



System Dynamics Group
Sloan School of Management
Massachusetts Institute of Technology

System Dynamics II, 15.872
Prof. John Sterman

Assignment 3

The Dynamics of Service Quality

Assigned: Session 5; Due: Session 7
Please do this assignment in a group totaling three people.

This assignment focuses on the dynamics of service quality. Services are an increasingly important sector of the economy and source of competitive advantage for firms of all types. You will learn about some of the dynamics of services by building a model of service delivery and quality using the financial services industry as an example. This assignment deepens your skills in modeling and analysis, including formulation of intangible variables and nonlinear relationships, analyzing model behavior, and model testing.

Background

The service sector now represents approximately 75 percent of US GDP and employment, and continues to grow as more heavy industry and manufacturing move offshore. As important as the service sector is, companies often fail to deliver high quality customer service. While the quality of most manufactured products has increased over the past few decades, the American Customer Satisfaction Index for important service industries shows great dissatisfaction: in 2009/2010, Airlines receive a score of 66 (out of 100); banks get a rating of 74; wireless phone service, 72; subscription television service, 66; health insurers, 75; hospitals, 73. Consumers give most service industry firms a grade of C or D, and satisfaction in many cases is not improving (for information on the ACSI, see <http://www.theacsi.org/>).

Nowhere is this dynamic more important than in highly competitive industries in which products and services have been largely commoditized. In those industries, service quality becomes a major differentiator and driver of sales. The personal computer industry provides a good example. In his book, *Direct from Dell*, Michael Dell notes “*We’ve found that pricing is only one-third of our customers’ decision-making process; the other two-thirds represent service and support*” (p. 143). High quality service delivery increases customer loyalty, leads to more repeat business and favorable word of mouth referrals to others who may then become customers. Poor service can destroy a firm’s brand and erode sales.

Services differ from manufacturing because they are produced in the context of a personal interaction between the customer and the server. Services are intangible, and the quality of a service interaction is necessarily a subjective judgment made by the individual customer. Feelings and emotions matter in the service encounter. Because customers have different backgrounds, knowledge, needs, and expectations, services are harder to standardize than manufacturing. Services cannot be inventoried, so balancing capacity and demand is more

difficult than in manufacturing. Perceptions of procedural fairness and respect are important. Customers do not evaluate service quality solely in terms of the outcome of the interaction (e.g., did the doctor correctly diagnose my illness, did I get better?) but also consider the process and experience (e.g., did the doctor listen to my concerns, offer empathy and understanding, take the time to hear me out, and treat me with respect—or did the doctor rush through the appointment as quickly as possible to stay on schedule and get to the next patient?).

Medical treatment is a classic example of a high-contact service—it involves intimate interaction between medical professionals on the one hand, and you, the patient, on the other. High contact services are those in which the process of service delivery and interpersonal interactions between customer and server are important to the customer’s experience and judgment about service quality, and in which such interaction is necessary to the delivery of the service. In this assignment, we’ll focus on service quality in the retail financial services industry. Financial services (including retail banking, investment management, and insurance) are also high-contact. For many people, financial transactions are a source of anxiety. The array of account types, investment vehicles, loans, etc. is bewildering. Fee structures and risks are complex, and the fine print difficult to understand. Whether you are entrusting your money to a bank or seeking to borrow, you want to be sure you are doing the right thing. How the firm’s employees treat you can matter as much or more than the interest rate you can get on your money market fund or the fees you will pay on your checking account. At the same time, many financial institutions have cut back on direct, face-to-face service to cut costs by consolidating more of their service functions into back offices, online banking and call centers where customer requests are increasingly handled by telephone and internet. Managers are told “do more with less.”

