School of Public Policy
George Mason University

Report On Peak Load Cost Modeling

Robert Cohen
Charles McBride
John Panzar
John Waller

October 7, 2011

Table of Contents

Report On Peak Load Cost Modeling

1. Summary and Principal Findings

Working under a competitively awarded contract, the George Mason University (GMU) technical team has conducted a study of peak load cost methods and issues. The principal findings are:

- Extensive literature is available on peak load pricing but virtually none exists on peak load costs

- Peak load costs can be analyzed as the difference in optimized operations with and without real world constraints on the use of labor and capital inputs in operations under conditions of uneven demand for these inputs because of mail arrival patterns, service requirements and labor agreements.

- Service standards and mail arrival patterns are also constraints whose costs can be analyzed in a similar fashion

- Costs with and without constraints can be calculated with further development of the analytical and linear programming models developed for and presented in this report for a sample of cases.

- Much of the data for these linear programming models can be obtained from current data available to the Postal Service.

- Other analytic techniques can be identified, such as econometric analysis, but the data requirements may be more burdensome that the linear programming approach.

- The models can also be used to estimate cost savings from the significant changes in service standards and network consolidations that the Postal Service has announced.

This report presents the models in an analytical form and as (numerical) linear programs in Excel with examples of their use for mail processing, delivery and transportation cases.

2.   Introduction

The work was performed in the four tasks identified in the RFP and used to organize the team's technical approach in the winning proposal.

1. Discuss Peak Load Issues with Commission Staff (7 days after contract award)
2. Describe Peak Load Cost Methodology (within 20 days of contract award)
3. Describe the Data Necessary to Execute any Peak Load Cost Modeling Strategies (within 20 days of contract award)
4. Report on all Known and Proposed New Methodologies for Calculating Peak Load Costs (within 20 days of contract award)

Task 1 involved a meeting with Commission staff that occurred on September 9 in which the Commission staff clarified the objectives of the study. Tasks 2 and 3 are the subject of this report. It characterizes peak load costs and describes the linear programming modeling techniques adopted and identifies an alternative econometric approach.   The authors gratefully acknowledge the assistance from Professor Frank Wolak in reviewing an earlier draft of an econometric analysis approach section and augmenting it with text contained in Appendix A, Section 2.1.2. Task 4 was fulfilled, with a briefing to the Commission on September 30. A copy of the Task 4 briefing is attached in Appendix D.

Appendix A contains a mathematical analysis the Peak Load Cost issue. Appendix B presents a numerical approach using linear programming techniques. Appendices A and B have independent characterizations of the peak load cost issue. While this creates some redundancies, it also permits each appendix to stand-alone. Appendix C describes the Excel

worksheets used to demonstrate the feasibility of producing numerical estimates of peak load costs using linear programming.

3.   Approach

Our approach was to first draw on the project's staff's extensive experience with Postal Service operations and costs to develop a preliminary definition of peak load costs. At the same time we conducted a literature search to identify previous modeling work that could be applicable to the Commission's request. It was concluded that postal peak load issues are fundamentally different from peak loads in other contexts such as electricity demand. A Google Scholar search of the postal literature yielded *301* papers with *peak load pricing* in the title, but only *one* paper with *peak load costs* in the title:

> **Ofelia Betancor, Juan Luis Jiménez, and M. Pilar Socorro,** "The costs of peak-load demand: An application to Gran Canaria airport," *Journal of Air Transport Management*, Volume 16, Issue 2, March 2010, Pages 59-64

For a potential peak load cost modeling approach, the following was found applicable

> **Michael A. Crew, Paul R. Kleindorfer and Marc A. Smith, "Peak-Load Pricing in the Postal Services," Economic Journal, 100, (September 1990), 793-807.**

We also met with postal officials on two occasions (at District offices and at USPS headquarters) to review USPS modeling efforts and available data on operations. The discussions identified several proprietary workload assignment models that might have some applicability in mail processing, delivery and transportation scheduling. Publicly available details concerning these models were not identified but knowledgeable contacts have been identified should further development work be pursued under Task 5.

We reviewed recent documentation on Peak Load cost issues, including:

"Advisory Opinion on Elimination of Saturday Delivery" in Docket N2010-1, especially Chapter VII Peak Load Issues. Issued March 24, 2011 and available at prc.gov.

"Cost of Service Standards" report issued by the U.S. Postal Service Office of Inspector General, Report number RARC-WP-11-008. Of special interest is Chapter 2,The "Peak Load" Issue and Volume and Workhour Profiles for the Postal Service, in an attached analysis by Christensen Associates. Both the report and supporting analysis are available at uspsoig.gov.

Much of this background research helped formulate the discussion with Commission staff required in Task 1. The first discussion occurred in a meeting on September 9 at the Commission. Subsequent telephone conservations and email exchanges occurred throughout September. A meeting at the Commission was held on September 23, and a briefing of Commissioners and staff took place on September 30 to summarize our findings.

John Panzar developed the analytical models for this report and Charles McBride developed the linear programming models. The model development effort helped the team to identify specific data requirements and methods for addressing the questions that the Commission has raised.

Shortly after this study was initiated the Postal Service announced plans to significantly shrink the mail processing network and modify delivery service standards. Since many peak load issues occur because of existing service standards, the planned elimination of overnight service is likely to mitigate some peak load costs and create costs savings. The proposed models can be tailored to assist the Commission in the analysis of these Postal Service plans as part of the expected filing of a request for an advisory opinion

4.   Impact on Study of Postal Service's Announcement of Plans to Modify Operations and Service Standards

The essential feature of the new plan is to change the current 1 to 3 day service standard to 2 to 3 days (i.e. eliminate the overnight delivery standard for preferential mail.)[1] This in turn will allow the processing of originating and incoming mail on the day shift of the next day and a reduction o the number of processing plants by more than half.[2] Without the overnight commitment, there should not be a significant hourly peak in mail processing which occurs now. Currently, the Postal Service must process all mail originating in a plant's service area in time to meet the critical dispatches for mail with an overnight service standard during the latter half of tour 3 so it can be delivered the next day. There is a short window beginning about 4 to 5 PM and ending a few hours later to finish the outward processing of mail so that the portion with an overnight service standard can be put into containers and transported to the destinating plants or set aside for sequencing on tour 1. The destinating plants do inward processing on that mail and dispatch it to carrier stations for delivery the next morning in time to be delivered that day.

The overnight standard creates the hourly peak on tour 3 in the originating plant. There is a related peak in both the originating and destinating plants on tour 1 when all incoming and originating overnight mail must be sequenced starting around midnight and finishing about 4 AM to be ready for transport to delivery units. .

The proposed plan would store mail originating on day zero (the day it enters the plant) and process it the next day (day one) during the period when there is no night time differential (between 6 AM and 6 PM.)[3] The plan envisions outward processing to begin on

---

[1] It is not clear at the time of this report whether time value periodicals will still be able to be dropped at plants for overnight delivery. However, this is a small portion of total volume.

[2] The overnight service standard requires a reasonably close proximity between plants in an overnight service area so that mail can be dispatched and arrive at the incoming plant in time for processing and next day delivery. Consequently, the service area of plants can be expanded when the overnight standard is eliminated and this in turn allows for a reduction in the number of plants.

[3] Very little processing occurs during this period in the current operating plan. For example, the recent Christiansen Associates report shows that only one of the eight hours on tour 2 has more than 100,000 dollar weighted clerk IOCS tallies. The rest of the hours in that tour have about 50, 000 tallies. In contrast, all the hours in tours 1 and 3

tour 2 at about 8 AM and sequencing for carriers to begin about noon and continue till about 2 AM[4]. This means that originating mail on day zero would be processed on day one and delivered on day two. Mail that has a current overnight standard from other plants could arrive later (on day one) because it would be held for tour two processing at the originating plant. That mail would be delivered on day 2. The proposed plan would be compatible with the current two and three day standard. Most of that mail is not shipped from the originating plant until the next day currently, so there would be little change.

5.   What are Peak Load Costs?

Peak load costs occur because of constraints placed on the ability of the Postal Service to handle the flow of mail into and through the postal operations in the most efficient manner. The difference in costs between the constrained situation and the unconstrained, or less constrained situation, becomes the peak load costs due to the constraints being examined. Conversely, it is the cost savings from removing or mitigating the constraints. For example, without overnight service standards, mail could be stored and processed without the cost impact of night differentials or expanded processing capability.

With this definition, the cost of the constraints is the difference between the cost of the optimal operating cost with and without the constraints. To perform the calculation, one needs a cost function to be minimized with associated constraint equations. If the cost function and associated constraints can be expressed as linear relationships, then linear programming techniques can be used to find optimal solutions for the cost function with and without the constraints.

Fortunately, linear relationships can be used for mail processing, delivery and transportation operations. To demonstrate the feasibility and value of this approach, the

---

have at least 150,000 and 11 of the 16 hours have more than 250,000 tallies. They range as high as 400,000 tallies. (See page 22.)

[4] Machine utilization would increase under the new plan. Currently, sorting machines run from about 5 PM to about 4 AM or about 11 hours a day. Under the new operating plan they would run about 18 hours.

study team has developed both analytical and empirical models for these postal operations that are presented here and in the Excel worksheets submitted along with this report.

The three operational areas identified in the RFP are used to organize the discussion here of peak loading issues and modeling in the mail processing, delivery and transportation functi

6. Modeling Approach

After a review of the modeling literature for estimating peak load costs, the study team has settled on linear programming approaches. A linear program calculates the lowest cost solution while satisfying the relevant constraints. In a sense it generates a theoretical or normative (should) cost result. This result can be compared to actual results to get a measure of efficiency. It can be used to estimate the cost of labor constraints by running it with and without the labor constraints. It can estimate the cost of peak loads by changing the dispatch schedules or the arrival pattern of the mail and running the model with a smooth workflow. It can estimate the cost of a change in service standards by running it with the old and the new service standard and then comparing the results. As discussed here and in the appendices, linear programming provides a straightforward approach to describing decision variables and constraints involved in mail processing, delivery and transportation functions.. Its use is described separately for each function and with examples of its use given in the appendices.

6.1 Mail Processing

We model peak load and other mail processing issues by constructing a linear programming model that minimizes staffing and equipment costs subject to perturbations in workload. These are caused by the fluctuations in the arrival of mail coupled with service standards that cause the Postal Service to handle more mail at certain periods than others. The linear programming approach (see Appendices A, B and C for more details, including preliminary examples) will use actual union contract constraints and will parameterize the number of machines available.[5] The

---

[5] The actual number of machines available would be used if a specific plant were being modeled.

user of the model will input: the total volume (if desired, by preferential and non-preferential and by shape), the hourly and daily arrival profiles for low and high volume periods of the year, the labor categories and wages, the sorting machine capacity, the labor constraints to be examined, and final dispatch times for each category of mail. The model then identifies the least cost labor schedule using the different categories of labor and the constraints.

The figure shown below from the Christenson report on changing service standards (Figure 14 in the report) illustrates the fluctuations in the demand for clerk labor for the Postal Service as a whole. It can be seen for example, that the Service uses approximately 8 times as many clerk hours between 10 PM and 11 PM on a regular work day as it does between 12 PM and 1 PM[6].

We can think of the figure as representing a typical plant's demand for labor given the arrival pattern of its mail and the dispatch schedules it is subject to. The model will calculate the lowest cost schedule of employees using all the labor categories defined in the clerk union's contract with the Postal Service and subject to the constraints in the contract. The labor categories include regular clerk, part time clerk, part time flexible clerk, regular overtime, penalty overtime etc. The constraints include such provisions as that each regular clerk must be given 8 hours work in a 9 hour period and must be given two consecutive days off, etc.

Figure 14 – IOCS Costs by Time-of-Day, Mail Handlers at Mail Processing Plants

---
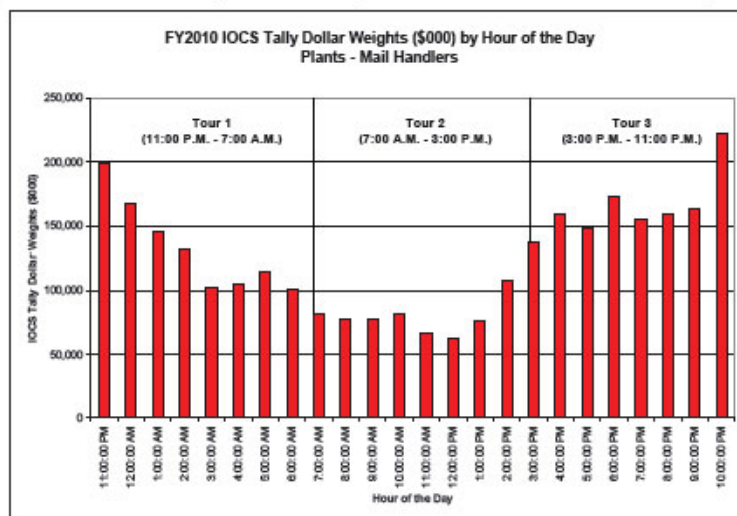
[6] Of course, the actual levels vary from day to day with the volume of mail.

The model can be used to estimate the costs for a day but a typical application would be for a week because some of the constraints can only be reflected in a week's period. The volume for each day of the week must be input. Seasonality can be represented by high, medium and low volumes as volumes are high in the fall and in the Christmas season while it is low in the summer. It should be born in mind that some of the labor constraints are different in the Christmas period than in other times of the year.

