This article was downloaded by: [155.246.103.35] On: 25 March 2017, At: 18:22
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

OM Practice—Balancing Risk and Efficiency at a Major Commercial Bank

Sunder Kekre, Nicola Secomandi, Erkut Sönmez, Kenneth West,

To cite this article:
Sunder Kekre, Nicola Secomandi, Erkut Sönmez, Kenneth West, (2009) OM Practice—Balancing Risk and Efficiency at a Major Commercial Bank. Manufacturing & Service Operations Management 11(1):160-173. http://dx.doi.org/10.1287/msom.1070.0207

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2009, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 11, No. 1, Winter 2009, pp. 160–173
ISSN 1523-4614 | EISSN 1526-5498 | 09 | 1101 | 0160

DOI 10.1287/msom.1070.0207
© 2009 INFORMS

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

OM Practice

Balancing Risk and Efficiency at a Major Commercial Bank

Sunder Kekre, Nicola Secomandi, Erkut Sönmez
Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
{sk0a@andrew.cmu.edu, ns7@andrew.cmu.edu, esonmez@andrew.cmu.edu}

Kenneth West
Treasury Department, PPG Industries, Inc., Pittsburgh, Pennsylvania 15272, kwest@ppg.com

Check processing institutions are being forced to downsize their workforce to cut cost and improve the efficiency of their operations as a result of continued growth of electronic payments, a consequence of the increasing popularity of debit/credit cards and use of online banking. For these institutions, these events are making more urgent the decision of how to staff a check-clearing house to trade off efficiency and the expected costs associated with the risks of delayed checks, which include fraud and float costs. In this paper, we discuss how a team of executives at a major commercial bank (CB) and Carnegie Mellon University students and faculty engaged in conducting a model-based study of the CB check-clearing operations. This project culminated in the development of a simulation optimization model to systematically analyze the nature of the highlighted risk efficiency trade-off at CB. The firm used the model recommendations to obtain operations downsizing guidelines for its senior managers during the implementation of a strategic workforce reduction program at their check-clearing house. The managerial insights from the team analysis, and the specific model-based recommendations, enabled CB executives to balance risk and efficiency while planning the reduction of their check-processing workforce.

Key words: check processing; workforce sizing; worker cross-training; cost of risk; simulation optimization; math-programming-based lower bound; empirical research
History: Received: December 23, 2005; accepted: October 29, 2007. Published online in *Articles in Advance* April 25, 2008.

1. Introduction

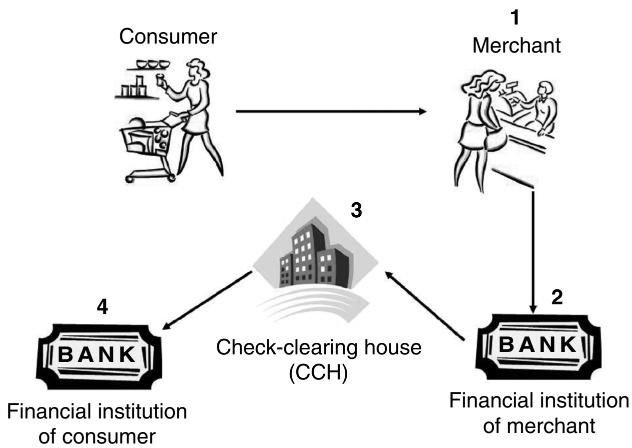
The site of our field study was a major commercial bank (CB) headquartered in the Northeast United States. CB provides various financial products and services for both corporations and individuals. For corporations it offers investment management, trust and custody, foreign exchange, securities lending, performance analytics, fund administration, stock transfer, proxy solicitation, treasury management, and banking services. At the individual offering, its portfolio includes a range of mutual funds and wealth management services.

A team of Carnegie Mellon University (CMU) students and faculty, including the authors, engaged with the cash and disbursements (C&D) division of CB. This unit serves corporations and is experiencing a sharp and steady decline in the overall volume of checks being processed at its facilities. This trend is representative of the industry and is projected to continue. CB senior management charged our team to support the bank operations and strategy group in modeling the operations of the C&D unit and to provide broad recommendations regarding restructuring of its workforce in light of the downward workload trends. While reducing the headcount seemed obvious because of declining check volumes, management was concerned with the degree of trimming, because excessive reductions could delay check-processing operations. The potential escalation of delayed checks and the consequent financial damages, such as the costs of fraud and float risks (explained below), could be substantial. Before making any decisions on

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

Figure 1 Check-Clearing Operations



the extent of reduction in workforce, management also wanted to examine new policies such as cross-training. In this paper, we report on the development of a system optimization model that quantifies the nature of this risk efficiency trade-off and provides recommendations for workforce sizing and the assignment of staff to workstations and shifts. CB senior management used the model guidelines to balance risk and efficiency while planning the reduction of its check-processing workforce.

Figure 1 shows the main steps involving check clearing. The customer pays the merchant for goods or services purchased by using a check (Step 1). The merchant deposits the check at its financial institution (Step 2). For a deposited check payable at other financial institutions, the merchant's financial institution sends the check to a check-clearing house (CCH) that clears and settles the check (Step 3). Finally, the CCH sends the check, or an electronic presentment file thereof, to the customer's financial institution (Step 4).

Checks generally arrive in bursts at the CCH, causing sudden surges in activity. If the CCH does not clear them before specific deadlines, it becomes liable for two charges associated with two sources of risk: fraud and float. The CCH fully incurs the cost of fraudulent checks that miss their deadlines (fraud cost) and is charged a penalty for tardy checks (float cost). The two costs must be balanced against the cost of maintaining sufficient staffing levels to avoid delayed checks. Scheduling too few workers increases the expected costs of the fraud and float risks, while assigning extra capacity increases the operating cost per check processed. Balancing this risk efficiency trade-off is complex because of the interaction of several variables such as arrival patterns, mix fluctuations in check types, changing levels of fraud risk associated with different streams of checks, and constraints on labor deployment. Furthermore, recent legislations, such as Check 21 (www.federalreserve.gov/paymentsystems/truncation/), have made check images legal documents. Their operational implication is that physical transportation of checks is being replaced by electronic image transmission. Therefore, the CCH internal operations, rather than the physical transportation of checks, essentially determine the fraud and float costs incurred by the CCH on delayed checks.

1.1. Research Questions

In our study, we focused on answering three research questions.

(1) How will changes in workforce levels affect the systemwide expected cost experienced by the CCH of CB (CB-CCH, for short), which includes wages and the costs of float and fraud risks? We modeled the trade-off between these cost components at CB-CCH and estimated their behaviors as functions of head-count reduction.

