
FINANCIAL SERVICES

Raising banks' productivity will take more than one-time cost cuts

Lean manufacturing approaches could bring big benefits

The case of "Chuck the check"




JAMEL AKIB

First National Toyota

Anthony R. Goland

John Hall

Devereaux A. Clifford



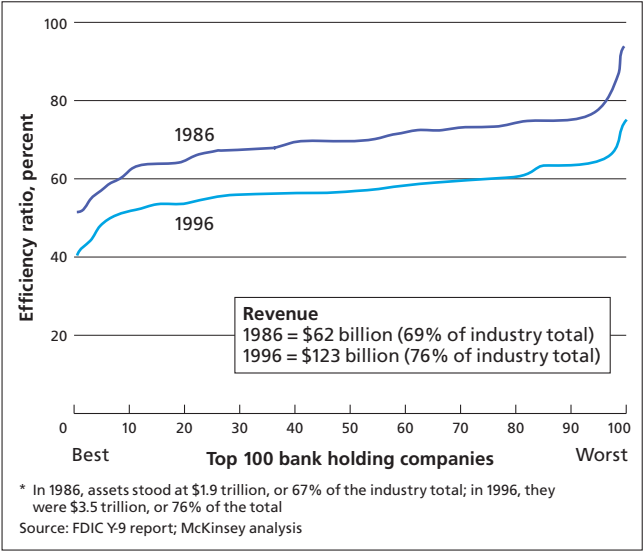
OVER THE PAST DECADE, US banks have improved their productivity by about 1 percent a year. They have done so mainly through short sharp shocks: one-off initiatives focused on cutting costs and reducing headcount (Exhibit 1). Though unpleasant, these exercises proved effective in boosting earnings, and hence stock price. Once they were completed, however, life usually returned to normal. Productivity often stagnated or even slid back until the next wave of cost cuts.

So while substantial costs have been taken out of the US banking system, scope remains to raise productivity

Tony Goland is a director in McKinsey's Pittsburgh office, John Hall is a consultant in the Toronto office, and Dev Clifford is a former consultant in the Cleveland office. Copyright © 1998 McKinsey & Company. All rights reserved.

Exhibit 1

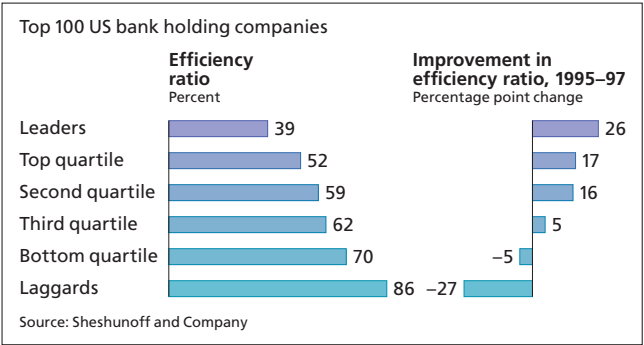
Efficiency ratios of top 100 US bank holding companies, 1986–96*



further; indeed, much of the merger and acquisition activity of the past two years has been based on this premise. Though variations in business mix justify some structural differences in operating costs, the size of the gap between the absolute efficiency and improvement rates of the best-performing banks and those of the rest of the pack gives some indication of the potential (Exhibit 2).

Exhibit 2

Differences in operational performance



How then should banks go about capturing the extra productivity? The answer is: by looking at how manufacturers do it. Rates of improvement recorded by world-class manufacturing companies far exceed those attained by banks (Exhibit 3). Much of the reason can be found in lean manufacturing, a system of operational management developed by the Japanese car maker Toyota.

Exhibit 3

Improvement rates of world-class manufacturers

	Years	Performance variable	Annual improvement rate, percent
GE Lighting	6	Cost per unit	8.0
Nucor	5	Cost per ton	10.6
PepsiCo	5	Order-to-delivery time	9.1
Toyota	6	Defects per million	6.4

At the heart of lean manufacturing lies the belief that simultaneous improvement in quality and efficiency is not a one-off goal, but a continuous pursuit that leads naturally to lower manufacturing cost. Lean manufacturing has reshaped industries from steel to semiconductors, bringing about dramatic shifts in costs, market share, and profitability.