Appendix A describes analytic models that the study team has constructed to demonstrate general modeling techniques. It contains examples of peak load cost models that examine the impact of service standards, smoothing of the mail arrival and the existence of excess capacity. Appendices B and C present specific linear programming examples involving two types of preferential mail (originating and incoming), with overnight processing requirements, and with nighttime differential pay. Appendix B develops the linear programming model and Appendix C describes the Excel worksheets containing several examples of the numerical calculation of optimal linear programming solutions for the constrained and unconstrained situations. The actual peak load costs for these examples are calculated in the Excel worksheets described in Appendix C and submitted as a separate Excel file, LPExamples.xlsx

6.2   Delivery

There are three major types of delivery routes: city, rural and highway routes. City carriers are paid an hourly rate and can receive overtime payments. They are represented by the National Association of Letter Carriers (NALC). Generally speaking, rural carriers are paid a flat amount based on periodic assessments of the volume of mail on their routes. They do not get paid overtime except in exceptional circumstances. They are represented by the National Rural Letter Carriers' Association (NRLCA). Highway routes are procured on a competitive bid basis and are paid a fixed amount. They are represented by the Star Route Association The discussion here focuses on City Routes whose local supervisors make day-to-day decisions on whether to modify route operations in light of changes in volume and carrier absence. They can authorize overtime if needed.

City delivery operations are organized by routes that operate out of delivery units that serve one or more 5-digit ZIP code areas. Mail from a local processing plant is transported to the delivery unit and then distributed to the carriers for delivery on their assigned route. Most of the mail coming from the processing plant is also organized by carrier and most letters are in delivery point sequence (DPS) order. Those delivery units served by a processing plant with Flats Sequencing System (FSS) sorting machines also receive flats in DPS order. About one third of the routes now receive flats in DPS form.[7] The DPSed mail is ready to be taken to the street for delivery. The carriers in the delivery unit must sort non-DPSed mail by hand.

With the advent of DPSed mail, the carrier time in the office has been reduced from about four hours to two or less. This has reduced the overall variability of carrier costs with respect to volume, since in-office time has nearly 100 percent variability while street-time variability is about one third of that. A major reason for the variability of street time is the coverage factor (the percent of delivery points on the route receiving mail and requiring deviation from the route to the customer's mail box), and another important reason is loading the mail into a mail box which requires fingering the mail to isolate the pieces in the carrier's bundle for a specific address (called load time).

Typically, a carrier is assigned to a specific route. During the last few years, most city routes have been redesigned on an annual basis in response to decreases in mail volume that in turn has reduce the number of city routes. Between periods of route redesign, routes are fixed. However, each route consists of well-defined segments, such as a distinct neighborhood or a certain set of streets. On a daily basis, supervisors have the authority to reassign segments to different carriers to smooth the volume load or to accommodate absent carriers. This process is known as pivoting and is a form of dynamic route adjustment. In some postal districts, carriers can be reassigned between delivery units to further assist in handling imbalances in the distribution of carriers and mail.

---

[7] The new operating plan envisions sequencing flats for more routes with the same number of machines because of a longer sorting window.

Unlike mail processing in which mail volumes show hourly peaks, delivery volume variation is solely by day and season. As shown in Table IV-4 reproduced from the Commission's Advisory Decision in Docket N-2010-1, mail arrival varies by day of the week with Monday being the highest volume day. Saturday's average total volume is lighter at 855,548. Total city carrier workhours are virtually constant on the weekdays, ranging from 900,748 to 933,941 per day. This might support the Postal Service assertions that street time is virtually fixed as long as volume is below the capacity of the route. However, the productivity in terms of pieces per workhour varies significantly from a Tuesday low of 354 to a Monday high of 451.

Table IV-4

FY 2009 City Delivery Volumes and Carrier Street Hours by Weekday

	Col. 1 Volume	Col. 2 Hours	Pcs. Per Hr. Productivity (Col. 1 / Col. 2)
Monday	421,077,876	933,941	451
Tuesday	326,474,124	921,006	354
Wednesday	329,095,036	919,062	358
Thursday	334,472,819	908,866	368
Friday	335,126,740	900,748	372
Saturday	322,839,071	855,548	377
Total	2,069,085,665	5,439,171	380
Day After Holiday	487,192,012	952,020	512

Source: PRC-LR-N2010-1/2.

Once a set of routes is established for a year, the management problem is one of assigning carriers and other personnel to participate in the in-office and street time work so that costs are minimized subject to service and operational constraints. Issues that have to be addressed are the productivity of carriers, variability of carrier time with volume, use of overtime, use of part-time and auxiliary personnel, pivoting segments of the route, delaying some of the mail to delivery on a subsequent day.   Table IV-6 from the PRC Docket N2010-1 Advisory Opinion shows the percent of mail that is deferrable varies by day from a Monday low 25.6 percent to a Tuesday high of 39.1.

Table IV-6

Percent of FY 2009 City Carrier Average Daily Volume by Mode and Day

| | Weeks Without Holidays | | | |
	DPS Letters	Cased Flats	Sequenced Mail	Deferrable Mail (Cased Flats + Sequenced)
Monday	67.6%	21.9%	3.7%	25.6%
Tuesday	54.2%	24.5%	14.7%	39.1%
Wednesday	56.7%	23.5%	12.9%	36.4%
Thursday	62.2%	22.9%	7.8%	30.7%
Friday	63.3%	22.3%	7.3%	29.5%
Saturday	64.6%	22.8%	4.9%	27.7%

| | Weeks With Holidays | | | |
	DPS Letters	Cased Flats	Sequenced Mail	Deferrable Mail (Cased Flats + Sequenced)
Tuesday	67.4%	21.4%	4.5%	25.9%
Wednesday	53.8%	24.5%	15.1%	39.5%
Thursday	59.1%	24.4%	9.4%	33.8%
Friday	61.9%	22.9%	8.0%	30.9%
Saturday	64.3%	23.3%	4.9%	28.3%

Source: PRC-LR-N2010-1/2.

The study team has begun to demonstrate the feasibility of building models to calculate optimal allocations of labor resources in light of the actual constraints represented by the various operating conditions. For example, the requirement that carriers work an eight-hour day five days a week with known overtime rates can readily be built into the model, along with the options to augment with other types of labor, like delivery assistants or routers. One of Professor Panzar's modeling examples deals with the different types of labor.

An issue that will have to be parametrically analyzed is the impact of variability. The Service has basically claimed that street time is fixed relative to volume. However, cost analysis endorsed by the Service and the Commission indicates a street time variability of about 33 percent. Also, the fact that volume decreases have forced the Service to alter route configurations on an annual basis further demonstrates that volume variability is an important factor in carrier operations. If Task 5 models are developed, a means of testing for various variability levels should be built into the models.

The above data is for averages over a year. It is important to note that volumes differ by day of the week, month and season. The linear programming delivery model that the study team has identified can take into consideration these differences by using different volume sets.

The appendices to this report present preliminary models to demonstrate how models would be developed for the delivery function. Appendix A contains two exemplar analytic models developed by John Panzar. The first analytic model examines the peak load costs where the number of routes can vary by delivery day. The second analytic model examines the peak load costs when mailers are induced to smooth the flow of mail by delivery day. Appendix B presents a linear programming model that keeps the number of routes the same each day but allows carriers with different salary levels with constraints on the number of hours and days worked for one of the salary levels.

The examples in the appendices can readily be expanded to include other real world constraints and opportunities for significant operational changes as proposed by the Postal Service. Further development work planned under Task 5 in the original Commission RFP, but not included in this contract work statement, can build on these exemplar prototype models.

6.3   Transportation

The Postal Service's highway transportation system is perhaps the most complex in the world. It currently connects over 500 plants with each other (using hub and spoke and direct connection) and with some 30,000 thousand post offices. For the most part, the Service must contract for round trips (back haul) which frequently mean that capacity use is uneven as mail flows are not symmetrical. In addition, the transportation system must allow the Service to meet its service standards and this frequently requires trucks to be dispatched with less than full loads. We have developed an exploratory transportation model using linear programming that efficiently allocates trips to demand between points subject to service standards and their

required final dispatch times in each processing facility. The fully developed model could be used, for example, to estimate minimum costs for intra-SCF and Intra-BMC transportation involving a single point and many delivery units. This would be especially relevant to the proposed new operating plan that will eliminate the overnight standard for First Class. Other ways the model could be used is to estimate the reduction in the minimum cost for inter-SCF and inter-BMC highway transportation under the assumption that the number of processing plants will be cut in half as has been announced by the Postal Service. The model could also be used to estimate cost savings from moving mail volumes to highway from air assuming that the two-day service standard would be relaxed between designated city pairs. Finally, the models could be used to measure the cost of peak loads by examining costs with and without volume peaks.

Generally the model will examine daily scenarios with hourly volume availability in the staging area.[8] Minimum costs for weekly schedules could be estimated using a series of daily runs. As with mail processing and delivery, seasonal effects could be captured by running the model with high, medium and low volumes. Different size plants could be accommodated with separate runs using different representative volumes.

The study team has been informed by Postal Service staff that virtually all the truck trailers that it uses are about 42 feet long and that it must contract for back haul in most instances. Thus it appears that scheduling different size trailers is not an option.[9] The Postal Service now generally contracts for two years at a time. It has the ability to call for extra trips on a route and to cancel a trip. The study team would need cost information to incorporate this information into the model[10].

The model would require volume data between points including the arrival rates of mail into the staging area. It would also need the last dispatch of value information for each route.

---

[8] "Staging" refers to the area were mail is stored for dispatch after it has been processed..

[9] Moreover, the Postal Service staff informs us that the major cost in trucking is the driver and the using smaller trailers would yield little savings.

[10] The study team was also informed that these options were seldom used for reasons that were not explained.

7.   Data Requirements

Data requirements flow from the definition of the variables and constraints in the models. Based on the team's preliminary modeling work, initial data requirements can be identified to analyze peak load costs and savings from the new operating plan recently announced by the Postal Service. Data is required for the current operating plan and estimates for the proposed operating plan.

The number of facilities from which data should be gathered will define the number of facilities that can be modeled. In a world with unlimited recourses, data from all the facilities would be used. Of course, it is impractical to develop that many models and it would be an undue burden on the Postal Service to generate that amount of data. We think that data from a small, medium and large facility would be sufficient to generate usable results.

7.1   Mail Processing Models

The mail processing models require data on the arrival of mail by the hour and the dispatch schedule by the hour.. In addition we would need to know if these patterns are relatively constant over time. Processing peaks in a given plant occur by hour of a day, by day of the week, by day of the month, and by month or season as driven by arrival rates and dispatch schedules.   The volume flows and mail processing requirements differ by preferential or non-preferential status, by shape and by form of mail presentations (single-piece, pre-sort). In addition data is needed on the number of machines of various types and the productivity of the plant (pieces per man hour) productivity.

As a starting point, the following data should be obtained for a typical plant, such as the Merrifield, Virginia Processing and Distribution Center.

Mail arrival and flow rates by:

- Hour of the day for a week in a low volume week, like August, and a high volume week, like mid December, and also a normal-volume week.
- Broken out by
  - Preferential or non-preferential category mail
  - Shape and
  - Source (collection, bulk, and managed mail from other plants)
- Volume flow matrix of mail by shape from/to all MODS operations

Labor

- Categories of labor and average salary
  - Clerks
    - Full time
    - Part time
    - All categories identified in the new APWU contract.
  - Mail Handlers

- Productivity – average pieces per man hour for each MODS operation

- Physical Characteristics
  - Number and type of machines
  - Throughput in terms of pieces fed and pieces successfully processed per hour

Much of this information is readily available from the current records of a plant. For instance, mail arrival rates for bulk and managed mail should be available from a mixture of IMb data, Mail Permit Statements, FAST records, and mail surface visibility reports. Barcodes on containers are scanned, as the containers are unloaded from trucks to provide arrival time and information on contents. Some software programs may need to be developed to integrate the data from these sources. Some of this data is already in MODS. Data on the collection mail, could be approximated by a short special study of Commission staff visiting the selected plant to interview the dock workers and possibly counting containers coming in during the primary arrival time of late afternoon and early evening. The count could be used to corroborating the estimates from the dockworkers, or making a suitable adjustment. It might also be possible to rely on informed estimates by knowledgeable staff at the facility.

Work hour data can come from Time and Attendance Data and from databases of the employee roster of a plant. Physical characteristics can come from the variety of databases used to track the existence and location of equipment. Much of the data may already be available in the Enterprise Warehouse databases. The Commission should be given access, on a limited bases, to the databases currently available to most Postal Service employees.

7.2   Delivery Models

Variation in mail arrival rates for the delivery function occurs by day of the week, month and year. City carriers are paid on an hourly basis, work 8-hour days 5 days a week. Most carriers are assigned to a specific route for long periods. One or more carriers in a facility rotate between routes to fill in for carriers taking a day off during the 6-day delivery week and may receive a higher wage than other carriers. Overtime occurs with supervisor approval. Data on the availability of the different types of carriers available to serve the routes in a delivery unit and their wage rates is needed for the models. As a starting point, the following data should be obtained for a sample of delivery units for a week in a light and heavy volume month.

Mail arrival rates

- Volume of mail by shape by day of the week
    - Deferrable and non-deferrable (not equivalent to preferential and non-preferential since delivery point sequence mixes preferential and non-preferential)

Labor
- Labor available by different categories allowable in NALC contract
- Clerks available for sorting assistance
- Wage rates
- Man hours used for each labor category by day

This data should be available through the DOIS and TACS systems. The same source the Service used to produce the data submitted in Docket N2010-1 and cited in the figures in the above section on delivery models could be utilized.

7.3   Transportation Models

Highway transportation is driven by service standards and cubic foot capacity of vehicles.