(2) How should CB deploy its workforce, i.e., assign workers to workstations and shifts, to minimize total expected cost? We devised strategies for workforce deployment that adjust to changing volume and risk patterns experienced at CB-CCH during a day. CCH assigned workers to shifts by solely considering check volume patterns. However workload in check processing varies among and within shifts because there are periods of sudden overloads. Furthermore, check characteristics, like dollar value or fraud risk, vary from shift to shift. An alternative worker-to-shift assignment policy would jointly consider both check volumes and risk potential. Managing for both volume and risk potential could lower the total float and fraud expected costs. We assessed the incremental gains of this joint policy and quantified the estimated net savings in expected operating costs.

(3) What are the benefits of cross-training at CB-CCH? The wages of cross-trained workers are higher because they reflect the costs of worker retraining, learning multiple task sets, and loss of specialization. These workers can be deployed across multiple

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

workstations, resulting in higher labor utilization and faster clearing operations. This aspect is important because of the changing load patterns experienced during a day. As waves of checks propagate through the facility, a cross-trained workforce can "follow" the workload wave pattern. Consequently, the same service level can be achieved with fewer workers. However, the issue is whether the savings from headcount reduction are large enough to offset the higher wage rates associated with cross-training. We estimated the benefits of cross-training for CB-CCH.

1.2. Contributions to Operations Management Practice

This paper contributes to the operations management (OM) practice literature by providing a detailed description of recent new challenges faced by CB-CCH, and a thorough account of the novel application of simulation optimization to provide solutions and managerial insights to CB on how to deal with them. In other words, our contributions to the OM Practice literature include both introducing in the OM literature novel aspects at the *business problem* level and showing the practical effectiveness in dealing with these issues of a *methodology* that is not typically employed in this literature. We elaborate on this statement below.

At the business problem level, by focusing on specific challenges faced by CB-CCH, we study a general problem encountered by any CCH. While it is true that versions of this problem have been studied in the literature (reviewed below), to the best of our knowledge, none of the existing papers comprehensively considers all the relevant costs associated with check processing at a CCH. Thus, we analyze a significantly richer and realistic version of this business problem. Furthermore, most of the previous models assume homogeneous checks, but in practice, check streams vary during the day and across days in terms of volume, dollar values, and fraud risk potential. The novelty of our model formulation lies in capturing this "heterogeneity" in check streams, in terms of volume, dollar value, and risk potential, while also accounting for all the relevant costs associated with managing the CCH operations.

It is also important to note that while costs analogous to float also arise in manufacturing, e.g., from long processing delays leading to high inventories and low responsiveness (poor customer service), the fraud cost in our setting has a different interpretation because it is a *contingent* liability that is incurred only if checks (jobs) miss deadlines. Such an asymmetric cost characteristic has drawn little attention in manufacturing research.

At the methodological level, we formulate, solve, and analyze our model using simulation optimization. This methodology is well suited to deal with the complexities of our business problem and model formulation. Simulation allows us to realistically represent the *network* of check processing activities. Direct optimization, by advanced metaheuristic methods, of the resource levels assigned to the activities in this network allows us to integrate detailed modeling of the network with the optimization of its sizing. While we employ an off-the-shelf metaheuristic solver to carry out this optimization, we develop a new math-programming-based lower bound that allows us to assess the quality of the solution generated by this solver. We also compare this solution against that yielded by the lower bound model, a mixed integer program (MIP) that ignores the sequencing constraints that link the different activities in the processing network. Furthermore, we assess the impact of uncertainty on the solution to our model.

1.3. Relevance

Our model and analysis are *managerially* relevant because they explore novel ways of managing the CB-CCH operations and had significant practical impact. Applied to our model formulation, simulation optimization searches for the best workforce size and its deployment to workstations and shifts to respond to changing volume, value, and risk patterns while minimizing systemwide expected costs. This is a significantly different approach to managing a CCH operations than both the purely volume based approach that had been traditionally employed by CB and the approach of Krajewski et al. (1980) that ignores fraud risk costs. Our analysis reveals that ignoring fraud risk cost while deploying workers sharply increases the expected cost of risk assumed by CB-CCH.

Of particular interest to CB, our modeling approach indicates the critical workforce level below which

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

total expected cost starts escalating dramatically, and it is flexible enough to allow us to analyze the potential benefit of cross-training. Our solution and analysis critically informed the decisions of CB senior management regarding workforce sizing at their CCH. In particular, they alerted them to the steep escalation in liabilities at workforce levels that were being considered initially. As a result, they decided to implement a higher overall workforce level than originally planned.

Our work is also relevant at the *modeling* level, because it brings to light the importance of detailed modeling of the processing network activities in making workforce sizing decisions at CB-CCH. We find that employing the optimal solution obtained by the deterministic MIP used to compute our lower bound, which ignores the sequencing constraints that link different activities in the processing network, to size the workforce yields an expected total cost that is 61% higher than that of our model solution. We also discover that modeling uncertainty has a marginal effect on the simulation optimization solution, a result driven by the small coefficient of variation that describes the flow of checks arriving to CB-CCH. Thus, the benefit of our modeling approach can be almost entirely ascribed to the detailed modeling of the sequencing constraints that characterize the CB-CCH operations rather than to the modeling of demand uncertainty. Moreover, our MIP-based lower bound reveals that the simulation optimization solution is nearly optimal, being 6.1% away from the computed lower bound.

Our analysis was specific to CB-CCH, but our managerial and modeling insights have broader significance for the strategic management of the CCH at other financial institutions, e.g., the Federal Reserve Bank, as they face declining volume trends and consolidate their operations.

1.4. Related Literature

Gomes and Meile (2002) classify the check-processing operations literature into two main areas: (1) routing and transportation and (2) analysis of check-clearing operations. In the first area, Cornuèjols et al. (1977) model the location of bank accounts to optimize float. Hill and Whybark (1982) examine route optimization and scheduling for transportation of checks to encoding centers. In the second area, Hess (1975) discusses the design and implementation of a check-clearing system for the Philadelphia Federal Reserve District. Body and Maber (1977) develop a two-stage forecasting method to estimate daily check volume. Murphy and Stohr (1977) discuss a dynamic programming model for check sorting. Maber (1979) presents a chance constrained model to optimally schedule encoding operators while considering daily workload uncertainty. Although Maber's model captures uncertainty, it considers only encoding and assumes a single deadline for all the checks processed in a day. However, check-processing deadlines depend on the check type and arrival time. Our model captures both of these dependencies.

Krajewski et al. (1980) develop a linear program for encoder shift scheduling. The model minimizes weekly regular and overtime wages and float cost to decide the number of full-time and part-time encoder clerks. This model ignores fraud risk cost, an important cost component in practice, which is included in our model.