Case Study: UniversalGloboBank

Consider a large retail bank we will call “UniversalGloboBank” or UGB. Some years ago, to lower costs, UGB created a number of “lending centers” (LCs) to handle their retail and small business loan operations, including credit cards, lines of credit, personal loans, etc. A typical LC serves several hundred thousand to about a million customers in a particular region, and operates much like a call center. Requests for service arrive at the LCs from existing or potential customers, or via referral from bankers in branches. For example, an existing customer may call or go online to request an increase in the credit limit on her credit card. A new potential customer may apply at a branch, by phone, or online for a personal loan or car loan; the application is sent to the LC for consideration. Hence work arrives at the LCs by phone (customer inquiries or calls from bankers in one of UGB’s branches), by mail, email and online chat (customer requests and communications with branches), by web (loan applications completed by customers or bankers in branches), and in the form of automated computer-generated reports identifying problematic accounts that require action, such as overdrafts, delinquencies, etc. LC employees must evaluate the request, including checking credit scores, account histories and references. Most requests require LC employees to produce an email, letter or phone conversation with the customer. Often they must call the customer or others to get missing information or correct erroneous information. LC staff are also trained to use customer calls to learn more about the customer’s needs and financial situation, a process called “profiling”, and to offer them additional products or services, a process called cross-selling. For example, through profiling, the LC employee may learn that the customer runs a small business, owns a home, and has small children, then offer a line of credit for the business, a home equity loan, and a college-savings plan. Cross selling brings in significant revenue.

As in most call center operations, the managers of the LCs are evaluated on the basis of their costs and the average time taken to respond to requests. In turn, LC managers closely monitor costs, employee productivity (cases handled per employee per day) and the time taken to close out cases. The LCs have a strong norm that all cases are closed out in one day. Cost, closing time, and productivity data are reported to the LC managers daily. LC managers also receive feedback on customer satisfaction, but less often: they receive a monthly survey of customer satisfaction based on telephone surveys of a random sample of UGB customers. The surveys are done by a large public opinion research firm, which has been on retainer to the bank for years. LC managers report that the customer satisfaction data are out of date by the time they get them, and neither reliable nor useful.

LC employees report high pressure to close cases and boost productivity. They report that they often have to work uncompensated overtime to meet their targets. For example, two employees report

“I don’t claim it all in overtime. I tend not to claim for work I do before the eight o’clock start, nor for the lunch hour [an average of 5 hours/week].”

“[My coworkers and I] don’t always claim that overtime either. I suppose that [we’re] worried that someone would say ‘you are not working very cleverly’ or something. I never go out to lunch; I’m giving the bank five hours a week of [unpaid] overtime.

High schedule pressure (pressure to close cases within the one day target closing time) means employees often cut customer interactions short, or fail to follow all recommended procedures in checking credit references. LC staffers know that they are not able to provide good service or cross sell when schedule pressure is high:

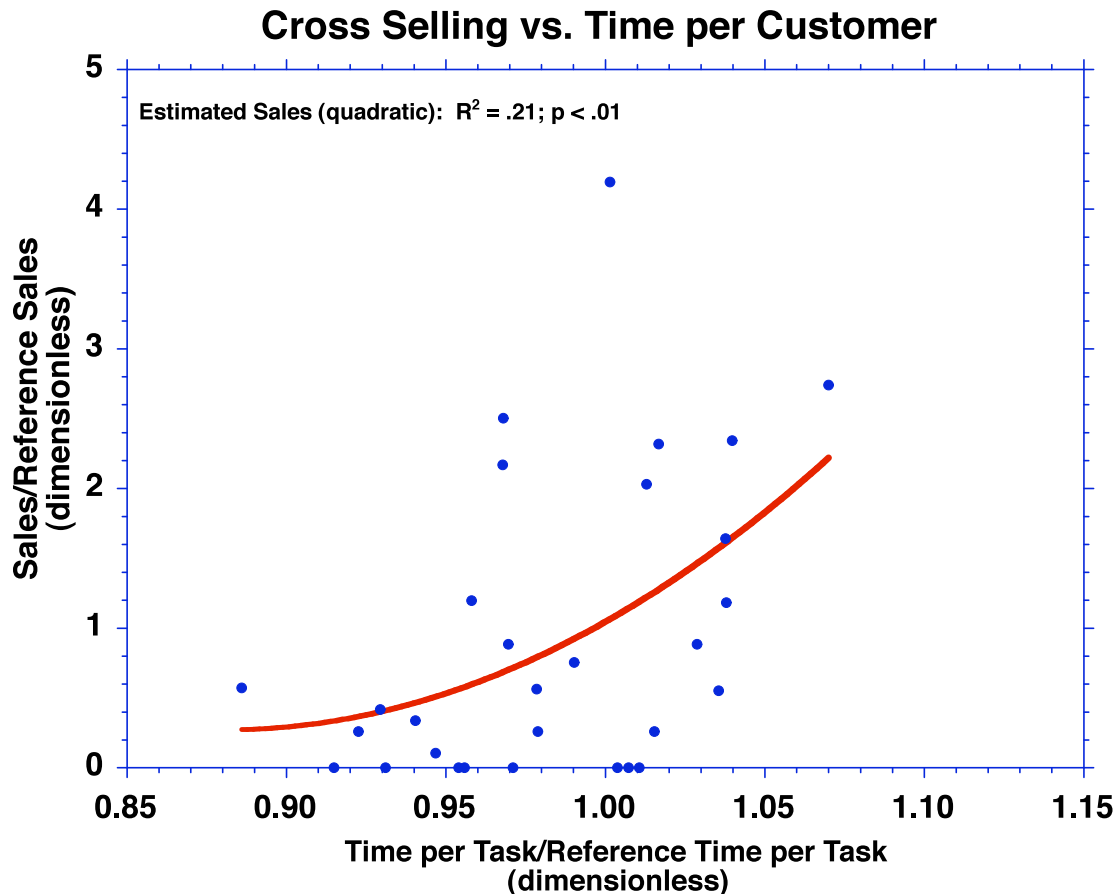
“The feedback you get back [from the customer] is ‘I’m dehumanized, I just became a number. I can no longer talk to you as a person, you just treat me as a number.’ [we] have lost the customers along the way.”