The following appear necessary to build the linear program models that we have described:

- Mail arrival rates at staging for highway transportation, by destination and in terms of cubic feet and total volumes by destination in terms of cubic feet
- Number used and capacity (cubic feet) for each available truck size
- Last dispatch of value by destination
- Plants that will remain after consolidation
- Which plants will be served by P&DC's and which will have direct connection
- Travel time matrix for the remaining plants
- Cost of air vs. cubic foot mile for highway transportation
- Volume matrix moving by air between plants

Potential sources for this data are:

Highway Contract Cost System (HCCS). A record from HCCS contains annual cost, annual miles traveled, number of trucks, cubic capacity of trucks, route length, highway cost account, etc.

Transportation Information Management Evaluation System (TIMES). The Transportation Information Management Evaluation System is a data collection system designed to track highway contractor performance as well as measure the utilization of the truck transportation network.

Surface Visibility. The Surface Visibility database describes the integration of multiple systems used to track mail transported within the postal network. Surface Mail container barcodes are scanned as containers arrive or are unloaded at postal facilities. Data is also integrated with the Facility Access and Shipment Tracking (FAST) system, which mailers use to provide advance notification of drop shipment mailings. The Surface Visibility scans verify acceptance and link containers to the mailer's electronic manifest.

Appendices

Appendix A: Analytic Models for Peak Load Costs

1. Introduction and Summary

Our objective in writing this report is to develop a methodology for measuring the "peak load costs" of various operations of the Postal Service.[11] There is an extensive economics literature on "peak load pricing," but not on "peak load costs."[12] Thus, the first order of business is to clearly delineate what is meant by peak load costs for purposes of this Report. This is best achieved with reference to the goals stated by the Commission in its RFP:

"These costs [of Postal Service operations] should be modeled in such a way that the effect on costs [of] volume variation, of relaxing preferential service standards, and of relaxing staffing and scheduling constraints, can be separately identified and estimated."

In response to continuing declines in mail volume, the Postal Service has proposed drastic changes in its operating procedures: e.g., 5-day delivery instead of 6; reduction of service standards for preferential mail, consolidation of mail processing operations. These changes involve major departures from the Postal Service's established operating procedures. Insights regarding the cost effects of these changes cannot be readily obtained directly from traditional USPS models and analyses. Therefore, our strategy is to develop theoretical models of the changed nature of Postal Service operations.

The basic approach is to develop an optimizing model capable of calculating values of "before" and "after" solutions and comparing the minimized costs of each. As indicated in

---

[11] See PRC RPF

[12] A Google Scholar search yielded 301 papers with *peak load pricing* in the title, but only *one* paper with *peak load costs* in the title: **Ofelia Betancor, Juan Luis Jiménez, and M. Pilar Socorro,** "The costs of peak-load demand: An application to Gran Canaria airport," *Journal of Air Transport Management*, Volume 16, Issue 2, March 2010, Pages 59-64. The measure of peak load costs obtained by the authors was calculated by comparing the total operating and congestion costs experienced by the airport and its users with their estimate of the costs that would be incurred under the counterfactual assumption that airport use was constant over the demand cycle. As we shall see, this is consistent with our characterization of one source of peak load costs, below.

the above quote, the term "Peak Load Costs," is best viewed as reflecting such a "before" and "after" comparison when changes are made in constraints related to short-run cyclical patterns in mail operations: e.g., changes in service standards, mail arrival profiles, and labor work rules.

The objective of our stylized models of postal operations is to formulate a nontrivial constrained optimization problem facing the operations manager (e.g., minimize operation costs subject to satisfying various service standards and work rules.) The next step is to solve the problem with and without various constraints. The reduction in the minimized cost is the cost of the constraint (or set of constraints) in question. We have adopted a Linear Programming approach to conduct this "before and after" analysis of peak load costs. This has the advantage of being implementable using readily available software (Microsoft Excel). The data requirements are likely to be less demanding than other modeling methodologies.

2. A General Statement of the Problem

It is quite straightforward to describe the operations of the Postal Service, or any other decision-making entity (DM) in quite general, abstract terms. The DM is called upon to choose the values of n decision variables, i.e., specify a vector of real numbers $(x_1, x_2, ..., x_n)$ = $\mathbf{x} \in \mathbb{R}^n$. Usually, the choices of the DM are *constrained* by technology, custom, politics and/or the laws of physics to be *feasible*. That is, they must be chosen such that $\mathbf{x} \in F \subseteq \mathbb{R}^n$. The *feasible set*, $F = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}$ is a feasible choice for the DM$\}$, describes the range of values of the decision variables that the DM is actually able to implement. The feasible set often depends on the value of various exogenous parameters, $(y_1, y_2, ..., y_k) = \mathbf{y} \in \mathbb{R}^k$, that affect the DM's operating environment: e.g., mail volumes, service standards, etc. This dependence is expressed by writing $F = F(\mathbf{y})$.

The choices of the DM have implications and may determine the values of many types of *outcomes*. For example, the Postal Service's decisions on how many hours of labor to employ and trucks to operate will, in part, determine the level of operating costs it incurs,

the number of electric bills that are paid, the amount of congestion on the nation's highways, and the "carbon footprint" of the Postal Service. In general, these *outcomes* will also depend upon the levels of other exogenous parameters, $(w_1, w_2, ..., w_s) = \mathbf{w} \in \mathbb{R}^s$; e.g., wage rates, fuel costs, etc.[13] That is, depending on the application, the analyst can specify various *outcome functions*, $O_j(\mathbf{x}, \mathbf{w})$, that map the DM's decisions to real numbers of particular interest.[14]

We are now in a position to characterize the general problem faced by the analyst. The environment facing the DM has is about to change, and the analyst is asked to determine what change can be expected to occur in the values of one or more outcomes of interest. In terms of our notation, the values of the exogenous parameters are changing from $(\mathbf{y}^0, \mathbf{w}^0)$ to $(\mathbf{y}^1, \mathbf{w}^1)$ and the analyst's task is to determine the value of $\Delta O_j = O_j(\mathbf{x}^1, \mathbf{w}^1) - O_j(\mathbf{x}^0, \mathbf{w}^0)$. Assume, for simplicity, that the analyst is fully informed about the past and future values of the exogenous parameters, $(\mathbf{y}^0, \mathbf{w}^0)$ and $(\mathbf{y}^1, \mathbf{w}^1)$.[15] There are clearly two challenges involved in this task. First, the analyst must, somehow, *identify* the outcome functions of interest. That is, he must specify, by assumption or empirical measurement, the exact nature of the relationships that are summarized by the outcome functions in question as well as the feasible set. Second, he must be able to accurately *predict* the values of the decision variables that the DM will select in the situation characterized by $(\mathbf{y}^1, \mathbf{w}^1)$.

2.1 Methodologies based upon the optimization hypothesis

In principle, the analyst could simply ask the DM to report his planned (feasible) response to the change in his operating environment. This would seem to obviate the need to predict $\mathbf{x}^1$. The analyst could then merely plug in the reported value into the outcome function of interest and directly calculate $\Delta O_j = O_j(\mathbf{x}^1, \mathbf{w}^1) - O_j(\mathbf{x}^0, \mathbf{w}^0)$. However, it is quite likely that the DM may be unable or unwilling to provide an accurate statement of his planned choice of $\mathbf{x}^1$. It simply may not be in his interest to do so.

---

[13] The vectors $\mathbf{y}$ and $\mathbf{w}$ may have elements in common. That is, both the feasible set and various outcome functions may depend upon the same exogenous parameters; e.g., the ambient temperature.

[14] That is, $O_j : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}$ for $j = 1, ..., J$.

[15] In the absence of this knowledge, another part of the task would be to predict, or forecast, the future values of the exogenous variables, $(\mathbf{y}^1, \mathbf{w}^1)$.

By far the most common approach to the prediction aspect of the task is to *assume* that the DM makes his choices with the goal of optimizing *one* of the outcome functions discussed above. For ease of notation, let the "first" of many possible outcome functions denote the firm's total expenditure on productive inputs: i.e., $O_1(\mathbf{x}^1,\mathbf{w}^1) = E(\mathbf{x}^1,\mathbf{w}^1)$. Then, the analysis proceeds under the hypothesis that the choices of decision variables that will pertain under parameter values $(\mathbf{y}^1,\mathbf{w}^1)$ are given by

$$\mathbf{x}^1 \in \mathbf{x}^*(\mathbf{y}^1, \mathbf{w}^1) = \operatorname{argmin}_{\mathbf{x}}\{E(\mathbf{x}, \mathbf{w}^1) : \mathbf{x} \in F(\mathbf{y}^1)\}$$

The optimization hypothesis can then be used to define a *value function* associated with the outcome function of interest. The value function associated with the DM's expense outcome function, $E(\mathbf{x},\mathbf{w}^1)$, and feasible set, $F(\mathbf{y}^1)$, is its minimum cost function:

$$C(\mathbf{w}^1, \mathbf{y}^1) = \min_{\mathbf{x}}\{E(\mathbf{x}^1, \mathbf{w}^1) : \mathbf{x}^1 \in F(\mathbf{y}^1)\}$$

At this point, the analyst's task is straightforward. Once the function C has been specified, all that is required is to evaluate it at two different points and calculate the difference: i.e.,

$$\Delta E = C(\mathbf{w}^1, \mathbf{y}^1) - C(\mathbf{w}^0, \mathbf{y}^0).$$

There exist a wide variety of analytical specifications and software implementations designed to solve such optimization problems.[16] The literature on applied optimization analysis is immense.[17] There are both sophisticated, general purpose software optimization packages (e.g., Mathematica, MATLAB) as well as custom built packages designed to solve particular problems. Given this wide universe of possible analytical and numerical approaches to the general constrained optimization problem, why have we chosen the tried and true Linear Programming approach to the problem?

The answer is quite simple. Our task is *not* to derive optimal decisions for the Postal Service. Rather, our goal is to develop an analytical tool to provide the Commission with insights into the likely cost savings that would emerge from a system wide change in Postal

---

[16] Obviously, if a technique can solve the optimization problem for original parameter values $(\mathbf{w}^0,\mathbf{y}^0)$, it can likely solve it for $(\mathbf{w}^1,\mathbf{y}^1)$ as well. The difference is exactly ΔE, completing the analysis.

[17] For a recent thorough introduction to the subject, see Wilhelm Forst and Dieter Hoffmann, *Optimization—Theory and Practice*, Springer 2010

Service operating procedures. *If*, for example, our goal were to design work schedules for mail processing employees at a particular center for a particular time period, we would seek to obtain (or have designed) the most directly applicable software optimization package.[18] We would then run that package for various relevant parameter values, thereby obtaining the desired "before and after" measure of the cost savings for that particular facility.

However, in order for an analytical tool be useful for evaluating system wide changes, it must be recognized at the outset that the analysis is, at best, an approximation of the real world operating environment. The flexibility and versatility of the Linear Programming methodology makes it especially useful for this task.

The next subsection describes the Linear Programming formulation we propose to use for a "before and after" analysis of Postal Service policy changes. The following subsection briefly outlines the characteristics of an econometric analysis of postal operations. We believe that such an analysis would complement ours.

2.1.1 Linear Programming

The methodological approach we adopt to perform the above calculation is Linear Programming. Linear Programming is designed to deal with a particular subclass of the general Mathematical Programming problem described above. It assumes that the objective function is a linear function of the decision variables, $\mathbf{w} \cdot \mathbf{x}$, and the feasible set is characterized by the set of linear inequalities given by the matrix equation $\mathbf{Ax} \leq \mathbf{b}$.[19]

2.1.2 Econometric estimation

An alternative, often complementary, approach is to specify an econometric model that allows the analyst to identify and estimate a behavioral cost function, *C(w,y)*. This function is called a behavioral cost function because it does not require the analyst to assume a specific objective

---

[18] We have asked Postal Service operations managers to identify such packages, but have not, as yet, received a response.

[19] In terms of the general model described above, the elements of the matrix **A** and vector **b** are made up of the elements of the parameter vector **y**.

function on the part of the decision-maker in choosing the elements of x.   The only assumption required is that the rules and procedures the decision-maker uses choose the elements of x, are the same over time.   Under this weak assumption, the behavioral cost function, $C(w,y)$, recovers the actual relationship between the decision-maker's actual cost of production and input prices, w, and the output vector, y.   Using this econometrically estimated behavioral cost function,, the analyst can compute *fitted* or *predicted* values of *C* at the parameters of interest and compute their difference, just as above.[20]

An econometric study requires data on costs, input prices ($\mathbf{w}_{it}$) and the vector ($\mathbf{y}_{it}$).   Estimating a cost function for daily mail processing is possible using MODS operations data because it is collected at the operating tour level each day at more than 300 mail processing facilities.   The relative homogeneity in mail processing facilities across the country makes it feasible to pool data MODS data across facilities and over time in order to estimate of behavioral cost function. Specifying a flexible functional form in y, will allow an assessment of the extent to which production cost increases at an increasing rate when daily mails volume are high, which would be indicative of the existence of peak-period costs.

The behavioral cost function model could also be estimated at different levels of time aggregation to assess the extent to which peak-period costs occur on a daily versus weekly or monthly basis.   For example, a daily analysis could look at the impact of dropping to 5-day delivery on the Monday mail processing peak.   Alternatively, the weekly or monthly analysis could look at the impact of 5-day delivery on the cost of processing the Christmas time mail peak.   Certain elements of x could be assumed to be fixed in the daily analysis, but this assumption could be relaxed in the weekly or monthly analysis to reflect the fact that more inputs are flexible over longer time horizons.   The daily or even tour-level analysis

---

[20] Empirical work on regulated and government-owned firms finds statistically and economically significant deviations from least-cost production, so that most econometrically estimated cost functions are behavioral cost functions that requiring only the existence of stable decision rules for their validity.