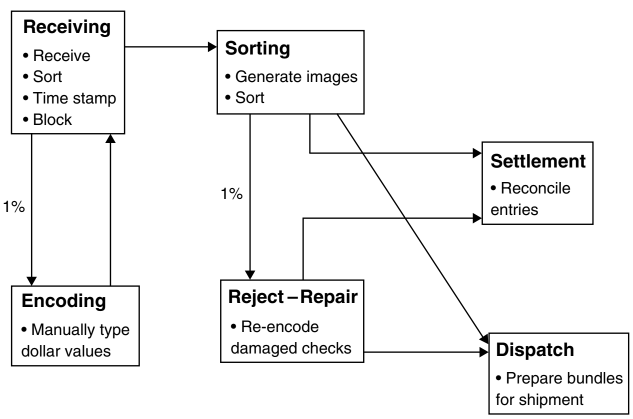
Davis et al. (1982) discuss a systemwide check-processing simulation model to analyze both transportation and processing of check-based payments. Davis et al. (1986) develop a simulation model to analyze the check processing operations of BancOhio National Bank. Markland and Nauss (1983) develop an integer programming model to minimize the cost of processing transit checks at Maryland National Bank (see also Nauss 1985). Gomes and Meile (2002) discuss a simulation model of check-processing operations and assess the benefits of alternative work flow scheduling rules motivated by new image processing technology. However, these authors do not optimize the resource levels of different activities as we do in this paper.

1.5. Organization

The remainder of this paper is organized as follows. In §2 we describe our analysis of the CB-CCH operations based on data provided to us by CB-CCH managers. In §3 we present our model and solution approach. In §4 we discuss the managerial insights that follow from the analysis of our model applied to the CB-CCH data. In §5 we elaborate on those modeling insights that are specific to our application and

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

Figure 2 **Check-Processing Work Flow**



implementation. We briefly conclude in §6, where we also discuss limitations of our work.

2. CB-CCH Operations and Data Analysis

In this section we discuss the check-processing work flow and our data analysis. While our discussion is specific to CB-CCH, it is representative of industry practices. Figure 2 illustrates the major workstations (displayed as boxes); we discuss the sorting step in detail below. Received and encoded checks are sent to the sorter room, where a number of reader-sorter machines generate images of and sort the individual checks. All checks are passed through these machines for what is called a *prime pass*. Subsequently, checks are sorted according to the following categories.

2.1. On-us

These are checks deposited at the same institution on which they are drawn, i.e., they are drawn on a CB account for disbursement. They arrive between 10:30 A.M. and 11:30 A.M. and have to be processed by 4:30 P.M.

2.2. Transit

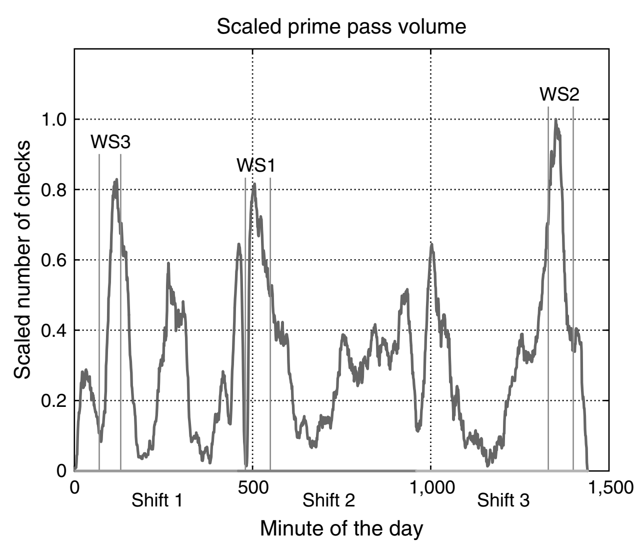
These are checks whose payer and payee have accounts at different financial institutions and are cleared and settled through clearing houses, such as one at CB or at a Federal Reserve Bank. These checks are further distinguished between "kill" and "rehandle." The former checks are transferred directly to the dispatch area because they are dispatched on one of the next courier pickups. The latter checks must be passed through the sorter multiple times to be grouped with other checks dispatched to the same location at a later courier time.

At this stage of the process, the checks are transferred to the dispatch area, while the data read from the magnetic information on the check is transferred to the settlement area.

We collected data on three aspects of the CB-CCH operations: check profiles, labor resources, and historical fraud patterns. This phase of the project involved querying databases populated by sorting equipment, interviews with check-processing managers, and time study observations. We obtained data on check characteristics such as volume, dollar amount, and check types from daily records. We "triangulated" the data to verify its accuracy. Management provided us with typical wage rates, productivity of workers at the respective workstations, and the historical profile of missed checks in terms of the fraud cost incurred. To preserve confidentiality, we properly mask these data in the ensuing discussion.

Figure 3 shows the scaled prime pass profile, i.e., the scaled number of checks per minute arriving at the CB-CCH sorter on a typical day. The same patterns repeated for every weekday, suggesting that each peak represented a burst of checks hitting the sorter. Hence, we partitioned a day into time buckets to capture these separate streams. Interviews with

Figure 3 **Scaled Number of Transit Checks Per Minute**



RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

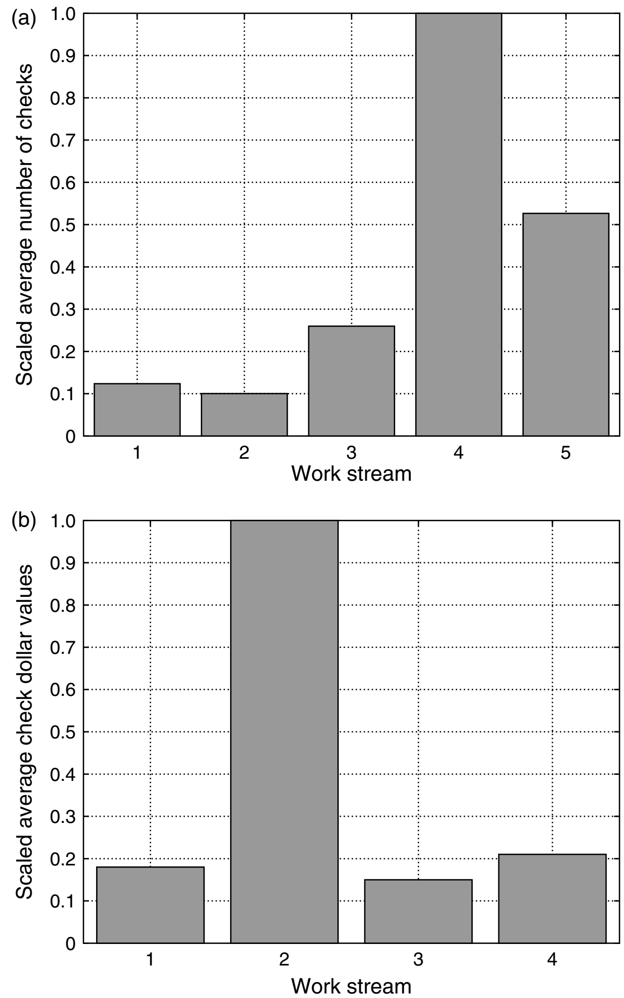
Table 1 Partitioning of Work Streams

Work stream	Time interval
1	8:10 A.M.–9:10 A.M.
2	10:40 P.M.–11:40 P.M.
3	1:10 A.M.–2:10 A.M.
4	Remaining time
5	10:30 A.M.–11:30 A.M.