Leading practitioners enjoy lasting advantages in terms of both processes (quality, inventory, and lead times, for instance) and output (cost, customer satisfaction, and market share).

Though the analogy between banks and manufacturing companies may seem tenuous, many areas of banking share important characteristics with manufacturing. Good manufacturers ensure that each of their plants operates at maximum productivity by regularly measuring performance and sharing ways of doing things better. Labor-intensive banking operations offer similar scope for measurement and improvement. The codification and transfer of best practices can raise productivity in credit evaluation and call centers, for instance, while the standardization of maintenance schedules and operator practices can lead to big improvements in machine-based activities such as check sorting.

World-class manufacturers are also masters at wringing every last unit of capacity from their operations before they commit further investment. In pursuit of the same goal, banks need to move toward viewing lending officers as a fixed pool of capacity that can be used for various tasks, including direct interaction with customers and selling. (Admittedly, optimizing this capacity may call for changes in organization, technology, and the way officers spend their time.) Another parallel between the two industries is the need for rapid delivery. For purchasers of manufactured goods, this is often a key buying factor, while for banks, compressing the time it takes, say, to approve a mortgage can confer competitive advantage.

In our experience, a typical consumer-oriented bank can realize one-off improvements in its efficiency ratio of 2 to 5 percentage points by applying lean manufacturing techniques in areas such as check processing, credit application and approval, and call centers. A disciplined focus on operational effectiveness may yield an additional year-on-year productivity increase; manufacturing companies proficient at continuous improvement regularly log annual increases of 5 percent. As pressure on the drivers of profitability intensifies, we believe banks should act quickly to embrace these techniques and move toward continuous productivity improvement.

Characteristics of lean manufacturing

Lean manufacturing was developed by Toyota to optimize production by eliminating waste. It entails building quality into manufacturing processes while recognizing the importance of cost reduction. It has four defining characteristics:

Waste awareness. Lean manufacturing aims to banish seven types of waste: overproduction, unnecessary motion, repairs, overprocessing, waiting (for parts or for a machine to finish its cycle), excess inventory, and inessential transportation. The responsibility for spotting waste and figuring out ways to prevent it lies with frontline employees. The solutions they propose are spread quickly around the organization.

Continuous quality assurance. Rather than inspecting work after it is completed and then performing any necessary rework, lean manufacturing techniques build in quality assurance at every step in the process. In some cases, frontline workers have the authority to stop a production line if they see it putting out substandard products. They are supported by standardized work instructions that prevent differences in execution and rigorous maintenance procedures that minimize breakdowns and other hiccups in machine performance.

Just in time. Just in time production aims to supply each process with exactly the right quantity of what it needs when it needs it. Although normally associated with external suppliers, it can also be extended to internal functions. Central to it is the idea of continuous flow: the elimination of stagnation in and between process steps so as to minimize in-process inventory, avoid waste in transportation and motion, and improve feedback on defective output. Flexibility (reducing set-up times to the minimum to promote effective capacity utilization and cross-training employees to enable them to switch between assignments) and minimum inventory (using tracking mechanisms to keep tabs on stock levels, avoiding bottlenecks to minimize work in progress, and timing deliveries to demand) are the other cornerstones of this approach.

Level production. Level production is the end state achieved by a lean manufacturing system. Peaks and troughs in volume are minimized to ensure optimum capacity utilization, and the process is designed to enable the fastest possible throughput. From an internal perspective, tasks are balanced between work stations and process steps so that utilization is maximized and overall cycle time kept to a minimum.

Three examples from check processing, loan processing, and call centers illustrate how lean manufacturing can be applied to financial services.

Keep the checks moving

Since it involves a physical process not unlike an assembly line, the handling of paper checks and credit card slips lends itself readily to lean manufacturing techniques. And their impact can be dramatic: the faster a bank moves checks through its system, the sooner it can collect its funds and the better its returns on invested capital.

Lean manufacturing has helped one North American financial institution to double the contribution from its check processing operation with very little investment. More important, it has produced a shift in mindset among managers and frontline operators that provides a foundation and impetus for ongoing improvement.