Other areas, besides mail processing. could be studied for the existence of increasing costs with mail volume increases, as long as there is available data on the costs of the activity, C, the input prices, w, and the vector of output and other service measures, y. For example, using a dataset on daily carrier route delivery times, an econometric model of the cost mail delivery could be estimated and a counterfactor analysis of the impact of 5-day delivery on delivery costs in the remaining days could be computed.   Another example is the cost of long-haul transportation. Using a dataset on USPS contracts for long-haul transportation an econometric model of the cost of long-haul mail transportation could be estimated to perform a counterfactual analysis of the changes in mail volumes across days of the week, weeks of the year, or months of the year on postal costs.

If the Postal Regulatory Commission is able to gain access to updated USPS datasets that were formerly used in Postal Rate Commission rate-setting proceedings, these dataset can provide a rich source of information to estimate econometric models of how postal costs are actually incurred and thereby provide a method for estimating the cost changes associated with significant changes in the level and distribution of postal mail volumes across days of the year

3. Peak Load Costs in Mail Processing

This section develops the model introduced by two decades ago by Crew, Kleindorfer and Smith (CKS)[21]. It highlights the peak load issues caused by the service standard requirements on preferential mail as well as their partial mitigation from the ability to defer non preferential mail. We analyze a simple two period, two input, two product version of their model. This enables us to develop simple algebraic solutions for the optimal policies, with and without constraints. Then, we derive simple formulas for the difference between the minimized costs with and without service standards, which allows us the measure. We also employ the model to derive a simple formula for calculating the costs savings that results from shifting a unit of preferential mail from the high cost period to the low cost period. The model presented here is quite simple, in order to highlight the underlying intuition. However, the most general version of the CKS model is also a Linear

---

[21] Michael A. Crew, Paul R. Kleindorfer and Marc A. Smith, "Peak-Load Pricing in Postal Services," *Economic Journal*, **100**, (September 1990), 793-807.

Program. Thus it is straightforward to add additional variables (classes of mail, time periods) and constraints (lengths of shifts, hours per week) for a more detailed quantitative analysis.

Following CKS, we assume that the day is divided into two time periods or Tours ($t = 1,2$) and there are two classes of mail, preferential mail and non preferential whose volumes in period t are given by P_t and N_t, respectively. (Let $P=P_1+P_2$ and $N=N_1+N_2$, respectively, denote the total volumes of preferential and non preferential mail and let $V=P+N$ denote the total volume of mail over the demand cycle.) We assume that most preferential mail arrives on Tour 1, so $P_1 > P_2$. (Let $\Delta = P_1 - P_2$ denote the size of the peak in preferential mail.) In view of recent trends in mail volumes, we also assume that $N > P$.[22]

Non preferential mail can be deferred one period without processing, but preferential mail must be processed during the period in which it arrives. (Let Y_t denote the volume of non preferential mail processed during period t.) It is assumed that labor can be adjusted over the day, but processing capacity cannot. (Let L_t denote the amount of labor employed in period t.) Each unit of mail processed requires on unit of labor and one unit of machine processing capacity (M). The price of a unit of machine capacity, which is available for use on both Tours, is denoted by m. The basic wage rate, paid to labor on Tour 2 is denoted by $w_1 = w$. Tour 1 labor is paid a wage premium of 100ρ per cent, so that $w_1 = (1+\rho)w$. Table 1 summarizes the notation used in our mail processing example.

---

[22] See, for example, U.S. Postal Service Office of Inspector General, Risk Analysis Research Center, Report Number: RARC-WP-11-008. As noted above, the model can be analyzed with many patterns of volume over service category and arrival times. In this 2×2×2 model, this realistic assumption serves to reduce the number of analytical cases that need be presented.

Table 1: Notation used in the Mail Processing Example P_t = volume of preferential mail in period t=1, 2.

$P = P_1 + P_2$ = total volume of preferential mail

$\Delta = P_1 - P_2 > 0$ measures the variability of preferential mail across periods

N_t = volume of non preferential mail in period $t = 1, 2$.

$N = N_1 + N_2$ total period volume of non preferential mail

$V = N + P$ = total volume of mail

Y_t = amount of non preferential mail processed in period $t = 1,2$.

M = machine processing capacity.

L_t = labor hired in period $t = 1, 2$.

$w = w_2$ = base (period 2) wage rate

ρ = shift wage premium

$(1+\rho)w = w_1$ = wage rate in period 1.

m = unit cost of machine processing capacity

Thus, in this example, the cost minimization problem facing the manager of the mail processing center is to:

$$\min_{L_1,L_2,M,Y_1,Y_2} \left\{ w(1+\rho)L_1 + wL_2 + mM \right\}$$

This minimization is subject to (1) Technological constraints that each unit of mail have available the required amount of labor and machine processing capacity; and (2) Quality of service (dispatch) constraints that requires preferential mail to be processed during the period it arrives and that all the mail be processed. That is,

$$M \geq P_t + Y_t; \qquad L_t \geq P_t + Y_t \qquad Y_1 + Y_2 \geq N \qquad\qquad t = 1,2$$

Clearly, this problem is a (very) simple Linear Programming problem. The objective function is a linear function of (three of) the (five) decision variables and the feasible set is determined by five linear inequality constraints.

This illustrative mail processing LP problem is "small" enough to derive simple, intuitive formulas for the optimal solutions in two parametric cases of interest. The logic of the

analysis is as follows. Because preferential mail must be processed during the period in which it arrives, it is certainly necessary that installed machine capacity must be at least as large as Tour 1 preferential mail volumes: i.e., $M \geq P_1 > P_2$. Then, the operations manager must decide how much of the non preferential mail to process in each period. Here, there is a tradeoff between efficient capacity utilization and incurring shift premium costs: *Either* additional machine capacity must be installed to process all non preferential volumes in period 2, *or* some of the non preferential mail must be processed using higher priced period 1 labor. The solution depends quite intuitively upon whether machine capacity is "cheap" or "expensive" relative to the size of the Tour 1 wage premium: i.e., whether $m < w\rho$ or $m > w\rho$.

Case (i): Capacity is *cheap*, so that it is optimal to install enough capacity so that *all* non preferential mail can be processed in period 2, thereby minimizing the amount of shift premium wages paid. The optimal solution requires $M = Y_2 = P_2 + N$, $Y_1 = P_1$ and the (minimized) costs of providing mail processing services are given by:

$$C_i(P_1, P_2, N; \rho, w, m) = w(1+\rho)P_1 + (w+m)(P_2 + N)$$

The intuition behind these results is as provided by the following thought experiment. *If* sufficient machine capacity were installed to serve Tour 2 preferential mail and *all* non preferential mail there would be enough capacity to serve Tour 1 preferential mail as well as long as total non preferential volumes are at least as large as preferential volumes.[23] Now, *imagine* shifting one unit of non preferential mail to Tour 1. Doing so would allow you to reduce machine capacity by one unit, saving an amount equal to m. However, you would then incur the wage premium on one additional unit processed during Tour 1, at an added cost equal to $w\rho$. When capacity is cheap ($m < w\rho$) this does not pay.

Case (ii): Capacity is expensive, so that it is optimal to schedule non preferential mail so as to fully utilize capacity in both periods. Then, the optimal solution requires $M = Y_1 = Y_2 =$

---

[23] That is, $M = P_2 + N$ implies $M > P_1$ whenever $N \geq P = P_1 + P_2$.

$(P_1 + P_2 + N)/2$, so that some non preferential mail is processed in period 1. The (minimized) costs of providing mail processing services are given by:

$$C_{ii}(P_1, P_2, N; \rho, w, m) = [w(2+\rho) + m](P_1 + P_2 + N)/2 = [w(2+\rho) + m]V/2$$

Again, the intuition behind the results can be obtained from a simple thought experiment. When machine capacity is expensive, it makes sense to conserve it by fully utilizing capacity on both Tours: $M = V/2$. Now, *imagine*, shifting one unit from Tour 1 to Tour 2 in order to avoid the Tour 1 wage premium. The resulting wage savings are $w\rho$. However, an additional unit of machine capacity must be installed to process this mail on Tour 2, increasing machine costs by the amount m. When capacity is expensive ($m > w\rho$), this does not pay.

3.1 What determines Peak Load Costs?

We have derived simple formulas for total costs that reflect various parameter values and constraints. Intuitively (as well as mathematically), there are three reasons to suspect that the resulting level of costs is higher than it might be in other circumstances.
Consider first the impact of the period 1 wage premium. Clearly, C_{1A}, C_{1B} and C_{1C} are all increasing functions of the wage premium, ρ. Thus it is tempting to consider identifying the total amount paid in shift premiums as "peak load costs". However, this is not appropriate. In a competitive labor market, the shift premium is merely a cost of doing business. Certainly, mail processing costs would be lower if period 1 wage rates were lower. However, this has nothing to do with the peaking structure of postal operations. Mail processing costs would also be lower if period 2 wage rates and/or equipment costs were lower. As we shall see, the amount of appropriately measured peak load costs may depend upon the value of ρ. But they will also depend upon the values of w and m. In terms of the present model, we are left with two "peak load" sources of added mail processing costs: service standards and the arrival behavior of preferential mail.

3.1.1 The cost of service standards.

All that is necessary to measure the cost of the service standard that requires preferential mail to be processed in the period in which it arrives is to compare the minimized costs

derived above to the costs that would be incurred if it were possible to "roll the mail:" i.e., process it whenever possible without regard to arrival time. Again, the result differs depending upon whether machine capacity is "cheap" or "expensive." When capacity is cheap and there are no service standards, it is optimal to install enough capacity so that all mail is process during period 2. On the other hand, when capacity is expensive, the optimal solution in the absence of service standards involves dividing total volume equally across the two periods. The "roll the mail" cost functions for the two cases are given by:

$$C_c(P_1, P_2, N; \rho, w, m) = \frac{(w+m)V}{2}$$

$$C_e(P_1, P_2, N; \rho, w, m) = \frac{[w(2+\rho)+m]V}{2}$$

It is now possible to derive formulae for the cost of service standards for the two cases:

(i) Cheap Capacity: In this case service standard costs are given by:

$$SS_{(i)} = C_{(i)} - C_c = w(1+\rho)P_1 + (w+m)(P_2+N) - (w+m)V \qquad = w\rho P_1 - mP_1$$

When capacity is cheap, "rolling the mail" results in an expansion of capacity, the costs of which partially offset the elimination of premium wages.

(ii) Expensive Capacity. The difference between minimized costs with and without service standards is

$$SS_{(ii)} = C_{(ii)} - C_e = w(1+\rho)P_1 + w(P_2+N) + mV/2 - (2w+w\rho+m)V/2$$

$$= wV + w\rho P_1 + mV/2 - (2w+w\rho+m)V/2 = 0$$

In this case, the high cost of capacity makes it optimal to process one half of total mail volumes in each period, both with and without service standard constraints. Shift premium costs of $w\rho(V/2)$ are incurred in either situation. Since the optimal mail processing configuration remains unchanged after the relaxation of the service standard, the appropriately measured service standard costs are zero.

3.1.2 Cost savings from smoothing the peak

Even with service standards in place, it may be possible to induce some preferential mailers to change their behavior so that their volumes arrive during Tour 2. If it is necessary to induce mailers to shift their deposits of mail by offering a discount, it is important to determine how much is saved by shifting a small amount (one unit) of preferential mail from Tour 1 to Tour 2

The cost savings from such smoothing can be analyzed more easily when costs are rewritten in terms of total volumes and peak size, Δ.

$C_{(i)} = w(1+\rho)(P+\Delta)/2 + (w+m)[(P-\Delta)/2+N]$;
$C_{(ii)} = w(1+\rho) (P+\Delta)/2 + w[(P-\Delta)/2+N)] + mV/2$;

When capacity is cheap, shifting 1/2 unit from Tour 1 to Tour 2 (so that Δ increases by one unit) necessitates an increase in capacity, which reduces the cost savings:

$dC_{(i)}/d\Delta = (w\rho-m)/2$

When capacity is expensive, the shift of 1/2 unit of preferential mail from Tour 1 to Tour 2 has no effect on optimal machine capacity so that:

$dC_{(ii)}/d\Delta = w\rho/2.$

3.2 Excess Capacity Considerations

The above analyses have been carried out from the economist's typical "timeless" or "long-run" perspective. That is, it is assumed that one is designing a facility "from the ground up," before any resources are committed to facilities, machines, or labor contracts. Thus, one is free to vary machine capacity from one scenario to another in order to achieve the least cost feasible configuration. However, under current circumstances, the proposals for service standard reductions and other cost saving initiatives are associated with significant declines in mail volumes. Thus, it is also necessary to consider the cost implications of

service standard changes in cases in which the postal operator has a large amount of available machine processing capacity that cannot be reduced over the planning horizon. Fortunately, this possibility is easily handled within the current framework. Formally, one assumes that there is an initial ("legacy") level of machine capacity, M^0, that has already been paid for and is freely available for use in both periods. Then the technological constraints on machine resources become $M^0 + M \geq P_t + Y_t$; $t = 1,2$. For a large enough level of legacy capacity (i.e., $M^0 > P_2+N$), the machine capacity constraints will not be binding in either period for the current volume levels. The solution to the problem then becomes even simpler, with the optimal values of the decision variables given by:

$$Y_1 = 0; \qquad Y_2 = N; \qquad L_1 = P_1; \qquad L_2 = P_2 + N; \qquad M = 0.$$

The minimized level of costs is then:

$$C^0 = w(1+\rho)P_1 + w(P_2 + N) + mM^0 = \frac{w\rho(P+\Delta)}{2} + wV + mM^0$$

3.2.1 Peak Load Costs under excess capacity

Because of the practical relevance of this situation, it is worthwhile to briefly summarize and restate the results for this case. The cost of service standards is just the difference between base costs, C^0, and the cheap capacity flow mail costs, C_c: i.e.,

$$SS^0 = C^0 - C_c = w(1+\rho)P_1 + w(P_2 + N) - wV = w\rho P_1$$

Thus, in the simple excess capacity situation, the peak load costs associated with service standards are exactly equal to the total amount of the shift premia paid to process preferential mail during Tour 1.