the sorting supervisors revealed that different work streams, representing specific sources and types of checks, were causing the volume fluctuations. Further discussions with operators confirmed that five work streams, broken up by prime pass volumes, would capture the majority of the typical load pattern shown in Figure 3. Table 1 presents the time intervals defining these work streams. Figure 3 highlights Work Streams 1, 2, 3, and 4 (WS1, WS2, WS3, and WS4, respectively; WS4 is not explicitly indicated because of its "remaining time" definition). For instance, WS3 mostly comprises checks from lockboxes, and WS2 originates from high dollar value commercial checks. WS5 is solely associated with on-us checks and is not represented in Figure 3, which instead focuses on transit checks. (These patterns apply only to weekdays and a similar grouping in terms of work streams (not reported here for brevity) was performed for weekends, with Saturdays and Sundays treated as separate days.) CB-CCH uses three shifts per day: day (8 A.M.–4 P.M.), twilight (4 P.M.–12 A.M.), and night (12 A.M.–8 A.M.). Figure 3 also illustrates the work-stream-to-shift mapping. Because shifts include multiple work streams, they differ accordingly in terms of check volume, dollar value, and fraud probability. We used the Arena input analyzer to compute the probability distribution of the interarrival times for each work stream. We chose an exponential distribution to model these times for each work stream after conducting Chi-square and Kolmogorov-Smirnov tests. We also computed the average number of checks and check dollar values for each work stream (see Figure 4). For simplicity, we assumed that these random variables were normally distributed (in both cases the coefficients of variation were small enough to make negative realizations extremely unlikely).

CB management provided us with summary measures related to productivity of operators at various workstations and their wage rates. Table 2 displays

Figure 4 Scaled Average Number of Checks (a) and Check Dollar Values (b) by Work Stream



disguised versions of these figures. Finally, there are two types of relevant deadlines at CB-CCH. The first imposes a time limit for dispatch; the second is the settlement deadline. The dispatch deadline depends on the type of check. On-us checks have a dispatch deadline of 4:30 P.M., transit-kill checks have a

Table 2 Scaled Processing Times and Salary Per Employee

Process	Checks/hour	Annual salary
Blocking	0.38	0.9
Sorting	1	1
Encoding	0.025	0.9
Reject-repair	0.0086	0.9
Settlement	0.34	1
Dispatch	0.35	0.9

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

dispatch deadline of two hours of entering the system, and transit-rehandle checks have a longer time window for dispatch, stretching to 12 hours after entry. CB mandates an eight-hour window for settlement.

3. Model

In this section we present our model formulation and approach to solving it numerically. This model is a stochastic integer program. We do provide a math programming formulation of this model, but the random variables that depend on the number of resources assigned to the activities in the processing network are cumbersome to express analytically. Thus, this formulation does not provide explicit expressions for these quantities, and we use simulation to model them.

We employ the following notation to formulate our model:

- $i = 1, \ldots, 6$: worker type ($1 =$ Blocking, $2 =$ Encoding, $3 =$ Sorting, $4 =$ Reject and Repair, $5 =$ Settlement, $6 =$ Dispatch);
- $j = 1, 2, 3$: shift ($1 =$ Night, $2 =$ Day, $3 =$ Twilight);
- $k = 1, \ldots, 5$: work stream;
- T: planning horizon (simulation length in days);
- $t = 1, \ldots, T$: day;
- c_i: salary of type i worker for T days;
- x_{ij}: number of type i workers assigned to shift j; $x := (x_{ij},\ i = 1, \ldots, 6,\ j = 1, 2, 3)$;
- $\mathscr{X}$: set of worker-to-shift assignment (nonnegative integer valued) vectors x that clear all the checks by the end of each week;
- N_{tk}: random variable number of checks that arrive in stream k of day t;
- $\mathcal{N}_{tk}$: the support of N_{tk} with $n_{tk} \in \mathcal{N}_{tk}$; also define $\mathcal{N}_k := \bigcup_{t=1}^{T} \mathcal{N}_{tk}$ with $n_k \in \mathcal{N}_k$ for $k = 1, \ldots, 5$, and $\mathcal{N} := \bigcup_{k=1}^{5} \mathcal{N}_k$ with $n \in \mathcal{N}$;
- V_{tk}: random variable dollar value of any check in stream k of day t; we assume that V_{tk} and N_{tk} are stochastically independent;
- r_k: probability that a check in stream k, of any day, is fraudulent;
- p: float penalty per dollar.

We also use the two following random variables, which we model through simulation:

$$\mathrm{SL}_{tkl}(x) := \begin{cases} 1 & \text{If check } l \text{ in stream } k \text{ of day } t \text{ is settlement late given workforce level } x \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathrm{DL}_{tkl}(x) := \begin{cases} 1 & \text{If check } l \text{ in stream } k \text{ of day } t \text{ is dispatch late given workforce level } x \\ 0 & \text{otherwise.} \end{cases}$$

The staffing model is

$$\min_{x \in \mathscr{X}} \sum_{i=1}^{6} c_i \sum_{j=1}^{3} x_{ij} + \sum_{t=1}^{T} \sum_{k=1}^{4} E\left[\sum_{l=1}^{N_{tk}} V_{tk} \cdot \mathrm{DL}_{tkl}(x) \cdot (r_k + p)\right],$$

$$\text{s.t.} \sum_{t=1}^{T} \sum_{k=1}^{5} \sum_{l=1}^{n_{tk}} \mathrm{SL}_{tkl}(x) \le 0.01 \sum_{t=1}^{T} \sum_{k=1}^{5} E[N_{tk}], \quad \forall\, n \in \mathcal{N}$$

$$\sum_{t=1}^{T} \sum_{l=1}^{n_{t5}} \mathrm{DL}_{t5l}(x) \le 0.01 \sum_{t=1}^{T} E[N_{t5}], \quad \forall\, n_5 \in \mathcal{N}_5.$$

The objective function captures the systemwide expected cost during the planning horizon: worker wages and the expected cost of delayed checks, with the latter component including fraud and float expected costs (WS5 is not included because late on-us checks do not generate such costs). It should be noted that expected float costs do not depend on lateness of a check because all the checks that miss their deadline get priority processing and are batched and cleared together by the next deadline. Therefore, the float cost, p, is the dollar penalty associated with one-time bucket delay when a check misses its deadline. The two constraint sets reflect CB-CCH operational requirements to ensure adequate service levels with respect to settlement and dispatch deadlines (such deadline violations are difficult to quantify in terms of monetary loss). The first states that the workforce should be set at a level such that the total number of settlement tardy checks in the planning horizon should be no more than 1% of the total expected inflow. The second imposes a service-level limit on the on-us checks that ensures a cumulative cap on the maximum number of checks that could be held up in the dispatch area.