Before launching its lean manufacturing effort, the bank had pursued a range of cost-focused projects to improve performance, and carried out competitive benchmarking against other processing operations in the industry. The results indicated that its check processing operation stood near the top of the business in productivity, and satisfied some senior managers that it was operating at full capacity. Since additional capacity could be sold profitably to other banks, these managers contemplated investing in new capital equipment to increase production.

But other managers thought they needed to look at the quality of performance more carefully. They decided to follow one check as it made its way through the bank's systems, dubbing the exercise "the journey of Chuck the check." They documented the time spent in actual processing and in waiting, rework, and handling. They also conducted an all-night vigil in the processing room. The exercise recorded the time spent in processing and non-processing activities at each stage in the check's journey over the course of 24 hours, and the total volume of checks being processed at the same time.

To their surprise, the managers discovered that they were nowhere near as efficient as they had thought. They found that almost half of the bank's notional processing capacity was consumed by non-processing activities such as fixing jams and setting up machines. Even with this loss of capacity, there were only two periods during the day when the number of checks to be processed exceeded capacity; for the rest of the day, capacity went unused. Further investigation revealed wide variations in productivity between individual operators on a single shift. When the work practices of the least and most productive operators were compared, it became evident that although all were engaged in the same task, differences in the way they performed it were creating huge swings in productivity.

The faster a bank processes checks, the sooner it can collect its funds and the better its returns on invested capital

The bank's response was to adopt a lean manufacturing approach. Its first step was to match the flow of incoming checks to processing capacity. At the end of each business day, the check processing operation was swamped with more checks than it could hope to handle. This bottleneck created the false impression that capacity was constrained. The bank applied just in time principles to the processing of incoming checks and spread the check flow evenly through the day. Previously, although there were two check shipments each day, the branches had held most of the checks until the final shipment to simplify their end-of-day reconciliation process. When this practice was abolished, the bottleneck vanished.

A second bottleneck occurred at the beginning of the day. Standard practice dictated that all checks presented for morning processing were sorted three

times in all. This prevented the processing operation from handling the morning check volume in time to meet the account posting deadline. However, many of the checks did not need to be completed by the morning deadline, and once the sorting of these low-priority items was shifted to later in the day when volumes were lower, capacity increased by 122 percent.

Eventually, the bank's managers realized that they could have their cake and eat it. By uncovering and freeing up "phantom" capacity that had previously been taken up by waiting time, maintenance, and rework, they could increase actual capacity by more than 25 percent without investing in additional equipment. The bank was able both to sell its services to other banks at an attractive price and to expand capacity during the most time-sensitive period of the day, when its services could be priced at a premium. In all, these one-off improvements resulted in a more than doubling of the margin contributed by the operation.*

Seeking faster approval

The \$900 billion residential mortgage underwriting business in the United States is also fertile ground for lean manufacturing techniques. Because customers usually take up the first mortgage for which they are approved, the faster and more accurately a lender can process an application, the better its chance of closing the deal. And since it is so easy for customers to shop around for a mortgage, productivity advantages that can be translated into lower prices represent a powerful competitive weapon.

Many well-qualified applicants experience long delays in getting their mortgage approved because loan officers submit applications with incomplete or incorrect information. They then have to call back the applicants to complete the information. Not only does the cost of each transaction soar as an application is submitted, revised, and partially processed again and again, but applicants understandably become annoyed. The problems are compounded at many institutions by decentralized systems that lack standardized processes, measurements, and training for frontline employees.

Lean manufacturing techniques can produce dramatic improvements for mortgage processors in this position. Less than a year after adopting them, one large bank saw direct costs fall by more than 30 percent, bringing it level with top-flight performers. It achieved this result by standardizing products and processes, introducing upfront quality checks of applications, simplifying supporting data systems, and harnessing scale economies through a tenfold increase in the number of applications handled per site per year. On top of these savings, the bank also cut average processing times by 20 to 30 percent,

* Despite this, the absolute cost saving was minimal. Whereas previous productivity efforts had focused on reducing overall cost, the lean manufacturing initiative concentrated on the margin produced per unit of capital.

and began approving many applications at the point of sale. It was able to exploit these benefits of speed and convenience to expand sales.

Similar results may be achieved in loan approval and other risk underwriting. What's more, they can be captured quickly: the process can be completed in as little as a year, and the bottom line show the benefits within eight months.