The analysis of peak shifting in the excess capacity case proceeds as before. Shifting one half unit of volume from Tour 1 to Tour 2 would decrease Δ by one unit. As a result the wage premium on ½ unit of volume would be saved. Equivalently, the cost savings from shifting the marginal unit of volume to Tour 2 is given by:

$$\frac{dC^0}{d\Delta} = \frac{w\rho}{2}$$

There is an important caveat to keep in mind, both in the context of these simple models and in practice. The available capacity volume, M^0, underlying the cost function C^0 was presumably determined by capacity installation decisions made in the past, when volumes were significantly larger: e.g., at $P_1^0 > P_1$; $P_2^0 > P_2$; and $N^0 > N$ (and, of course, $V^0 > V$). Suppose that those decisions were made optimally, i.e., to minimize costs to serve the initial volumes. Obviously, the amount of capacity installed depends upon which case pertains in the initial situation. However, the largest possible value of optimally installed initial capacity would be that corresponding to an initial "cheap capacity" situation: i.e, $M^0 = P_2^0 + N^0 \geq V^0/2$. But, the optimal level of capacity that one would need to install to avoid paying the Tour 2 wage premium under the "roll the mail" option would be V. Even if volumes decline significantly, it may still be the case that $P_2^0 + N^0 < V < V^0$.

4. A *very* simple model of peak load considerations in delivery

Our basic approach can be readily extended to analyze exogenous changes pertaining to the delivery function. Here, we present a very simple example driven by the variability of mail over the days of the week. Mail volumes over the two "days" of the week are given by V_1 and V_2, respectively. Again, total volume is denoted by $V = V_1 + V_2$, with the size of peak given by $\Delta = V_1 - V_2 > 0$. For purposes of this example, no distinction is made between preferential and non preferential mail. The operations manager must determine the number of routes, R, established to serve delivery area under study. The per period (*daily*) wage cost of a route is assumed to be w. The maximum amount of volume that can be delivered on a route each day is assumed to be r. For simplicity, there are assumed to no costs that vary with volume up to the capacity of a route, but the number of routes operated is constant across the week.

Given these assumptions, the operations manager's problem is to

$\min_R C = 2wR$ subject to: $R \geq V_1/r$ and $R \geq V_2/r$

Clearly, this is also a (very simple) Linear Programming problem. The objective function is a linear function of the (single) decision variable and there are two linear inequality

constraints. Obviously, the cost minimizing solution is for the operations manager to choose a number of routes that is just adequate to serve the high volume day, $R = V_1/r$, recognizing that there will be excess "delivery capacity" on the off peak day.   Then, the minimized level of delivery costs is given by:

$C = 2wV_1/r = w(V+\Delta)/r$

4.1 Peak Load Costs in the delivery model

There are two obvious sources of peak load costs in this simple delivery model. The first is the *work rule* constraint that the number of routes must be constant over the "week." The second source of added costs is the variability of mail across days. As in the mail processing example, the procedure to analyze these costs is derive the solutions to counterfactual "relaxed" and "smoothed" versions of the delivery model and subtract the resulting level of minimized costs from the level of minimized costs in the base case. First consider the case in which work rules have been relaxed so that the operations manager no longer has to operate the same number of routes each day of the week. Now, the decision problem is to

$\min_{R1,R2} C_x = w(R_1+R_2)$ subject to: $R_1 \geq V_1/r$ and $R_2 \geq V_2/r$

The obvious solution to this simple Linear Programming problem is for the operations manager to choose exactly the required number of routes for each day. Thus, the optimal numbers of routes are given by $R_1 = V_1/r$ and $R_2 = V_2/r$ and the level of minimized costs is given by:

$C_x = w(V_1+V_2)/r = wV/r$

Then, the peak load cost savings from relaxing delivery work rules is

$WR_d = C - C_x = 2wV_1/r - w(V_1+V_2)/r = w(V_1-V_2)/r = w\Delta/r$

Next, let us examine the cost savings that would result if mailers could be induced to shift ½ of a unit of mail from day 1 to day 2. Again, this would decrease Δ by one unit. Differentiating the minimized cost function with respect to Δ yields

$d\text{C}/d\Delta = w/r$

The cost savings from shifting a unit of volume from peak to off peak is efficient cost per unit.

5. Peak Load Costs in Ground Transportation: Relaxing Service Standards

The example we analyze involves two mail processing centers (A and B) that transport mail volumes to each other over a two period cycle. Because of service standards, each center must ship available volumes to the other in each period. In period 1, center A ships volume A_1 to center B and center B ships volume B_1 to center A. The trucks involved make a roundtrip loop, so the same total capacity, T_1, is always available on trip AB as on trip BA. Similarly, in period 2, center A ships volume A_1 to center B and center B ships volume B_1 to center A using T_2 units of trucking capacity. Let t denote the cost of employing a unit of roundtrip trucking capacity. While many combinations of relative volumes are possible, we analyze the case in which $A_1 > B_1$ and $A_2 < B_2$.

The transportation manager's scheduling problem in this example involves choosing the level of trucking capacity to dispatch in each period. That is,

$$\min_{T_1, T_2} t(T_1 + T_2) \text{ subject to } T_1 \geq A_1; \qquad T_1 \geq B_1; \qquad T_2 \geq A_2; \qquad T_2 \geq B_2$$

Again, this is a simple Linear Programming problem. The objective function and constraints are linear functions of the decision variables. Many combinations of relative volume levels are possible. However, we analyze the case in which $A_1 > B_1$ and $A_2 < B_2$. The optimal solution is, obviously, to choose a level of trucking capacity equal to the maximum directional demand in each period: i.e., $T_1 = A_1$ and $T_2 = B_2$. Then, the minimized level transportation costs is given by $C^0 = t(A_1 + B_2)$.

Now consider the new problem facing the transportation manager after the service standard has been relaxed, so that he no longer has to transport volumes during the period in which they become available. His problem is then to

$\min_T tT$ subject to $T \geq A_1 + A_2;$ $T \geq B_1 + B_2$

The obvious solution to this Linear Programming problem is to choose a level of trucking capacity sufficient to carry the larger of the aggregate directional volumes. That is, $T = \max\{A_1+A_2, B_1+B_2\}$ and $C_x = t \max\{A_1+A_2, B_1+B_2\}$. Then the cost savings from relaxing the service standards are given by:

$C^0 - C_x = t(A_1 + B_2) - t \max\{A_1+A_2, B_1+B_2\} = t \min\{A_1-B_1, B_2-A_2\} > 0.$

6. Towards empirical analysis.

As demonstrated in the Technical Appendix, the approach we have outlined can be readily implemented using Microsoft Excel. However, as yet we have produced no "real numbers:" i.e., calculations that purport to approximate actual Postal Service operations. To implement our approach quantitatively, we would need to "know" as much as possible about the parameters of the model; e.g., wage rates, labor rules, mail volumes, the time pattern of mail flows to processing plants and delivery units, interplant transport data, etc. The disaggregated the data the more useful it would be. For example, the analysis would be more accurate using a mail processing plant's *hourly* mail arrival pattern rather than its daily totals. Similarly, the delivery process is better analyzed using data disaggregated down to the level of daily volume on each route rather than aggregate volume data on a delivery unit over a month.

Not surprisingly, this is precisely the kind of data that an econometrician would seek into order to perform an empirical study. However, the information requirements of our Linear Programming approach are much more modest than those of a serious econometric study. Specifying and estimating an econometric model would require an extensive panel data

set[24] to be deemed successful. In contrast our approach requires only a point estimate or "best guess" for each parameter of interest. If a statistical estimate with a small standard error[25] were available, we would certainly make use of it. However, in the absence of tightly estimated statistical values, a single value from a "typical" plant or delivery unit would allow us to implement the analysis.

6.1 Examples of data required for an empirical analysis

This subsection provides a brief discussion of the types of data required for a thorough empirical/quantitative analysis and likely Postal Service sources. Additional details are provided in section 7 of the main report.

6.1.1 Mail Processing

As was illustrated in the mail processing example, mail arrival patterns by day and hour are among the most important drivers of peak load costs in mail processing. Thus it would be useful to have data on First Handled Pieces by plant. This should be available from MODS. Other useful disaggregate information on the time pattern of mail arrivals by plant should be obtainable from Permit Mail Statements and IMB data. Disaggregate data on mail processing labor hours worked should be available from the TACS data set.

6.1.2 Delivery

Weekly and seasonal variability seems to be the primary driver of peak load costs for delivery units. Mail volumes by day and route should be available from the DOIS data set. Again, labor hours worked by route should be available from TACS.

6.1.3 Highway Transportation

In the highway transportation example discussed above, peak load costs were ultimately driven by mismatches in directional mail flows between mail processing units.[26] Contract data from the HCCS data set would be required to determine the Postal Service cost of

---

[24] That is, observations on, say, Tour 1 labor hours for each of many mail processing centers, over a time period of many months: i.e., a "time series" of "cross-sections."

[25] For example, a sample average taken over a large number of plants.

[26] In transportation terminology, relaxing service standards may mitigate the *backhaul* problem.

hiring truck capacity. Capacity utilization levels by route should aid in quantifying the magnitude of the problem. This may be available from TRACS data.

## Appendix B: Linear Programming Models of Peak Load Costs

1. Introduction

Peak load costs for the Postal Service arise from the combination of uneven mail arrival rates, service standards, and labor/capacity constraints. Mail arrival rates vary by hour of day, day of week, day of month, and season of year. Service standards vary by type of mail, but can be approximately characterized as preferential and non-preferential. Both preferential and non-preferential mail can be usefully further subdivided into two types depending on the origin and destination of the mail. Originating mail (also called collection and acceptance mail) is mail being handled in its first plant and may be destined for either local or non-local plant areas. Incoming mail originates in a different plant's service area and destinates in the given plant's service area.

Most originating mail arrives after 5 PM each day. The preferential mail portion has service standards that dictate that it must be processed by a final dispatch time (also called cut-off time or dispatch of value) between 10 PM and midnight. As a result, there is an hourly peak during Tour 3. Non-preferential originating mail generally has the same final dispatch time, but it can be delayed until the next day if necessary. Incoming mail arrival patterns are similar to those of originating mail but overnight committed mail usually arrives after midnight and it has a final dispatch time in the early morning, so much of it is processed during Tour 1 which has a peak period prior to dispatch to delivery offices at about 5 AM.

Figure 1 below is an illustration of the patterns of cumulative arrival rates for originating (collection and acceptance), as well as typical final dispatch times for these two mail types. In this example, preferential and non-preferential mail arrival rates are combined. It can be seen that originating mail has the narrowest window for processing.

Figure 1. Example of Cumulative Hourly Arrival Rates with Final Dispatch Times for Originating and Incoming Mail
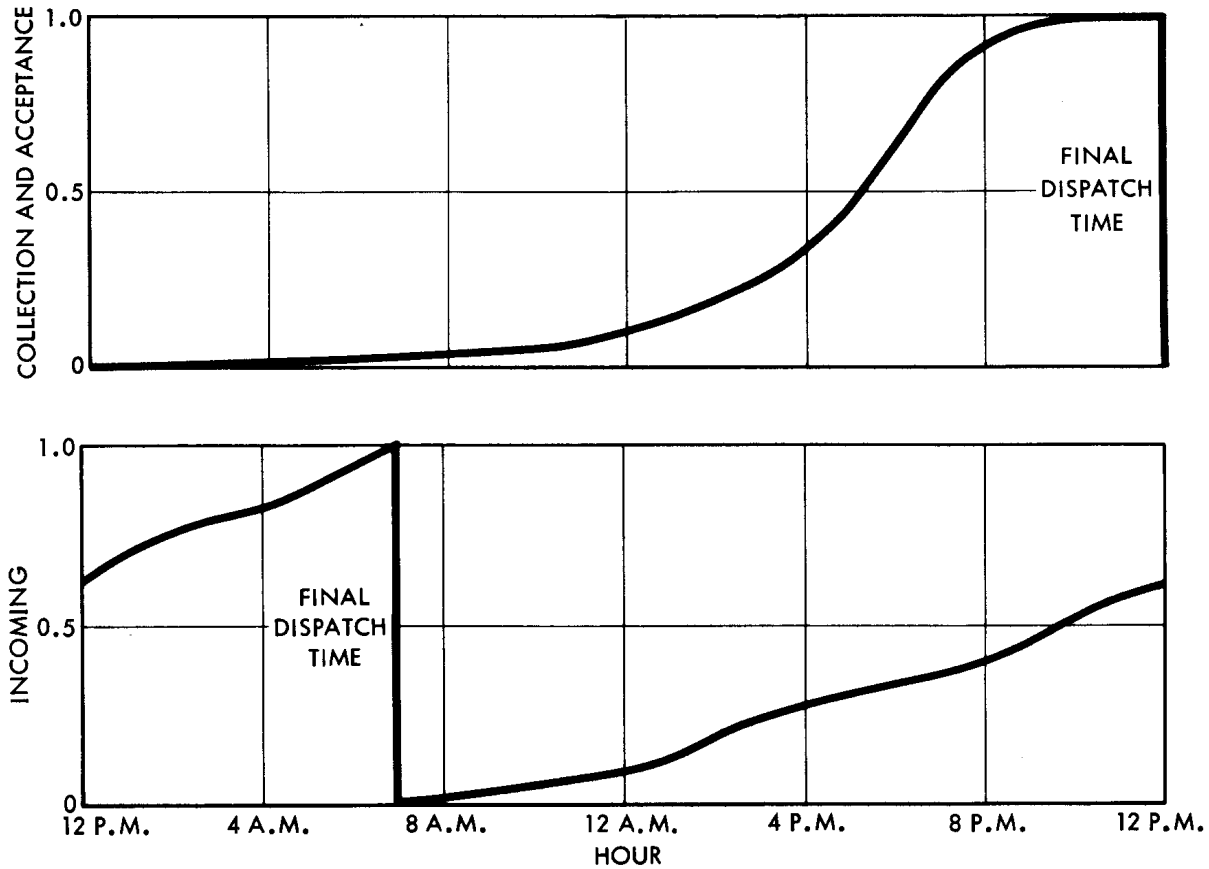
Delivery presents peak load issues similar to mail processing except that there are no hourly peaks. Work rules for the two functions are different also. Rural and city carriers are compensated differently and so present different issues. The former are paid by the route while the latter are paid by the hour. Establishing routes for rural carriers is done at a time of year according to the union agreement that represents neither a peak nor slack period. Thus the carrier must work longer hours during peak periods and fewer during slack periods. From the Postal Service standpoint the cost of peak and slack periods (especially seasonal ones) is minimized provided that the period when routes are established is one of average volume.