Our optimization model is a stochastic integer program that is difficult to solve exactly. Moreover, as stated above, we model random variables $\mathrm{SL}_{tkl}(x)$ and $\mathrm{DL}_{tkl}(x)$ via simulation. Therefore, we employ a simulation optimization approach to solve our model. Specifically, we used the Arena simulation software supplemented for optimization by OptQuest, which works in conjunction with Arena iteratively. When a

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

solution vector x needs to be evaluated, OptQuest calls Arena, which simulates the system accordingly. OptQuest then evaluates this vector x and integrates it with any previously obtained solutions to determine a new solution based on tabu search and scatter search principles (Glover and Laguna 1998, Laguna and Martí 2003). OptQuest continues to evaluate different solutions until a termination criterion is satisfied.

In the simulation step, we simulate the operational details of the system for a one year period (the length T of the planning horizon) using Arena. Checks arrive at CB-CCH in batches, and move between workstations in batches whose sizes change due to execution of the encoding, sorting, and reject-repair operations.

We validated our simulation model by using a confidence interval approach based on historical data. We used a three-month-period data set comprising arrival times, volumes, and dollar values of the processed checks. Management provided us with actual transit dispatch times, so we were able to compare the float cost for the actual system with that predicted by our model of the existing operations during the same time period. We used float cost only as a validation metric because this is a metric that CB-CCH meticulously tracks for performance measurement.

The system clears all the checks at the end of each week. For this reason, we divided the data into subperiods with one-week duration. We computed float costs for the three-month period (12 weeks) using our simulation model. Then, we compared the model performance with that of the actual system by constructing a confidence interval for the difference $\mu_X - \mu_Y$, where μ_X is the mean weekly float cost for the actual system and μ_Y is the same quantity for the simulated system.

To construct such a confidence interval, denote X_M and Y_M the float costs for the actual and simulated systems, respectively, in week M. Define $D_M := X_M - Y_M$, $M = 1, \ldots, 12$. The $(1-\alpha)100\%$ paired-t confidence interval is $\overline{D} \pm t_{N-1,\,1-\alpha/2}\sqrt{s_D^2/N}$, where $\overline{D}$ is the average of the D_M values, $t_{N-1,\,1-\alpha/2}$ is the upper $1-\alpha/2$ critical point for the t distribution with $N-1$ degrees of freedom (here $N=12$), and s_D^2 is the sample standard deviation of the D_N values (Law and Kelton 1991, p. 587). The computed 90% confidence interval for $\mu_X - \mu_Y$ is $[-24.409, 16.74]$. Because this interval contains 0, we cannot reject the hypothesis that the quantities μ_X and μ_Y are the same, and we concluded that our simulation model provided a fairly close representation of the CB-CCH operations.

In the optimization step, OptQuest offers different stopping rule options. We experimented with a fixed limit on the number of simulations, as well as run time, and we found that stopping after 265 nonimproving iterations, one of the default options in OptQuest, seemed to yield the best performance in terms of solution quality. Thus, we selected this as our stopping criterion.

4. Managerial Insights

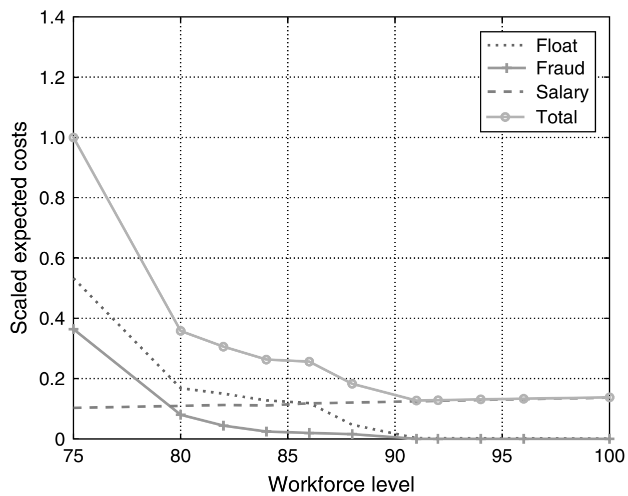
Our model-based analysis was critical to CB senior management in strategically planning and tactically implementing a workforce reduction program in the wake of declining business volume at their check-processing centers. Knowledge of how to manage the drivers of total expected costs, labor, and fraud and float risks, was crucial in targeting deployment of operators. Management placed a high degree of confidence in the recommendations and insights generated by our model and analysis and used them to guide executive decisions that were taking place concurrently with our field study. Thus, the output of our model-based study was an important input to the executive decision-making process. While our recommendations were implemented and had significant practical impact, in this section we focus on those managerial insights that follow from our analysis that have the most potential of being relevant beyond the specific application at CB-CCH.

4.1. Sensitivity of Expected Cost to Workforce Level

The first and foremost concern of CB management was to assess how expected costs would change before deployment of headcount reduction plans. This necessitated projecting the total expected cost function at varying levels of workforce, i.e., $\sum_i \sum_j x_{ij}$. Figure 5 displays this function, as well as its wage, float, and fraud components. (The cost items plotted in Figure 5 are the minimum expected costs for workforce levels $W = 75, 80, 82, 84, 86, 88, 91, 92, 94, 96, 100$, as obtained from solving our model with the following constraint added: $\sum_i \sum_j x_{ij} = W$.) The shape of the function confirmed the apprehension of senior

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

Figure 5 Scaled Expected Costs at Varying Workforce Levels



management with respect to the extent of workforce trimming.

Moreover, Figure 5 reveals an important insight. As we move from the right to the left and cut headcount, the total expected cost at first decreases, although not substantially. The minimum occurs at the workforce level. However, the adverse impact of reducing employee level rises dramatically below worker level 91 due to drastic increases in expected float and fraud costs. Below 82, the increase in such expected costs is very severe. The decrease in salary in contrast is very small in magnitude compared to the expected costs of the fraud and float risks assumed by CB. This suggests a threshold level of 91, below which fraud and float expected costs overshadow labor costs. Moreover, at this workforce level, the expected costs of fraud and float risks are small, which suggests that CB-CCH clears most of the checks before their deadlines without incurring any financial liabilities.