“We just don’t have the relationship basis to sell effectively. The customers have said that they become a number, and in a way they have. ...It is difficult to sell that way.”

Quantitative data back up these impressions. Figure 1 (below) shows average LC revenue from cross-selling as a function of the time per task (time spent on each customer request). The data have been normalized so that revenue from cross-selling equals its reference (average) value when the time spent on each customer equals its reference (average) value. The graph also shows the best-fitting quadratic curve. Although there is a lot of variability in cross-selling revenue, there is a highly statistically significant relationship: the longer each LC employee spends with each customer, the greater the revenue from cross-selling.

The budget for each LC determines how many workers they can hire. The budget for each LC is determined by the revenue generated within the region served by that LC. Employee turnover in the LCs is high, averaging 30-60%/year, and employee morale is often low. Absenteeism (workers who arrive late for their shift, or fail to arrive at all) is common.

Figure 1. Dependence of Cross-Selling on Time Spent with Customers



Getting Started

A. Begin your modeling effort by replicating the model shown in Figure 14-6 of *Business Dynamics* (p. 564). To speed your work, we have created a template for the model which you will find on the course website. The model, called “ServiceDelivery.mdl”, corresponds to the diagram shown in Figure 14-6, but does not include the equations. Use your judgment and the discussion on pp. 563-569 to formulate the equations for the model, paying special attention to dimensional consistency. For instance, the workforce is measured in people, the workday in hours/day, and time per task in person-hours/task. Knowing this you should be able to arrive at an equation for Potential Completion Rate that gives you units of tasks/day. The model includes two nonlinear functions (the effect of schedule pressure on workday and on time per task). Before formulating these, be sure to read pp. 551-563, with special attention to Table 14-1. Then read section 14.3, which details how the two nonlinear functions in the model are formulated. Use the values for the table functions given on pages 571 and 572 in your model for the effects of schedule pressure on workday and the effect of schedule pressure on time per task.