City carriers, however receive overtime payments especially during peak periods. These

routes may be divided into sections which can then be handed off to part time, temporary, or regular carriers and, of course, overtime can be used to meet peaks. In addition, routers can be used to perform the in office duties of city carriers. Thus, the Service has several means to minimize the cost of day of the week, day of the month and seasonal peak and slack periods.

As an example of delivery peak loads, Table 1 below shows typical variations in delivery workload from Monday to Saturday, as presented in the Commission's Advisory Opinion in the recent 5-day delivery case (Docket No. N2010-1). Since there is an opportunity to delay delivery of non-preferential mail by at least one day, the arrival rates at a carrier station are likely to be more variable than the daily delivered volumes.

Table 1. Delivery Volume Variations by Day of Week[27]
(Percentage Difference between Daily Volume and Average Daily Volume)

Day of Week	Percentage Difference from Average
Monday	+15%
Tuesday	+2%
Wednesday	+1%
Thursday	-2%
Friday	+8%
Saturday	-8%

There are also significant differences in mail processing and delivery arrival rates over longer time periods than a week. For example, volume is higher near the first of the month. Also, volumes are significantly higher in the Christmas period.

In spite of the large variations in hourly, daily, and longer period mail arrival rates, it would theoretically be possible to process and deliver mail efficiently if labor resources were only paid when needed, as with electric power consumption. However, this is not the case since there is a very complex set of rules that the Postal Service must follow when scheduling and using its labor resources. These include requirements that most employees be full-time regulars, with work schedules that require primarily 8-hour shifts (plus an hour for

---

[27] Source: PRC Opinion and Recommended Decision, Docket N2010-1, p. 120.

lunch and breaks) and consecutive five-day workweeks. There is some flexibility available through the limited use of part-time and casual employees, and through limited use of overtime for full-time employees, although the overtime rates (which do not include benefits) are 10-15 percent higher than the regular hourly rates (including benefits).

These labor constraints, together with the uneven mail arrival rates and service standards result in inefficiencies in using resources. If we consider the completely flexible (unconstrained) labor scenario as the minimum cost option, then we define peak load costs as the difference between the costs with all constraints included and the unconstrained case. Another useful measure would be the percentage of total costs that can be considered peak load costs.

Peak loads in transportation are also caused by the daily, weekly, monthly and seasonal variations in volume, but they involve trucking and air capacity issues rather than labor issues. Most of the Postal Service' air transportation is in the hands of FedEx who has a contract to transport most of the volume requiring air transportation. An important issue would be to estimate the additional cost reflected in the FedEx contract caused by the seasonal variations in mail volume.

The Postal Service is responsible for securing its own surface transportation. It does this by purchasing highway capacity over the many routes that it transports mail on a multi-year basis with the capacity sized to meet the peaks. This leaves overcapacity much of the time. It may be possible to buy reduced capacity on many routes and fill in from the spot market when required. Thus, a cost minimization analysis could shed light on the cost of peak loads in highway transportation. Rail transportation does not present peak load issues in that it can usually be purchased on the spot market over all routes.

2. Discussion of Proposed Methodologies for Analyzing Postal Peak Load Costs

2.1. Introduction

The RFP for this project specifies that the Contractor shall prepare a proposed methodology, or methodologies, for analyzing peak load costs under the constraints applicable to the Postal Service such as labor union rules and product specific service standards. The methodologies should be able to answer the following questions:

1. How might delivery frequency changes impact peak loads and what are the associated costs?
2. What savings will be gained from new flexibilities in the recent labor contract and when will they be achieved?
3. Can operational inefficiency be quantified?
4. How do unit costs change with seasonal fluctuations in volume?

Subsequent to the publication of the RFP for this project, the Postal Service announced plans for a major overhaul of its entire operating system as a means for cutting costs in its current declining-volume environment. The proposed changes involve extending the current First-Class delivery standards by one day, as well as consolidation of the mail processing network from the current 500+ plants to about 250. As a result, we also considered methodologies that would be useful in evaluating these major changes. We include this subject by adding a fifth question to the above list:

5. What additional savings will be gained from relaxing service standards for First-Class mail and from consolidating the processing network?

Question 3 will be addressed by comparing our estimated minimum costs that reflect all constraints (hourly, daily, and seasonal) with the actual costs for the same time period. Question 1 will be addressed by finding minimum costs reflecting only hourly and day-of-week constraints minus the minimum unconstrained costs. Question 4 will be addressed

by estimating minimum costs that reflect all constraints, then subtracting the minimum costs from Question 1. Question 2 will be answered by first changing all constraints to reflect the new labor agreements, then finding adjusted minimum costs reflecting all constraints, then subtracting the minimum unconstrained cost. Question 5 will be addressed by first changing the service standards and the number of processing facilities, then finding adjusted minimum costs reflecting all changes, then subtracting the minimum unconstrained costs.

We considered several peak load costing methodologies for addressing the above issues. The first was to use existing (commercial) mail processing, delivery and transportation models used by the Postal Service to assist field managers with developing labor and/or equipment schedules for mail processing, delivery, and transportation. Our discussions with knowledgeable USPS managers revealed that this approach would not be viable because these commercial models are considered confidential. Second, we investigated the development of an econometric approach to the peak load cost problem, which is described at pp ?? below. Finally, we developed a linear-programming approach for estimating postal peak load costs. Linear programming is a widely used tool for solving cost minimization (or maximization) problems with many embedded constraints.

In Section 2.2, we present an overview of the linear programming (LP) approach, along with a simple example of the use of LP in the mail processing, delivery, and transportation functions. In Part C, we discuss the mail processing, delivery, and transportation constraints in the current environment in some detail, as well as how these constraints would change in two scenarios: a) the current environment with the new labor flexibilities provided in recently-adopted union contracts, and b) the new environment with relaxed service standards and fewer facilities.

2.2. Linear Programming Approach

2.2.1 Overview of Linear Programming

Linear programming is one of the most common optimization techniques used in the field of Operations Research. It is used to find the minimum (or maximum) value of an *objective function*, which consists of a sum of terms involving a constant times a *decision variable*. One or more *constraints* are included in the specification of a problem which impose limitations on the values of the decision variables. The linear programming solution algorithm finds the values of all decision variables that minimize (or maximize) the value of the objective function while satisfying all constraints.

Although many linear programming software packages are readily available, it is also available as a tool in Microsoft Excel (Solver application). We feel that Excel would be the most efficient alternative because of its widespread availability, but it may not be feasible if the number of variables and constraints for the peak load cost problem is too large. To illustrate the Excel-based linear programming approach, we developed several detailed numerical examples of cost minimization problems in mail processing, using typical values for the required input parameters (including arrival rates, final dispatch times, labor constraints, and use of premium pay labor). These examples are described in Appendix A.

2.2.2 Mail Processing Example

The following simple example will be used to illustrate the linear programming approach for the mail processing function.[28] Assume we want to find the unconstrained minimum total one-day cost of processing two types of preferential mail -- originating and incoming -- at a single plant. Also assume that there is only one type of employee and that no labor constraints or premium costs are reflected. We will also assume that mail processing costs are 100% variable with respect to volume, so unit processing costs will not vary with

---

[28] More complex mail processing examples that incorporate labor constraints and premium pay are presented in Appendix C.

volume. Finally, we assume there is adequate existing sorting machine capacity to handle the input volume. So the objective function for this example can be expressed as:

Minimize $\sum_{i,j} (h(i, j) * r(i) * c$

where $h(i, j)$ = number of employee hours (in all operations) assigned for mail type i

during hour j (the decision variables chosen by the LP algorithm),

$r(i)$ = flow path productivity for mail type i (first handling pieces per employee hour), and

c = cost per employee hour.

First, we need a set of non-negativity constraints for the decision variables:

$h(i, j) >= 0$ for all i and j.

Next, we need a set of constraints that restricts the total processed volume for type i during hours 1 through k to be no more than the cumulative arrivals of type i mail at each hour k (see Figure 1):

$$\sum_{j=1}^{k} ((h(i,j) * r(i)) <= v(i, k) \text{ for all i and k}$$

where $v(i, k)$ = cumulative hourly arrivals for mail type i at hour k.

A constraint is also needed to ensure that all mail is processed during the 24-hour day:

$$\sum_{j} ((h(i, j) * r(i)) = v(i, di) \text{ for all i}$$

where   di   = final dispatch time for mail type i (see Figure 1), and

$v(i, di)$ = total daily arrival volume for mail type i.

2.2.3 Delivery Example

The linear programming model for minimizing delivery cost is similar to the mail processing model described above, except that hourly arrival variations are not an issue. The following example illustrates a simple delivery model for calculating minimum total cost for two days at a carrier station[29]. Assume that there is only one type of mail, and there are two types of employees, full-time (type 1) and part-time (type 2). Both types of employees must work eight-hour days. Full-time employees have a higher wage rate than part-time employees. Full-time employees must work on both days. Part-time employees can work either no days or one day. All routes have the same volume on a given day. Since the delivery function is only about 50% variable with respect to volume, the daily productivity for each route will depend on the daily volume for the route. The objective function for this example would be:

$$\text{Minimize } \sum_{i,j} (d(i, j) * c(i, v(j)))$$

where $d(i, j)$ = number of carriers assigned for employee type i on day j

(the decision variables chosen by the LP algorithm),

$c(i, v(j))$ = cost per type i employee day on day j with volume $v(j)$

= $cf(i) + cv(i) * v(j)$,

$v(j)$ = daily volume per route

$cf(i)$ = fixed cost per type i employee day, and

$cv(i)$ = variable cost per piece for a type i employee

Again we need a set of non-negativity constraints for the decision variables:

$d(i, j) >= 0$ for all i and j.

---

[29] This is a simple approximation of a weekly cost minimization model.

We also need constraints requiring that the number of full-time employees ($i = 1$) be the same on days 1 and 2:

$$d(1, 1) = d(1, 2)$$

A constraint is also needed to ensure that all mail is delivered on each day:

$$\sum_i ((d(i, j) * r(i, v(j))) = v(j) \text{ for all } j,$$

where $r(i, v(j))$ = productivity for employee type i on day j (delivered pieces/day).

2.2.4 Transportation Example

The following example presents a simple linear programming model for calculating minimum total daily intra-SCF transportation cost under relaxed service standards. Assume that the processing center (plant) in the SCF must deliver sorted mail to all delivery units in the SCF area. To meet the current preferential mail service standards, there are typically three daily truck runs from the plant to each delivery unit. The first run occurs around 2 am, and contains sorted non-automated letters, flats, parcels and carrier-route bundles for the delivery unit that must be merged and sequenced manually by carriers at the delivery unit during the first or second hour of their tour before they deliver the mail. The second run is made around dawn and takes the delivery point sequenced mail to the delivery unit. There may also be a third run just before the carriers go on the street that contains any left over or late-arriving items, such as parcels and Priority mail.

Each of the three possible truck runs typically has excess capacity, which could be better utilized if the preferential mail service standards were extended by one day, as planned in the new USPS operating environment. In this case, there would be an opportunity for intra-SCF transportation cost savings by reducing the number of daily truck runs from the plant to the delivery unit as much as possible (conceivably, one trip could be adequate).

For this example, assume that each truck has the same capacity, and that the maximum number of truck runs to a delivery unit is three. In this situation, the objective function would be:

Minimize $\sum_{i,j} (f(i, j) * c(i))$

where $f(i, j)$ =   cubic feet of mail contained in the jth trip from the plant to delivery unit i (the decision variables chosen by the LP algorithm), and

         $c(i)$ =   cost of a truck run from the plant to delivery unit i

With this objective function, the LP algorithm will minimize the intra-SCF transportation cost by putting as many cubic feet as possible in each truck run, so that fewer total runs will be needed (i.e., $f(i, j)$ will be zero for $j = 2$ and 3 to the extent possible).