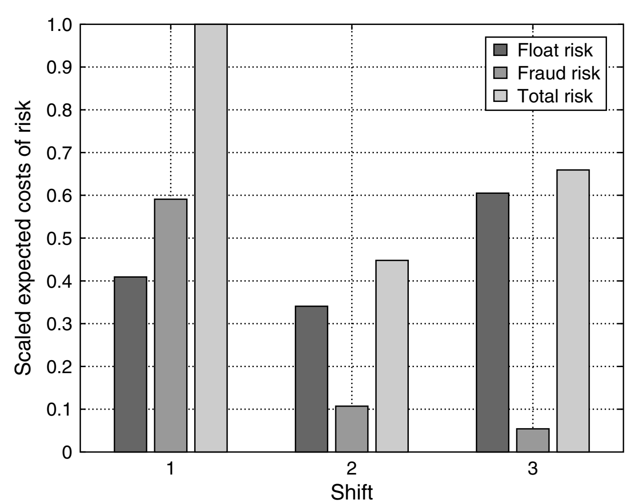
The managerial insight for CB senior management is as follows. The planned workforce reduction is based on estimated onward trends in the volume of processed checks. Firing too many workers and ending up with check overloads can be financially detrimental overall. Thus, being conservative and having more than adequate capacity is far better than lacking capacity. In other words, it is wiser to pay labor costs rather than taking risk. On the other hand, management does need to cut the current excess labor and plan for a headcount of about 90. It should be noted that while our analysis (Figure 5) does not show any huge monetary savings from such reduction, there are several hidden costs of over-employment that are not explicitly considered in our model.

4.2. Relevance of Considering Fraud Cost

CB senior managers were cognizant of the fact that different assignment policies could give rise to varying cost figures, even if workforce size were kept constant. For example, they correctly, but qualitatively, argued that a good assignment policy should consider the differences in check characteristics like dollar value and fraud probability. We used our model to quantify the benefit of considering the work stream risk potentials while assigning the available workforce to the various steps of check clearing.

The float cost is higher when CB-CCH misses a deadline for a high-dollar-value check. The same is true for fraud cost, when CB-CCH delays a check with high fraud probability. As discussed in §2, this probability depends on the work stream. We defined total expected cost of risk assumed by CB as the sum of the expected costs associated with the float and fraud risks. We computed the total expected cost of risk for all the individual shifts and show it in Figure 6. Note that the expected float cost for Shift 3 is higher than for the other shifts because it includes WS2. This work stream comprises high-dollar-value commercial checks. The probability of any of these checks being fraudulent is low. In contrast, the expected fraud cost

Figure 6 Scaled Expected Costs of Risk

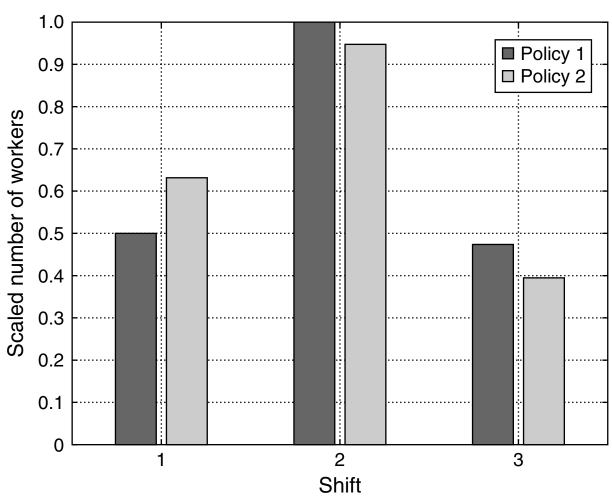


Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

for Shift 1 is significantly higher because it includes WS3, which is formed by low-dollar-value checks coming from lockboxes with higher fraud probability. This analysis provided valuable insights to CB senior managers and confirmed that assignment should be based on both volume and risk costs in each shift.

However, given that there were five work streams, each with different workload and risk profile, an optimization routine was needed to balance the labor requirements across shifts with different volume and risk profiles. This motivated us to analyze different assignment policies: one that considered fraud cost, and one that ignored it (as previously done by Krajewski et al. 1980). The results of the analysis showed that fraud cost could not be ignored in workforce assignment. This analysis was performed using a two-step computation. We found workforce assignments for two different policies. In the first policy, we fixed the workforce size at 75 workers and considered only float cost. In other words, we ignored the cost of fraud and found the workforce assignment that minimized expected float cost using our simulation optimization model. In the second policy, by keeping the workforce size at 75, we included fraud cost and recomputed the assignment that minimized the sum of expected float and fraud costs. As can be seen from Figure 7, when we include fraud cost, the number of workers assigned to Shift 1 by the optimization

Figure 7 **Difference Between Assignment Policies that Ignore and Consider Fraud Cost (Policies 1 and 2, Respectively)**



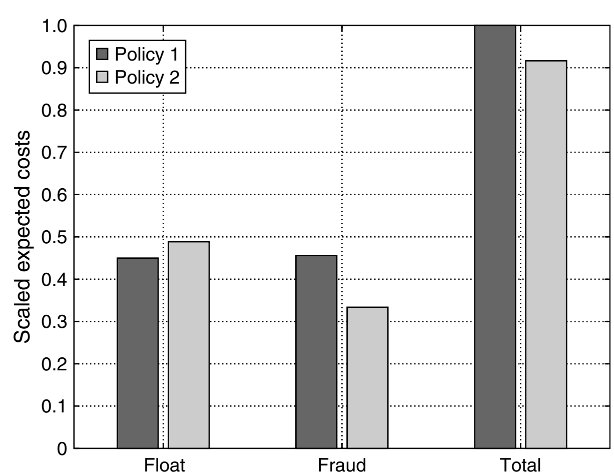
Note. Policy 2 moves workers to the shift with more fraud cost potential (Shift 1).

model increases. Also note that the number of workers assigned to Shift 2 is greater than that in other shifts because this shift includes on-us checks that have to be cleared faster as an operational requirement (the second service level constraint set of our optimization model). Moreover, the check volume of Shift 2 is also greater than that of the other shifts.

Figure 8 illustrates the total expected cost and its breakdown by expected float and fraud costs in the previous analysis. The financial impact of considering fraud cost in an assignment policy is very significant. The total expected cost reduces substantially (by approximately 8.30%). Most of the gains are accrued by reducing its fraud component by 26.82%. Note that the optimal assignment achieves this reduction in expected fraud cost by allocating workers to work streams with higher fraud cost potential. In contrast, expected float cost actually increases by 8.58%, resulting in total expected savings of 8.30%. This aspect of workforce rebalancing in response to consideration of fraud cost was very insightful to CB senior managers. Note that in both assignment policies the labor costs were the same as we maintained a workforce level equal to 75.