To speed your work, we have also modified the model slightly compared to Figure 14-6:

- ⇒ The model shown in Figure 14-6 measures time in weeks. For our purposes, we will measure time in days. Consistent with the use of the day as the unit for time, instead of “workweek” and “standard workweek”, the model variables have been changed to “workday” and “standard workday”. In the model posted to Stellar, we have pre-set the time step to 0.125 days, and the length of the simulation to 365 days (see the “Settings” menu).
- ⇒ Based on the description of UGB’s policies above, we know that the Target Delivery Delay is one day. Observation of LC workers shows that the Minimum Delivery Delay is one-quarter day. The standard workday is 8 hours/day. Standard Time per Task is one person-hour/task. The nominal headcount of a typical LC is 100 people.
- ⇒ To aid model testing, the model includes two test generators: one for the Task Arrival Rate and one for absenteeism. The test generators are found on the Task Arrival and Absenteeism view of the model. Using the test generators you can select a variety of inputs for the task arrival rate or absenteeism, including steps, pulses, ramps, cycles, and random variation. The test generator for task arrivals is set up so that the initial Task Arrival Rate equals the initial value of the Standard Completion Rate: tasks arrive at the rate equal to the LC’s ability to process those orders at the rate given by their initial head count and the standard workday and standard time per task.

Absenteeism is modeled as:

$$\text{Net Labor} = \text{Labor Force} * (1 - \text{Absenteeism})$$

$$\text{Absenteeism} = \text{MAX}(0, \text{Input}_0)$$

Where the Labor Force is the nominal LC headcount, Absenteeism is the fractional reduction in the actual number of employees working at any time, and Input_0 follows pulses, steps, ramps, cycles or noise according to your choices. The MAX function ensures that absenteeism lowers net labor (people sometimes fail to appear for their scheduled shifts), but does not increase net labor (people do not show up to work when they are not scheduled to do so, and net labor cannot exceed the labor force).

- ⇒ The test generators for task arrivals and absenteeism allow you to include random variations as a test input. The structure to model noise is called “pink noise” because realistic noise processes are autocorrelated. To use the noise input, you must set both the standard deviation for the noise and the Noise Correlation Time. The correlation time captures how much persistence there is in the noise from day to day. Please read *Business Dynamics* Appendix B to learn more about pink noise and autocorrelation. The data for UGB show that the noise correlation time for task arrivals is 7 days, and the correlation time for absenteeism is 14 days. These values have been set in the model; you can vary them to explore the sensitivity of the models response to different degrees of persistence in the noise inputs.
- ⇒ Formulate the initial task backlog to ensure that the model always starts in equilibrium regardless of the initial Task Arrival Rate. To do so you must set the initial value for every stock to an algebraic expression, not a number. Use the equilibrium condition for backlog (task completion = task arrival) to derive an algebraic expression for the equilibrium backlog, assuming that the task completion rate equals its desired rate.
- ⇒ As always, you must fully document your model by writing brief but informative comments in the comment field for *every* variable and constant in the model.

❑ 1. Explain the shape of the table functions for the effect of schedule pressure on workweek and time per task *using language a manager would understand*. Pay special attention to reference lines and the behavior of the functions for extreme values of schedule pressure.


❑ 2. As described above, LC managers receive frequent feedback on labor productivity and monitor it closely. Add a new variable to your model to compute labor productivity. Labor productivity is the number of tasks completed per day per person.

B. A valuable method to explore the behavior of models is to start in equilibrium and then shock the system with a known perturbation such as a sudden increase in workload. Use the test input generator to run this test by having the task arrival rate step up by 20% on day 5. Run the model and answer the following questions.

❑ 1. How does the simulated LC respond to the sudden increase in tasks arriving? Explain this response in terms a manager familiar with the industry would understand. Specifically, how do LC employees respond to the increase in workload? What happens to productivity?

❑ 2. What happens as the size of the step in task arrivals increases? Does the model reach equilibrium for all step sizes? Why or why not? Explain the relationship between task arrivals relative to the organization's nominal capacity (the standard completion rate) in terms managers would understand.

⇒ Use Vensim's Synthesim mode to quickly find the values of equilibrium delivery delay, workday, time per task, productivity, and other variables.

⇒ Launch Synthesim by clicking on the  button in the top toolbar. Use the slider for Step Height in the test generator to set different size increases in the arrival rate. You can set exact values for any input slider by clicking on the arrow at the end of the slider, then entering your desired value in the dialog box that appears. Once you set the step height, use the table tool to find the final (equilibrium) value of delivery delay.

What do you conclude about the performance of the system in this initial model as the workload increases? You may want to explore the response of the system to other test inputs, including random variations in task arrivals. Read about noise inputs in *Business Dynamics* Appendix B.