Again we need a set of non-negativity constraints for the decision variables:

$$f(i, j) >= 0 \text{ for all } i \text{ and } j.$$

Constraints are needed to ensure that the cubic feet on each truck run do not exceed the truck's hauling capacity:

$$f(i, j) <= C \text{ for all } i \text{ and } j,$$

where C = cubic feet capacity of a truck.

Finally, constraints are needed to ensure that all mail is sent to each delivery unit on each day:

$$\sum_j f(i, j) = F(i) \quad \text{for all } i,$$

where $F(i)$ = total cubic feet for delivery unit i.

3. The Effect of New Labor Rules, Service Standards, and Network Size on Peak Load Costs

In this section, we discuss the impact on mail processing, delivery, and transportation peak load costs of:

- More flexible clerk scheduling rules contained in the new APWU contract
- Proposed relaxation of First-Class mail service standards
- Proposed consolidation of the mail processing network

3.1 Mail Processing

Current Environment

A review of the labor staffing constraints described in the former (2006-2010) APWU National Agreement, together with discussions with field managers and limited sensitivity analysis resulted in the following preliminary conclusions:

- Currently, a major source of mail processing peak load costs is the requirement that most workers be full-time regulars with two consecutive days off.

- Current preferential service standards require that much of the mail be processed between 6 PM and 6 AM, and so the contractually-required night differential (about 6%) adds significant costs.

- Regular overtime is an efficient way to provide capacity in peak-volume periods, since the overtime wage rate is only a few percent higher than the fully-loaded straight-time wage rate.[30]

- Penalty overtime is not an efficient way to provide capacity in peak-volume periods, but does not have to be paid during the month of December. Thus, the large holiday peak at the end of the year can be handled more efficiently by using extra overtime hours.[31]

- Part-time employees are a very efficient way to handle peak volume periods because their average wage is somewhat lower than full-time employees and because they do not have to be paid for eight hours a day, five days a week. However, there are contractual limits on their use which prevent the Postal Service from fully utilizing part-time labor in peak volume periods.

As described below, it appears that the new APWU labor constraints and the relaxation of First-Class service standards are the major changes as far as the mail processing function.

---

[30] For normal work hours (eight per day), regular employees receive straight-time pay plus all benefits, including the employer's contribution to Medicare and Social Security. Up to two additional work hours per day can be scheduled and paid at the overtime rate, which is 1.5 times the straight time rate, but only Medicare and Social Security are included as benefits. As a result, overtime hours only cost the Postal Service slightly more than normal hours.

[31] Penalty overtime must be paid for work hours in excess of 10 per day. The penalty overtime rate is 2.0 times the straight time rate, and also includes Medicare and Social Security. The penalty overtime wage is thus considerably higher than the normal wage rate.

These two changes should allow considerable cost reductions compared with the current environment. It is not clear that consolidation of mail processing facilities by itself would have a significant effect on mail processing peak load costs.

New APWU Labor Constraints

The new APWU National Agreement includes a Memorandum of Understanding which creates a new work force – Non-Traditional Full Time (NTFT). These employees can have much more flexible work schedules than traditional full-time employees. Their scheduled work week can vary from 30-48 hours; their daily work can vary between 4-12 hours; and their number of scheduled days per week can be more or less than five. This could allow, for example, schedules with four twelve-hour days per week, or six eight-hour days per week, or six five-hour days per week. The additional scheduling flexibility and resulting cost reduction opportunities provided by NTFTs can be estimated using an LP approach similar to that shown in the Mail Processing Example described above.

Relaxation of First-Class Service Standards

The proposed relaxation of current First-Class service standards should allow most mail processing work to take place from 6 AM to 6 PM, since late-arriving mail can be deferred to the next day. This change would avoid most night differential pay, so mail processing costs would be reduced considerably. These savings can also be estimated using a modified LP approach during the next stage of this project.

As an aside, the financial effect of any change in service standards can be estimated through the LP approach, provided that the analyst can predict the final dispatch times required to meet the new standards.

3.2 Delivery

Current Environment

A review of the labor staffing constraints described in the (2006-2010) NALC National Agreement, together with discussions with field managers showed that the contractual rules for delivery staffing are very similar to those for the former APWU National Agreement. The conclusions for the delivery function are thus the same as listed above for mail processing.

As far as potential opportunities for reducing delivery peak load costs, it appears that only the relaxation of First-Class service standards would have a significant effect. At the present time, there is no new NALC contract, so it is unclear whether any new flexibility in scheduling carriers will be forthcoming. As with mail processing, it is not clear that consolidation of the postal network would have a significant effect on delivery peak load costs.

Relaxation of Service Standards

It is expected that the extension of the First-Class service standards should result in a more even flow of mail to delivery units during the week, since managers would have the opportunity to level peaks by deferring delivery of some mail. The resulting savings can be estimated during the next stage of the peak load cost project using the delivery LP approach described above.

3.3 Transportation

Current Environment

A review of the current utilization of USPS-owned and contract transportation has shown there is considerable excess capacity in both local and non-local surface transportation. In some cases, excess capacity may be unavoidable, but it is clear that in many cases preferential service standards are driving the need for more frequent truck runs with less efficient capacity utilization.

It appears that both the relaxation of First-Class service standards and consolidation of mail processing facilities would allow significant reductions of transportation peak load costs. It is not clear that any new APWU or NALC labor constraints would have a significant effect on transportation peak load costs.

Relaxation of Service Standards

As described in the transportation LP example above, First-Class service standards drive the need for more frequent truck runs than would otherwise be necessary. More frequent runs mean lower utilization of truck capacity, both for local and non-local surface transportation. Relaxation of First-Class service standards would result in fewer truck runs and higher average capacity utilization. The resulting transportation peak load cost savings could be estimated an LP approach similar to the one described in the transportation example above.

Consolidation of the Mail Processing Network

Currently, there are 500+ major mail processing facilities nationwide. In principle, each of these network nodes must be connected to every other node by one or more transportation links. Each link requires one or more dedicated trucks, and each truck contract has a fixed cost component. Reducing the number of nodes (plants) to about 250 would result in many fewer transportation links, and thus fewer truck contracts, and thus less fixed costs. These transportation peak load cost savings could also be estimated an LP approach similar to the one described in the transportation example above.

Appendix C: Linear Programming Numerical Examples

This Appendix describes several numerical mail processing cost minimization examples which use the linear programming (LP) approach. All examples are contained in a Excel workbook, LPExamples.xlsx, submitted with this report. The hourly and daily volume profiles are presented on the Arrivals worksheet. The Patterns worksheet contains input data on the different employee work schedules available to process mail at each hour of the day. Each numerical LP example is then presented in a separate worksheet. The LP optimal solution for each example was found using the linear programming software in the Excel Solver module. It should be noted that all input data on arrival rates and wages are approximate.

1. Patterns Sheet
 The Patterns sheet shows the key element to this LP scheduling approach. Each work pattern has its own column, with zero or one values. A one indicates an employee is scheduled to work during the hour specified by the row. Each of the 24 patterns on this worksheet identifies a consecutive 8-hour shift with a different starting time. Each pattern can also have a different cost per hour. Other patterns can be added as necessary to represent other work schedules and associated costs, e.g. a pattern with just one hour representing overtime, or a pattern with 3 consecutive hours representing a part-time worker. This is what gives this approach the ability to handle a wide variety of employee schedules and wages.

2. Model1a Sheet
 The Model1a sheet shows the optimal solution for scheduling one day of originating mail processing with the hourly arrival rates shown on the Arrivals sheet and the 24 employee work schedules shown on the Pattern sheet. No constraints or premium rates are considered other than those imposed by the midnight dispatch of value and the consecutive 8-hour shift requirement for each employee. For this example, the following parameters were used: a total daily volume of 10 million pieces, an hourly wage of $40.00,

and a flow path productivity of 200 pieces per hour. The flow path productivity is the weighted (by operation unit cost) average sum of handlings in all operations divided by total first handling pieces. This results in a flow path cost per piece of $0.20, so the minimum cost with no constraints is $2.0 million. Column C shows the cumulative arrivals by hour assuming 10 million total pieces and the hourly arrival profile shown on the Arrivals sheet. Row 8 shows the cost per employee for each of the 24 patterns (8 hours times $40 per hour). The rectangular area in cells E11:AB34 shows the product of the hourly productivity and the 0-1 pattern matrix from the Pattern sheet.

The Excel Solver application was used to find the number of optimal (minimum cost) number of employees required to start their 8-hour shifts at each possible hour (see cells E6:AB6). It can be seen that the solution required 600 employees to start their shifts at 6 am (variable x7), 400 employees to start at noon (variable x13), and so on. The resulting pieces processed by hour are shown in Column B and the cumulative pieces processed are shown in Column D. The total volume processed during the 6 pm - 6 am night premium period is shown at cell B40 (5.38 million pieces). The total cost with this solution was $2.0 million, which matches the minimum unconstrained cost. If an arrival profile with a narrower peak arrival area was used, some overtime hours would be required (an example using overtime will be discussed in the Model2a section below).

3. Model1b Sheet

 This example is exactly the same as Model1a, except that the total volume was set at 13.3 million pieces to simulate a typical Monday volume. The optimal solution is the same as Model1a with the required employee numbers scaled up by the ratio of the total volumes (1.33). Again the optimal solution for this model matches the minimum cost solution. This would happen for any total volume, as long as the arrival profile and final dispatch time stayed the same.

4. Model1c Sheet

 This example is the same as Model1a (10 million pieces daily volume) except that a 10% night premium for work between 6 pm and 6 am is applied to each employee pattern. An

hourly night shift wage of $42.50 was used for this example. The costs for each employee pattern in Row 8 are calculated as a weighted sum of the scheduled night premium hours and day hours. For example, Pattern 1 includes 6 hours of night work and 2 hours of day work, so the weighted sum is (6 x 42.50 + 2 x 40), or $335.00.

The minimum cost for this case of $2.066 million is shown at cell A6. This is about 2.3% higher than the minimum cost solution. Note that the LP algorithm found a solution with less volume processed during the night premium hours (see cell B40) than for Model1a.

5. Model1d Sheet

This example is the same as Model1a, except that the incoming mail arrival profile and final dispatch time is used instead of the originating mail profile. The total volume is the same at 10 million pieces, but the flow path productivity is assumed to be higher at 400 pieces per hour. The optimal solution, $1.0 million, again matches the minimum cost solution.

6. Model1e Sheet

This example is the same as Model1d, except that night premium costs are included for incoming mail processing. Total costs are $1.044 million, which is 4.4% higher than Model1d. The incremental effect of the night premium for incoming mail is higher than for originating mail, because more night shift hours are required for incoming mail.

Model1f Sheet

This example is a combination of Model1a and Model1d in that both originating and incoming mail are included in the same model. The solution shows that the model solution for both originating and incoming mail matches the minimum cost for both mail types.

7. Model1g Sheet

This example is the same as Model1f except that night premium costs are reflected.

8. Model2a Sheet

This example illustrates how multiple days of the week as well as hourly variations can be included in the optimization process, although only a 2-day scenario is used for this case. The model is for originating mail only with no night premium costs, but overtime is reflected with an assumed wage 5% higher than straight time ($42.00/hr) and penalty overtime at an assumed wage 50% higher than straight time ($60.00/hr). Rows 8-41 (Day 1) reflect a typical Monday volume level of 13.3 million pieces, while Rows 44-77 (Day 2) reflect a typical average day volume of 10.0 million pieces. It is assumed that the number of full-time regular employees is the same on both the high- and low-volume days. So the optimization problem involves a tradeoff between using the lower number of full-time employees required for Day 2 volume together with significant overtime on Day 1 versus using the larger number of full-timers required for Day 1 and underutilizing them on Day 2.

Columns AC:AZ have been added to this sheet to incorporate 24 patterns for the first hour of overtime; columns BA:BX have been added to incorporate 24 more patterns representing a second hour of overtime; and columns CA:CV have been added for a third hour of (penalty) overtime. For example, column AC (variable x1a) contains only one non-zero value representing a ninth hour of work for Pattern 1 employees with regular 8-hour shifts starting at midnight. Similarly, column BA (variable x1b) contains only one non-zero value representing a tenth hour of work for Pattern 1 employees. Finally, column CA (variable x1c) also contains only one non-zero value representing an eleventh hour of work for Pattern 1 employees. Constraints for these overtime variables are added in the solution process to ensure that there can't be solutions with scheduled overtime hours for employee patterns with no scheduled full-time employees, and that there can't be a second hour of overtime scheduled unless a first hour has been scheduled, etc.

It can be seen from cells B3 and B45 that the unconstrained minimum cost solution for the two-day period is $4.67 million. The optimal LP solution shown at cell A8 is $4.86 million, so the day-of-week peak load cost in this example would be $0.19 million. Comparison of the model solution for each day shows that there was not enough regular overtime available to handle all the extra 33% of Day 1 volume, so 3,394 additional full-time

employees (compared with the Day 2 requirements) were added for both days as well as a near-maximum use of 3 hours of overtime per full-time employee.

9. Delivery and Transportation LP Models

No numerical examples were developed for the delivery and transportation functions due to lack of time, but the examples presented in the main paper (pp. 48-51) could easily be put into the same LP matrix format used for the mail processing examples.

For the delivery example, the rows of the matrix would correspond to the days of the week, and the columns to employee types. For the transportation example, the rows would be the delivery units and the columns would be the capacity used for each possible truck run.

Appendix D   Slides from Panzar's September 30 Commission Briefing

Postal "Peak Load Costs"

Professor John C. Panzar

Postal Regulatory Commission

September 30, 2011

1

Introduction and Summary

- Why study "peak load costs"?
  - What are "peak load costs"?
- Stylized models of mail processing and delivery
  - Costs of service standards
  - Costs of variable flows
- Towards an econometric analysis

2

Why study "Peak Load Costs"?