The conclusion from Figures 7 and 8 is that a good workforce assignment policy should consider both check volume and risk potential. The counterintuitive insight is that an increase in expected float cost can lead to a lower expected fraud cost. This finding went against conventional wisdom, perhaps because

Figure 8 **Expected Costs of Ignoring and Considering Fraud Cost (Policies 1 and 2, Respectively) in the Assignment Policy**



Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

most banking operations and the existing literature have focused primarily on managing the float cost component.

4.3. Benefit of Cross-Training

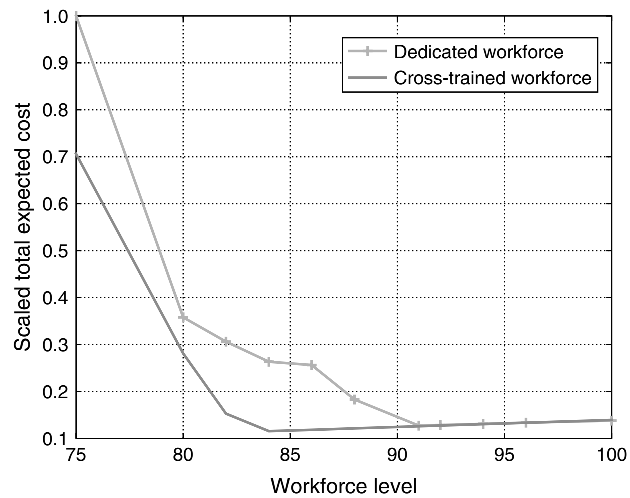
Cross-trained multifunctional operators can bring further cost reductions. The current flow of work occurs cyclically and must move sequentially through a number of steps. This creates a different, but predictable, need for resources as a function of time of day for each specific activity. By cross-training, management has the flexibility to share workers across workstations and assign them based on changing workloads patterns imposed by work streams during the day. For this reason, a team of cross-trained operators can achieve higher service levels compared to a same-size team of dedicated operators. However, cross-training is expensive and possible only for some groups of machines (see Table 3). The price to be paid for this flexibility is the higher salary of a cross-trained worker than that of a dedicated worker.

Figure 9 shows the sensitivity of total expected cost to cross-training in the workforce. As can be seen from this figure, cross-training is attractive in a selected workforce level region. Cross-training is overall beneficial in the 80- to 90-worker zone, approximately, with the exception of workforce level 75, which is however far from optimal. In other words, with this exception, cross-training is most valuable for workforce levels that fall in between overload and excess capacity situations. This can be explained as follows. With 90 or more workers, the system processes nearly all the checks on time and within the deadlines, the expected cost of risk is very small, and processing capacity is adequate. For this reason, cross-training of workers does not bring substantial reductions in total expected cost at these workforce levels. When worker level is smaller than 80, utilization of workers is very high

Table 3 **Cross-Trained Worker-Types-to-Workstations Mapping**

Cross-trained worker type	Blocking	Encoding	Dispatch	Settlement
1	✓	✓		
2	✓			✓
3	✓		✓	
4	✓	✓		✓

Figure 9 **Benefit of Cross-Training**



and the system is already congested. In this overload situation, cross-training cannot prevent the rapid escalation in total expected cost. Shifting workers cannot substantially relieve the high congestion-related expected costs. If the worker level is between 80 and 90, there is a significant potential for increasing utilization by leveraging cross-training to move workers between machines depending on changing workload patterns, thereby reducing the expected cost of risk. As a result, cross-training leads to significant expected cost reductions if it is introduced in the right capacity band.

We also examined the drivers of the benefit from cross-training at the optimal workforce level. This benefit can be split into two components: the first is the ability to adapt to systematic changes in workload volume over time, i.e., the deterministic variation of average load over time; the second arises from the ability to adapt to statistical variations around this deterministically changing average. We refer to the first component as *systematic changes in workload patterns* and to the second component as *risk pooling*. We computed the improvement from cross-training without stochastic variability in check volume and found it to be 9.83%. We then computed the percentage improvement from cross-training in the wake of this stochastic variability in the volume of checks. We found that the beneficial effect of cross-training when both the systematic and risk pooling components are present to be 9.28%. Note that on adding volume

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

uncertainty, the expected cost without cross-training increases by 1.06%. Taken together, these results seem to suggest that the lion's share of the benefit of cross-training comes from the first component, i.e., moving workers to accommodate systematic changes in workload patterns. The reason for this outcome is that when uncertainty in volume is added on top of systematic variation, the adverse impact of uncertainty is small (1.06%), so the cross-training benefit from risk pooling is intrinsically limited.

The managerial implications of these findings are that the benefit of cross-training is substantial and arises primarily from leveraging the movements of workers to respond to systematic workload variations. However, it should be noted that our results reflect an environment characterized by low uncertainty. We cannot, therefore, estimate what the risk-pooling-related benefit of cross-training would be in higher uncertainty environments, but it seems fair to state that it would be expected to be higher.

5. Modeling Insights

Several papers discuss the modeling lessons that typically can be learned from applications (e.g., Pidd 1999, Secomandi et al. 2002). These include having a clear financial objective, obtaining early visible commitment from senior management, and adopting a staged modeling approach. We also found these aspects to be relevant in our study. In this section, we discuss the main modeling insights that were specific to our application.

5.1. Lower Bound Tightness

To assess the quality of the solution of our model generated by OptQuest, we computed a lower bound on its expected cost as follows. We formulated and solved a relaxed and deterministic MIP by relaxing the sequencing constraints on the processing operations and by removing the uncertainty in check volume and mix in our simulation optimization model. The online appendix presents this model formulation and proves that its optimal solution yields a lower bound on the true optimal expected cost. Notice that we maintained the integrality conditions on the workforce-level decision variables. This MIP addresses only the "loading problem" and ignores the specific sequencing of jobs across work centers. In contrast, our simulation optimization model recognizes these dependencies and ensures that, besides loading, sequencing constraints are also satisfied. Therefore, it provides a better representation of the actual system. We solved the MIP to optimality using CPLEX and obtained a solution whose cost was 6.1% below the expected cost of the solution to our model obtained by OptQuest. This finding established that the OptQuest solution to our model was nearly optimal. Given that OptQuest is a metaheuristic solver and the MIP ignored the sequencing constraints, we were positively surprised by the high quality of the OptQuest solution and the tightness of the lower bound.