C. Expanding the model boundary: The initial model does not capture a variety of important feedbacks affecting the performance of the LC. In this section, you will relax some of these assumptions, one at a time.

C1. Revenue: As described above, UGB derives significant revenue from cross selling. Use the information below and the case description above to formulate equations for LC revenue.

⇒ Total LC revenue consists of a base revenue level plus revenue from cross selling. Base revenue comes from account maintenance fees, other fees and interest income generated by the customer base in the LC's service region, and averages \$12,000 per day (about \$4.4 million/year).

- ⇒ Revenue from cross selling is determined by the number of customer inquiries (tasks) completed each day and the average cross sell per customer inquiry (cross sell per task). Average cross selling revenue per task depends on the time spent per task, as shown in Figure 1. Formulate cross sell revenue per task using a table function. Formulate the table using the principles for table functions described in the text. In particular, normalize the function as follows: set cross sell revenue per task so that it equals a reference value when time per task equals a reference value. Define these reference values as constants.
 - ⇒ The data indicate that when time per task equals 1 person-hour/task, then average cross sell revenue per task is \$15/task. Use these as the reference values. Then use the data and best-fitting curve in Figure 1 to specify the values of the table function relating time per task relative to the reference value to cross sell revenue per task relative to its reference value.
 - ⇒ Your function should correspond to the values shown by the best-fit curve in the figure, but be sure to pay special attention to the shape of the curve outside the range of historical data. Consider extreme conditions: what must cross sell revenue per task be if time per task is zero? What must happen to cross sell revenue as time per task becomes many times greater than the reference?
- ❑ 1. Run the model with various size step increases in task arrivals. What happens to cross sell revenue? Why? Explain in terms of the feedback structure of the system, but in terms a manager can understand.
 - ❑ 2. Now run the model with a constant task arrival rate, but random variations in absenteeism. A standard deviation of 0.02 (2%) in the noise input to absenteeism is reasonable. What is the impact on revenue? Why? Explain.

C2. Organizational Norms for Time per Task: So far the standard time per task has been treated as a constant. Constant standards are appropriate in some settings, such as manufacturing, where processing times are tightly determined because so much of the work is automated and routinized. In high-contact service settings, however, it is very difficult to determine an appropriate standard for the time each employee should spend with each customer. Customer needs and knowledge are heterogeneous, server skill varies, and the customer's perception of quality depends strongly on how much time and attention they receive from the server. In such settings, workers' norms for the appropriate amount of time to spend on each case tend to adjust over time to the actual amount of time servers spend with each customer.

- ⇒ Norms that adjust to past performance are known as "floating goals". Floating goals are common: when sales people exceed their sales quotas, management tends to raise them; students sometimes adjust their aspirations for grades to the actual grades they receive; your belief about how much income you need to live comfortably tends to adjust to your actual income; your belief about your optimal weight tends to adjust to your actual weight. Read more about floating goals and how to model them in section 13.2.10, pp. 532-535.
- ⇒ In service settings, where it is difficult or impossible to determine the "correct" standard for the time each server should spend with customers, norms for customer service tend to adjust strongly to actual performance.

- ⇒ Modify your model to capture variations in the standard time per task. In particular, if LC employees find they are consistently spending more (less) time with each customer, the standard time per task—what they and management consider to be appropriate—will gradually rise (drop) until it equals the actual time they are spending.
- ⇒ Specifically, model the standard time per task as a stock that adjusts towards the current time per task over some adjustment time. Fieldwork suggests the average adjustment time for standard time per task is 90 days. Set the initial standard time per task at 1 person-hour/task. Document your formulation, check it for dimensional consistency and test it to make sure that it behaves plausibly.

- 1. How does your additional structure change the response of the system to a 20% step increase in task arrivals? Does the model reach equilibrium? Note: you may need to run your model longer than 365 days to determine if the system reaches a new equilibrium.
- 2. Explain the behavior of the system in terms a manager can understand. Discuss how the equilibrium of the system changes relative to the original model with a fixed standard time per task. How does the organization adjust to the increase in workload? What are the impacts of these differences on productivity? On cross-selling? What do these changes represent in terms of customer service quality, and what other feedbacks would you expect these changes to have in the real system? (A simple causal diagram will be helpful here.)
- 3. Now run the model with a constant task arrival rate, but random variations in absenteeism (use a standard deviation of 0.02). What happens to the norm for time per task, to productivity, and to cross selling? Why? Explain in terms of the feedback structure of the system.