Our operators are standing by

Call centers are growing in importance as a sales and service channel for retail banks. They already represent 5 to 10 percent of non-interest expenses, and in many cases call volumes are growing by more than 20 percent a year. They have tended to proliferate in an *ad hoc* fashion in response to market demand, with the result that many banks have a collection of separately managed, business unit–driven operations that are often sub-scale, inconsistently run, and erratic in performance. Many also fail to disseminate knowledge about best practice.

One regional retail bank discovered it had more than 40 call center operations managed by at least ten different business areas. Management practices and tools, performance standards, and infrastructure all varied, and overall productivity and quality were well below leading industry benchmarks.


There are dozens of management practices, most based on lean manufacturing principles, that can improve call centers' performance. Many have become common management tools or standard practice in call centers. Examples include:

- ◆ **Automated workforce scheduling** – software tools that enable management to use call history to predict how many agents they need and provide a just in time inventory of customer service capacity. This tool helps to reduce the time agents spend waiting for calls and, when combined with the use of part-time employment, is effective in minimizing labor waste.
- ◆ **Automated voice response units**, which offload simple customer enquiries from call center agents to an interactive computer.
- ◆ **Call scripting**, which provides a standardized best-practice approach to customer interaction, helping to minimize labor waste by keeping the length of calls on target and ensuring consistency in their quality.
- ◆ **Automated call routing**, which channels call traffic between locations to match demand as closely as possible to agent availability.
- ◆ **First-call resolution procedures**, which try to ensure that a customer's concern is handled on his or her first call. This reduces labor waste and improves service quality.

- ◆ **The dedication of agents to telephone duties**, rather than having them also perform back-office activities such as researching customer inquiries or handling correspondence. This eliminates wasteful set-up activities such as moving to another workstation, launching a different computer application, or putting away and retrieving support documents.
- ◆ **Workstation sharing** in which agents on different shifts use the same workstation, helping to minimize capital waste.
- ◆ **Automated predictive dialers** for outbound calls that automatically select and dial target customers and prospects from lists so that agents don't waste time finding and dialing numbers.

These techniques should help banks make cost savings of 15 to 25 percent – and become more responsive to customers at the same time. Largely by implementing approaches based on lean manufacturing, one large retail bank cut the cost of each call it handled by 20 percent, enabling it to accommodate a 15 percent increase in call volumes without hiring more staff. Half of the savings came from reducing the time agents spent on unproductive tasks, and another 20 percent from shortening the average length of calls. At the same time, service levels rose by 30 percent.



All these cases point to the remarkable improvements in productivity, service, cost, and quality that lean manufacturing techniques can bring about at financial institutions. This approach has already transformed the way manufacturing is carried out across the globe. It could do the same for the banking industry. 

BANKS NEED FEWER CHECKS, NOT FEWER BRANCHES

Branch density, staff productivity, a nation's stage of economic development, and the share of its financial assets controlled by banks have all been held responsible for variations in the cost of national banking systems. Of these factors, the first two have attracted the attention of individual banks searching for cost cuts, because they are matters that lie within their control.

But there is another factor that banks control, but to which they have paid scant attention: the pronounced link between the type of payment system used in a given country and the cost structure of its banking. Specifically, the more extensive the use of checks for non-cash payments, the higher the cost of the banking system. So while painful branch and staff cuts may, in some cases, have helped reduce costs, a more effective means to that end might have been to wean customers away from their checkbooks.

To take the aggregate evidence first: those OECD countries where checks are widely used for non-cash payments (the United States, Canada, Australia, and the United Kingdom) tend to have more costly banking systems than the European countries that rely to a greater extent on electronic transaction methods or non-bank giro systems (Switzerland, the Netherlands, and Germany). This link is stronger than that between the cost of national banking systems and any of the other factors traditionally used to explain differences in that cost.

Take branch density, for example. Some high-cost countries, such as the United States and the United Kingdom, have lower branch density than low-cost countries such as Switzerland and Belgium. What is true for an entire banking system also holds for individual banks: check use is important in determining overall cost structure. Banks with a low level of check use and a high level of electronic transactions operate with substantially lower costs than banks with a high level of check use.