- In response to continuing declines in mail volume, the Postal Service has proposed drastic changes in its operating procedures: e.g.,
  - 5 day delivery
  - Reduction of service standards for preferential mail
- These changes involve major departures from the Postal Service's established operating procedures.
- Insights regarding the cost effects of these changes can not be readily obtained directly from traditional USPS models and analyses.

3

Goals of Analysis: develop theoretical models of changed operations

- Basic approach is to develop an optimizing model capable of:
  - Calculating "before" and "after" solutions
  - Comparing the minimized costs of each
- Makes possible qualitative and quantitative evaluation of "Peak Load Costs:" i.e., the impacts of changes in
  - Service Standards
  - Arrival profiles
  - Work rules

4

Objective of a stylized models of postal operations

- Capture a nontrivial optimization problem facing the operations manager
  - Eg., trade-off between capital and labor
- Characterize the *constraints* facing the manager
  - Eg., service standards, work rules
- Solve the problem with and w/o constraints
- Reduction in minimized cost is the cost of the constraint in question

5

Stylized model of mail processing
Extension of Crew and Kleindorfer 1990

- Two types of mail with a daily demand cycle:
  - Preferential mail must be processed during the period it arrives
  - Non preferential can be deferred one period
- Two periods ("tours") in the daily cycle
  - Period/Tour 1 exhibits: peak in preferential mail arrivals and a wage premium
- Two inputs required in fixed proportions
  - Labor can be adjusted between tours,
  - Machine capacity is the same for both tours

6

3

Notation used in analysis

- N_t = volume of non preferential mail in period t = 1, 2.
- $N = N_1 + N_2$ = total period volume of non preferential mail
- P_t = volume of preferential mail in period t=1, 2.
- $P = P_1 + P_2$ = total volume of preferential mail
- $\Delta = P_1 - P_2 > 0$ measures the variability of preferential mail across periods
- $V = N + P$ = total volume of mail
- Y_t = amount of non preferential mail processed in period t = 1,2.
- M = machine processing capacity.
- L_t = labor hired in period t = 1, 2.
- $w = w_2$ = base (period 2) wage rate
- ρ = shift wage premium
- $w_1 = (1+\rho)w$ = wage rate in period 1.
- m = unit cost of machine processing capacity

7

Intuitive(?) description of model: Objective function and decision variables

- Operations manager seeks to minimize the total costs of processing *a given volume of mail:*

$$C = wL_2 + w(1+\rho)L_1 + mM$$

- His decision variables are
  - Number of machines installed (M)
  - Labor employed in each period (L_t)
  - Non pref mail processed in each period (Y_t)

8

4

Constraints facing the decision maker

- All the non pref mail must be processed:

$$Y_1 + Y_2 \geq N_1 + N_2 = N$$

- Sufficient labor must be employed each period to process its designated mail:

$$L_1 \geq P_1 + Y_1; \qquad L_2 = P_2 + Y_2$$

- Machine capacity must be sufficient to process the mail assigned to each period

$$M \geq P_1 + Y_1; \qquad M \geq P_2 + Y_2$$

9

Intuitive(?) solution of the model

- This is a *very* simple LINEAR PROGRAMING (LP) problem.
- It is simple enough to "intuit" the solutions for two parameter configurations that determine whether machine capacity is "cheap" or "expensive."
- Cheap Capacity: $w\rho > m$. It pays to add a unit of machine capacity so that an additional piece can be processed on Tour 2, avoiding a unit of the Tour 1 wage premium.
- Expensive Capacity: $w\rho < m$. It does not pay to add a unit of machine capacity to avoid incurring the per unit Tour 1 wage premium.

10

Case (i): Cheap Capacity

- *If* machine capacity were installed sufficient to serve Tour 2 preferential mail and *all* non pref ($M=P_2+N>P_1$), there would be enough capacity to serve Tour 1 preferential mail as well.
  - (This will always be the case when non pref volumes are greater than or equal to preferential volumes: i.e., $N > P = P_1 + P_2$)
- *Imagine* shifting one unit of non pref to Tour 1. You could install one unit less of capacity, saving m, but you would incur the wage premium on one additional unit, costing $w\rho$.
- When capacity is cheap ($m < w\rho$) this does not pay.

$$C_{(i)} = w(1+\rho)P_1 + (w+m)(P_2+N)$$

11

Case (ii): Expensive Capacity

- When machine capacity is expensive, it makes sense to conserve it by fully utilizing capacity on both Tours: $M = V/2$.
- Now, *imagine*, shifting one unit from Tour 1 to Tour 2 in order to avoid the Tour 1 wage premium.
  - The resulting wage savings are $w\rho$.
  - However, an additional unit of machine capacity must be installed to process this mail on Tour 2, increasing machine costs by m.
  - When capacity is expensive ($m > w\rho$), it's not worth it.

$$C_{(iv)} = w(1+\rho)P_1 + w(P_2+N) + mV/2$$

12

6

Peak Load Cost Savings from Elimination of Service Standards

- In order to measure the costs savings of, for example, the elimination of service standards for preferential mail, one needs to compare the costs with and w/o said standards.
- Costs w/o service standards are readily calculated because mail is now homogeneous in the sense that the arrival pattern has no impact on costs: i.e., it is optimal to "roll the mail." There are two relevant cases:
  - When machine capacity is "cheap," $m < w\rho$, the focus is on avoiding the Tour 1 wage premium: $M = V$ and *all* mail is processed on Tour 2.

$$C_c = (w + m)V.$$

  - When machine capacity is "expensive," $m > w\rho$, the optimal policy is to fully utilize capacity: $M = V/2$, with one half the total volume processed on each Tour. Thus

$$C_e = (2w + w\rho + m)V/2$$

13

Case (i) Cheap Capacity.

- The difference between minimized costs with and without service standards is

$$SS_{(i)} = C_{(i)} - C_c = w(1+\rho)P_1 + (w+m)(P_2+N) - (w+m)V$$

$$= w\rho P_1 - mP_1$$

- "Rolling the mail" results in an expansion of capacity, the costs of which partially offset the elimination of premium wages.

14

Case (ii) Expensive capacity.

- The difference between minimized costs with and without service standards is

$$SS_{(ii)} = C_{(ii)} - C_e = w(1+\rho)P_1 + w(P_2+N) + mV/2 - (2w+w\rho+m)V/2$$

$$= wV + w\rho P_1 + mV/2 - (2w+w\rho+m)V/2 = 0$$

- In this case, it was already optimal to process one half of total mail volumes in each period, thereby incurring shift premium costs of $w\rho(V/2)$. The optimal mail processing configuration remains unchanged after the relaxation of the service standard, so the appropriately measured service standard costs are zero.

15

Counterfactual: Smoothing the peak

- Even with service standards in place, it may be possible to induce some preferential mailers to change their behavior so that their volume arrives during Tour 2.

- If it is necessary to induce mailers to shift their deposits of mail by offering a discount, it is important to determine how much is saved by shifting a small amount (one unit) of preferential mail from Tour 1 to Tour 2

- The cost savings from such smoothing can be analyzed more easily when costs are rewritten in terms of total volumes and peak size, Δ:

16

8

Savings from a one unit reduction in the size of the Tour 1 peak.

- When capacity is cheap, shifting 1/2 unit from Tour 1 to Tour 2 (so that Δ increases by one unit) necessitates an increase in capacity, which reduces the cost savings:

$$C_{(i)} = w(1+\rho)(P+\Delta)/2 + (w+m)[(P-\Delta)/2+N];$$
$$dC_{(i)}/d\Delta = (w\rho-m)/2$$

- When capacity is expensive, the shift of 1/2 unit of preferential mail from Tour 1 to Tour 2 has no effect on optimal machine capacity:

$$C_{(ii)} = w(1+\rho)(P+\Delta)/2 + w[(P-\Delta)/2+N)] + mV/2 ;$$
$$dC_{(ii)}/d\Delta = w\rho/2$$

17

Excess capacity considerations

- Above analysis takes the economists long-run, "drawing board" perspective
  - An *ex ante* planning model; before any labor has been hired or machines purchased.
- In the present situation, it may be reasonable to assume that the Postal Service has an existing stock of machines, M^0, that can not be reduced in the short-run.
- Fortunately, this situation is readily analyzed with our model.

18

9

A short run analysis

- Technically, all that is required is to change the capacity constraint equation to

$$M + M^0 \geq P_1 + Y_1; \qquad M + M^0 \geq P_2 + Y_2$$

- Then, for a large enough legacy stock of machines (i.e., $M^0 > P_2 + N$), it is optimal to purchase no new machines, so that

$$C^0 = w(1+\rho)P_1 + w(P_2+N) + mM^0$$

- Existing machines are a fixed cost, marginal costs are "as if" capacity is "really" cheap: i.e., $m = 0$.

19

Short-run analysis of the cost of Service Standards

- In the short-run we are (likely) in the cheap capacity case so the appropriate measure is

$$SS_{(i)|m=0} = C_{(i)|m=0} - C_{c|m=0} = w(1+\rho)P_1 + w(P_2+N) - wV = w\rho P_1$$

- Again, total Tour 1 shift premia measure the costs of relaxing service standards.
- *CAVEAT*: The above equation is based upon the assumption that $M^0 > V$.
  - Suppose M^0 was set optimally for initial volumes $P_1^0 > P_1$, $P_2^0 > P_2$ and $N^0 > N$. This ensures that, $M^0 \geq V^0/2 > V$, but it is quite possible that $M^0 < V$.
  - Intuitively, eliminating Tour 1 processing by shifting all mail to Tour 2 will require more machine capacity than was previously optimal. There is no guarantee that there will be enough excess capacity in the short-run.

20

10

Numerical Analysis of Mail Processing

- Linear Programming with Excel
- Example with
  - 8 hour shift constraints
  - Service Standards
  - Shift premium

21

A *very* simple model of peak load considerations in delivery

- Mail volume variability is over "days" of the week. (No distinction re pref versus non pref.)
- Two "days" in the week, volumes V_1 and V_2.
  - Total volume = $V = V_1 + V_2$
  - Size of peak = $\Delta = V_1 - V_2 > 0$
- R = number of routes for delivery area
  - *Daily* wage cost of a route = w
  - *Capacity* of a route is r units of volume
- There is no "volume variability" up to the capacity of a route, but the number of routes operated is constant across the week.

22

11

Analyzing the delivery model

- Given these assumptions, the operations manager's problem is to

$$\min_R C = 2wR$$

subject to: $R \geq V_1/r$
$R \geq V_2/r$
- Obvious (simple) solution
  - Optimal number of routes is $R = V_1/r$
  - Minimized level of costs is

$$C = 2wV_1/r = w(V+\Delta)/r$$

23

Counterfactuals in the delivery model

- Two obvious causes of "peak load costs" in this simple delivery model:
  - The *work rule* constraint that the number of routes is constant over the "week"
  - Mail variability across days
- To analyze these costs, derive the solutions to counterfactual "relaxed" and "smoothed" models and subtract from the level of minimized base case costs.

24

Relaxing delivery work rules

- Now, the operations manager's problem is to
$$\min_{R1,R2} C_x = w(R_1+R_2)$$
- subject to: $R_1 \geq V_1/r$ and $R_2 \geq V_2/r$
- Obvious (simple) solution
  - Optimal route numbers are $R_1 = V_1/r$ and $R_2 = V_2/r$
  - Minimized level of costs is $C_x = w(V_1+V_2)/r = wV/r$
- Then, the peak load cost savings from relaxing delivery work rules is

$C - C_x = 2wV_1/r - w(V_1+V_2)/r = w(V_1-V_2)/r = w\Delta/r$

25

Costs savings from Peak Smoothing

- Next, examine the cost savings that would result if mailers could be induced to shift ½ of a unit of mail from day 1 to day 2.
  - Again, this would decrease Δ by one unit.
- Differentiating the minimized cost function with respect to Δ yields
$$dC/d\Delta = w/r$$
- The cost savings from shifting a unit of volume from peak to off peak is efficient cost per unit.

26

13

Quantitative limitations of our approach

- Economists refer to the approach outlined above as a *normative analysis*.
- It is about what "ought" to happen.
  - We specify an optimization problem and compare the solutions before and after relaxing various constraints.
  - This approach assumes that the subject of the analysis is minimizing costs both "before" and "after" the proposed changes.

27

Outline of an Econometric Approach (What would Frank do?)

- Given enough data, one could also use the present (or similar) theoretical model to specify and estimate econometric equations that could be used to provide quantitative measures of the cost effects of interest.
- Frank has in mind a highly disaggregated set of panel data (i.e., data that varies over both time and facility) on postal operations
- Obviously, an empirical analysis using historical data can not *directly* quantify counterfactual effects.
- However, historical data can be used to estimate model *parameters*, which can be used for extrapolation.

28

14

Examples of data required for an econometric analysis

- Mail Processing
  - Mail arrival patterns by day and hour
    - First Handled Pieces (FHP) by plant – MODS
    - Permit Mail Statements
    - IMB data
  - Labor hours worked – TACS
- Delivery
  - Mail volumes by day and route – DOIS
  - Labor hours worked by route – TACS
- Highway Transportation
  - Contract data – HCCS
  - Capacity utilization - TRACS

29

Conclusions

- Models can be developed to analyze
  - Cost savings from proposed operating changes by the Postal Service
  - "Peak Load Costs"
  - Impacts of declining volumes
    - Excess capacity
- Model formulation helps identify data requirements for
  - Linear Programming analysis
  - Econometric analysis.

30

15