5.2. Relevance of Modeling the Sequencing Constraints

To establish the relevance of our detailed modeling of the CB-CCH operations, we evaluated the optimal MIP workforce plan by feeding it into the simulation step of our simulation optimization model (see §3). Notice that this allowed us to impose on this MIP solution the sequencing constraints that are ignored by the MIP. The simulation showed that this workforce plan had an expected cost that was 61% higher than that of our OptQuest solution. The implication of this result is very significant. Ignoring the sequencing constraints in the MIP results in a workforce plan that dramatically escalates delayed checks and, hence, expected cost of risk. In particular, while our model yields similar *total* workforce level as the MIP, in the MIP solution the assignment of workers to stations and shifts is far from optimal, which results in such a dramatic degradation of performance. While we expected that the true expected cost of the MIP solution would be different from that of the simulation optimization solution, we were surprised by the particularly poor performance of the MIP solution relative to that of the latter solution.

5.3. Relevance of Modeling Uncertainty

We also analyzed if a deterministic simulation approach, based on replacing the random variables in our simulation optimization model with their expected values, could be used without introducing excessive approximations in the solution generated while also yielding good cost estimates. In other

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

words, differently from the MIP described above, this deterministic approach retains the detailed modeling of the system operations but suppresses any source of uncertainty. To our surprise, we discovered that the solution of the deterministic model was within 2.5% of that of the stochastic model in terms of objective function value *and* recommended the same workforce level and assignments. Relative to this deterministic model, adding mix uncertainty (in terms of check dollar values) increases expected cost by 1.5%. Adding check volume uncertainty on top of mix uncertainty further escalates cost by 1%. The "high volume" characteristic of the system is the main driver for this solution similarity. The coefficient of variation of the interarrival times is very small, making a deterministic assumption very reasonable (Berman et al. 1997). Thus, modeling uncertainty, in both volume and mix, changes the total expected cost by only 2.5%. This finding had useful practical implications for our computational study because it improved by an order of magnitude the computational time of obtaining a solution. Thus, it helped us swiftly conduct a detailed analysis of other operating scenarios. (For accuracy, we retained all the relevant sources of uncertainty in the analysis conducted in §4.)

6. Conclusions

Our model solution and analysis were critical to inform the decisions of CB management regarding their CCH workforce level sizing. However, not all the options considered were adopted by CB management. For example, the estimated monetary benefits from the cross-training analysis indicated that the expected gains were not high enough to warrant the reorganization and training costs needed to implement it.

We conclude by discussing some limitations of our work. (1) Our model assumes that the productivity rate of workers is constant. Such factors as machine breakdowns, uncertainty of worker productivity rates, and fatigue can change the capacity during a day, especially when business is shrinking, but our model does not consider this aspect of the business problem. (2) Our model assumes that a cross-trained worker keeps working at a workstation until the queue of that station becomes empty. When this happens, the cross-trained worker moves to the station with the longest queue. To be more realistic, we would need to impose rules for moving cross-trained workers rather than bind them at a station until idle. (3) Our model employs a yearly planning horizon. Extending this period beyond one year would better capture currently declining check volume trends. (4) Our model assumes that workers are assigned to shifts optimally. If optimal deployment is not applicable for a given CCH, the workforce level suggested by our model should be adjusted upward.

These are clearly simplifying modeling assumptions, which we made because we had only limited time availability for developing our model and making our recommendations. In spite of these limitations, our model proactively guided CB management to make workforce reduction decisions, which confirmed its practical usefulness. Further research could relax some of our modeling assumptions.

Electronic Companion

An electronic companion to this paper is available on the *Manufacturing & Service Operations Management* website (http://msom.pubs.informs.org/ecompanion.html).

Acknowledgments

The authors thank the Senior Vice President of Operations at CB and her staff for their help and support during our field study. They also thank the previous and current editors-in-chief (Garrett van Ryzin and Gerard Cachon), the associate editor, and three referees, whose constructive criticism led to a significantly improved version of this paper.

References

Berman, O., R. C. Larson, E. Pinker. 1997. Scheduling workforce and workflow in a high volume factory. *Management Sci.* **43** 909–932.

Body, S., V. A. Maber. 1977. A two stage forecasting approach at the Chemical Bank of New York for check processing. *J. Bank Res.* **5** 101–107.

Cornuéjols, G., M. L. Fisher, G. L. Nemhauser. 1977. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Sci.* **23** 789–810.

Davis, S. G., N. Ceto, J. Rabb. 1982. A comprehensive check processing simulation model. *J. Bank Res.* **13** 185–194.

Davis, S. G., G. B. Kleindorfer, G. A. Kochenberger, E. T. Reutzel, E. W. Brown. 1986. Strategic planning for bank operations with multiple check-processing locations. *Interfaces* **16**(6) 1–12.

Glover, F., M. Laguna. 1998. *Tabu Search*. Kluwer Academic Publishers, Boston.

RIGHTSLINK

Gomes, P. J., L. C. Meile. 2002. Leveraging the potential of process technology through workflow scheduling. *Internat. J. Service Indust. Management* **13** 7–28.

Hess, S. W. 1975. Design and implementation of a new check clearing system for the Philadelphia Federal Reserve District. *Interfaces* **5**(2) 22–36.

Hill, A. V., D. C. Whybark. 1982. Chexpedite: A computer-based approach to the courier problem. *Decision Sci.* **13** 251–265.

Krajewski, L., L. Ritzman, P. McKenzie. 1980. Shift scheduling in bank operations: A case application. *Interfaces* **10**(2) 1–8.

Laguna, M., R. Martí. 2003. *Scatter Search: Methodology and Implementations in C*. Kluwer Academic Publishers, Boston.

Law, A. M., W. D. Kelton. 1991. *Simulation Modeling & Analysis*. McGraw-Hill, Inc., New York.

Maber, V. A. 1979. A case study of encoder shift scheduling under uncertainty. *Management Sci.* **25** 623–631.

Markland, R. E., R. M. Nauss. 1983. Improving transit check-clearing operations at Maryland National Bank. *Interfaces* **13**(1) 1–9.

Murphy, F. H., E. A. Stohr. 1977. A dynamic programming algorithm for check sorting. *Management Sci.* **24** 59–70.

Nauss, R. M. 1985. Optimization of bank transit check-clearing operations. *Management Sci.* **31** 1072–1083.

Pidd, M. 1999. Just modeling through: A rough guide to modeling. *Interfaces* **29**(2) 118–132.

Secomandi, N., K. Abbott, T. Atan, E. A. Boyd. 2002. From revenue management concepts to software systems. *Interfaces* **32**(2) 1–11.

Downloaded from informs.org by [155.246.103.35] on 25 March 2017, at 18:22 . For personal use only, all rights reserved.

RIGHTSLINK