quarter, then that person was split into 0.6 and 0.4 of a skill A and B resource, respectively, for the upcoming quarter. (Given the dynamics of service engagement demand, however, this was neither an accurate representation of likely skill deployment in the next quarter nor the best way to deploy multi-skill resources in order to minimize shortages and overages.) Then skill shortages and overages were directly obtained as the difference between the supply and demand for each skill. This was followed by ad hoc investment decisions to hire and train resources in order to reduce a certain fraction of the skill shortage and overage estimates. A relatively small fraction was typically used due to perceptions of inaccuracies in these previous ITS approaches. In addition to its lack of OR models & methods, the previous ITS approaches resulted in a planning cycle that was very long and resource intensive. This is in stark contrast to the deployment of OTM which supported the ability to run the OR models & methods more frequently and in immediate response to unexpected events in order to accordingly adjust and fine-tune human capital management & planning investments and strategies.

The business impact of OTM on the effective and efficient management & planning of the ITS human capital supply chain has been significant. As part of the quarterly reviews with ITS executives, we have consistently found the reductions in skill shortages and overages obtained from our multi-skill shortage & overage optimization to generally fall within the range 10-80% and 30-150%, respectively, in comparison with the previous approach based on historical averages. Furthermore, detailed analyses have been performed over the past few years to estimate and review the business impact of OTM on the ITS human capital supply chain. These business benefits are in comparison with the corresponding results that would have been obtained under the ITS approaches used prior to the deployment of OTM. The methodology used for this comparison consists of estimating the revenue and cost impact of predicted skill shortages and overages as well as of the investment decisions made to address these shortages/overages, which was performed under both OTM and the previous ITS approaches. Specifically, these business impact computations were based on: (a) the increased costs for human resource investments and the decreased profits for lower utilization when there were skill overages; and (b) the increased costs for hiring and training and the decreased revenues for insufficient human resources to fulfill the demand when there were skill shortages. The initial set of business impact comparisons reviewed with executives on a quarterly basis demonstrated that the financial impact of OTM in its first year of deployment were above $11 million within a single quarter in the U.S.A. alone, in terms of cost savings and increased revenue. Similar results have been obtained as part of the worldwide OTM deployment over the past couple of years during which the findings of our business impact studies, reviewed on a quarterly basis with ITS executives, consistently demonstrated that the quarterly impact of combined relative cost and revenue benefits under the OTM solution were commensurate with 2-4% of ITS quarterly revenue targets over that under the previous ITS approaches.

In addition to the impact of OTM on the ITS business, we have developed prototype implementations of several OTM capabilities & components within IBM software platforms to support future IBM solution engagements (both products and services) in the marketplace. We also have been and are continuing to collaborate with various IBM product groups on the development of pilot solutions for a broad set of human capital management & planning clients, which are fully based on the core OTM models & methods.

Although this paper focused on one particular implementation of OTM, the possibilities for using the OTM framework and methodology in a broader application setting, beyond information technology services, are numerous. Concepts of labor and people are central to all services organizations, and companies in healthcare, finance and insurance, retail or public sector, are asking the very same questions addressed by the OTM models & methods. What is the outlook for our nursing workforce 3 years from now if we keep our current hiring, training and attrition policies? Will we have enough insurance agents with a particular set of skills to meet product demand? How much sales can we drive with the current salesforce? How many financial analysts should be hired next year to deliver on our growth objectives? As a result, the issue of human capital management & planning is becoming one of the most important factors on the agenda of any CEO, and the ability to manage skills and human resources more effectively and efficiently is becoming the critical driver of success for most services organizations, especially those with a large number of employees and diverse product, solution or project portfolios. Analysts research indicates that, despite the majority of organizations making significant financial investments in training, development and recruitment of their people, they spend up to an additional 8% of their total wages and salaries, on average, to manage human capital issues, many of which could be avoided or turned into bottom-line contributions by using more advanced solutions for human capital management. The integrated OTM suite is a novel first step in this direction, and in collaboration with IBM product groups we continue to enrich our models & methods to create a general human capital management & planning platform for a broader client base.

References

Anselmi, J., Y. Lu, M. Sharma, M. S. Squillante. 2009. Improved approximations for the Erlang loss model. *Queueing Systems Theory Appl.* **63**(1–4) 217–239.

Bertsimas, D., J. N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Athena Scientific, Belmont, MA.

Bhadra, S., Y. Lu, M. S. Squillante. 2007. Optimal capacity planning in stochastic loss networks with time-varying workloads. *Proc. ACM SIGMETRICS ’07*, San Diego. Association for Computing Machinery, New York, 227–238.

Brockmeyer, E., H. L. Halstrom, A. Jensen, eds. 1948. *The Life and Works of A. K. Erlang*. Academy of Technical Sciences, Copenhagen.

Erlang, A. K. 1917. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektroteknikeren* **13** 5–22.

Hu, J., B. K. Ray, M. Singh. 2007. Statistical methods for automated generation of service engagement staffing plans. *IBM J. Res. Development* **51**(3) 281–293.

Hu, J., M. Singh, A. Mojsilovic. 2008a. Categorization using semi-supervised clustering. *Proc. ICPR 2008*. Tampa, FL, 1–4.

Hu, J., M. Singh, A. Mojsilovic. 2008b. Using data mining for accurate resource and skill demand forecasting in services engagements. *Proc. KDD Workshop on Data Mining for Business Applications*. Las Vegas.

Jung, K., Y. Lu, D. Shah, M. Sharma, M. S. Squillante. 2008. Revisiting stochastic loss networks: Structures and algorithms. *Proc. ACM SIGMETRICS ’08*, Annapolis, Md. Association for Computing Machinery, New York, 407–418.

Kelly, F. P. 1991. Loss networks. *Ann. Appl. Probability* **1**(3) 319–378.

Lasserre, J. B. 2001. Global optimization with polynomials and the problems of moments. *SIAM J. Optim.* **11**(3) 796–817.

Louth, G., M. Mitzenmacher, F. P. Kelly. 1994. Computational complexity of loss networks. *Theoret. Comput. Sci.* **125**(1) 45–59.

Whitt, W. 1985. Blocking when service is required from several facilities simultaneously. *AT&T Bell Laboratories Tech. J.* **64**(8) 1807–1856.

Woods, R. A. 2009. Industry output and employment projections to 2018. *Monthly Labor Rev.* **132**(11) 53–81. Bureau of Labor Statistics, U.S. Department of Labor.