C3. Budget Constraint on Hiring: The number of workers in the call center is still exogenous in your model. Relax this assumption by modeling the dynamics of the labor force.

- 1. As described above, Lending Center managers must strive to staff their centers so that headcount is sufficient to complete work at the desired rate, but they also face a budget constraint. Model the labor force of the LC as a stock with explicit hiring and quits. The LC managers replace employees who quit and adjust the labor force to a desired level. The desired level is determined by the number of people needed to meet the desired task completion rate or the number of people the LC can hire given their budget, whichever is less. The number of people needed to meet the desired task completion rate is based on the desired completion rate, standard workday and standard time per task. In modeling how many people the LC can afford to have on staff, assume that the average daily cost per worker is \$160/day (about \$58,000/year, a figure that includes salary, benefits, and overhead costs). The budget for the LC is a certain fraction of the revenue generated in their service area. Two-thirds (67%) of that revenue is allocated to the budget of the LC, with the rest going to UGB to cover indirect costs and contribute to profit. Note that when time per task equals its reference value, the budget should be sufficient to support the number of people needed.
- ⇒ You can use the labor sector you developed in the Widgets model to model the labor force. Remember that time is measured in weeks in your Widgets model, but days in the service delivery model: make sure you set the parameters appropriately.
- ⇒ Compare the head count determined by the budget with the head count required to process the work at the desired rate. Be sure that the LC's initial budget is sufficient to provide


enough workers to complete the work at the desired rate and with the standard workday and initial standard time per task.

- ❑ 2. Test the model with a variety of step increases in incoming tasks. Explain, in terms managers would understand, which feedback loops are most important in generating the behavior you observe. Consider the behavior of productivity, revenue, head count, and other variables. *How do you think LC managers would interpret the results?* Explain.
- ❑ 3. Now test the model in response to other inputs. Try a ramp in task arrivals. A ramp with a positive slope corresponds to growth in the overall number of customers in the LCs service area. Try a slope of 0.0002/day (a little more than 7%/year growth in volume). How does the LC meet the increase in work volume? What other impacts does growth have?
- ❑ 4. Keeping the task arrival rate constant, run the model with random absenteeism (use a standard deviation of 0.02). Explain the resulting behavior, considering productivity, revenue, head count, and other variables. Show graphs of model behavior to illustrate (you should plot those variables needed to show why your explanation is correct; use your judgment).

D. Policy Implications: Based on your model results, what policies do you recommend to UGB’s senior management to avoid the pitfalls your model indicates may exist?

- ⇒ Your policy recommendations must be specific and implementable. It is not acceptable to say “maintain quality standards” or “keep standard time per task constant.” The first suggestion is not sufficiently operational: *how* would you maintain quality standards? The second is infeasible: because the mix of products offered is constantly changing, along with customer knowledge and expectations, a fixed standard is impractical. If product complexity grows, the appropriate standard time per task may increase; if technology improves so that more of the work can be automated, then standard times might fall without compromising service quality and cross-selling. Furthermore, each individual server forms their norm for time per task based on their own experience; management has only limited influence in the goal setting process.
- ⇒ Effective policies will consist of one or more of the following:
 1. Changes in model parameters that strengthen or weaken existing feedbacks.
 2. The elimination of an existing feedback loop.
 3. The addition of a new feedback loop or loops.
 4. Changes in the goals of the different loops.

In all cases your policies must be implementable. Explain why you think your recommended policies would work, and how they could be implemented in the real world.

E. Include a full documented listing (using the Document All  button) of your final model in your writeup. Submit both a paper printout of your writeup (with model diagram) and an electronic copy of your model file (the .mdl file) and writeup on the class website.

MIT OpenCourseWare
<http://ocw.mit.edu>

15.872 System Dynamics II
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.