BEFORE THE
POSTAL RATE COMMISSION
WASHINGTON, D.C. 20268-0001

OFFICE OF THE CONSUMER ADVOCATE
NOTICE OF POSTING OF LABOR DEMAND MODEL FOR MAIL SORTING OPERATIONS ON POSTAL RATE COMMISSION WEBSITE AND NON-EXPERT'S GUIDE TO THE MODEL
(May 31, 2002)

The Office of the Consumer Advocate (OCA) hereby notifies the postal ratemaking community that it has posted a paper entitled "An Empirical Model of Labor Demand for Mail Sorting Operations" in the OCA section of the Postal Rate Commission (PRC) website. Mark J. Roberts, a professor of economics at Penn State University, is the author of the paper. Professor Roberts will conduct a morning seminar and afternoon workshop on Thursday, June 20, to present his results and methodology and answer questions about the study.

Non-Expert's Guide to Professor Roberts' Econometric Analysis

I am going to attempt to summarize and "translate" a technical paper into something comprehensible to the lay reader. I have struggled myself with papers of this type over many rate cases and will try to explain the data patterns identified by Professor Roberts so that a wider audience can follow his effort and understand his conclusions. Professor Roberts is not responsible for any errors I may commit.

Purpose of the Econometric Analysis. The purpose of Professor Roberts' analysis is to estimate the extent to which mail processing labor use (reported as hours worked) increases or decreases as mail volumes increase or decrease. The extent of this increase or decrease – known as volume variability in PRC proceedings (and

marginal cost analysis in other economic fora) – is used by the Commission as the method for attributing mail processing costs to each subclass or service. Cost segment 3 mail processing costs are a significant component of the costs of most subclasses and services. Consequently, the percentage of these costs that is attributed (due to a volume variability analysis) can have a significant impact on the rates developed for each of the subclasses and services.

Professor Roberts develops an econometric model that essentially looks at how labor hours in mail processing (and, thus, labor costs,) change with respect to changes in volume. A number of real-world factors can affect how labor use changes. For example, installing new labor-saving devices such as Bar Code Sorters (BCS) or Flat-Sorting Machines (FSM) can have an effect on labor use that is simultaneous with the volume effect. Professor Roberts, consistent with past Commission practice, must *isolate* the volume effect from all others. He constructs econometric models (formulas) that identify and control for the effect many observable factors besides volume, including the types of capital equipment and mix of technologies used in the plant, have on labor hours. For those factors that have an important effect on mail processing hours expended, but which are unknown or unobservable, he can isolate and minimize *their* effect on labor use by using a combination of econometric techniques that are described in detail later in this Guide.

Data Used by Professor Roberts. Professor Roberts uses the data employed by Postal Service witness Bozzo in the last rate case, Docket No. R2001-1. Although Professor Roberts might have preferred additional data sets or data somewhat different than that collected and furnished by the Postal Service, he believes that the data

available to him were sufficient to draw reliable conclusions about the volume variability of labor use.

The data that Dr. Bozzo used in Docket No. R2001-1 and that were furnished to OCA in response to discovery requests were primarily data on output (the number of pieces of letters, flats, and packages) and hours of labor for specific operations within 321 mail processing facilities.[1] Certain measures of capital equipment, such as the expenditures on BCSes or FSMs in a plant were also available. Data had not been provided for other types of facilities, such as Bulk Mail Centers (BMCs), nor for retail facilities.

Economic Model Used in the Analysis. Professor Roberts' approach begins with a description of the production process in a mail-sorting facility. The plant uses inputs of labor hours in different sorting operations and capital to turn raw, unsorted (or only partly sorted) letters and flats into outputs of sorted letters and flats.[2] This is the object to which all processing plant activities are directed. The Postal Service collects and receives letters and flats from an enormous number of diverse originating points and must separate and combine these mailpieces into the groupings that eventually will be routed to specific destinating points. Each movement of a letter or flat through a processing plant is directed at that purpose. Professor Roberts makes that the basis for each of his econometric models. Because letters are handled separately from flats

---

[1] Dr. Bozzo indicated that these 321 plants generally consisted of Processing and Distribution Centers (P & DCs) and Processing and Distribution Facilities (P & DFs). Docket No. R2001-1, Tr. 11A/3856-57.

[2] Plants also have a parcel mailstream that Professor Roberts has not analyzed because the Postal Service was unable to provide the type of output data required for his model. Professor Roberts has also performed a preliminary analysis of the Priority Mail processing stream in plants, which is discussed later in this Guide.

in both manual and automated operations in the plant, Professor Roberts can separate the mail by shape and analyze the labor used in sorting each shape.

The Postal Service's sorting of each distinct mail shape (i.e., letters and flats) is generally accomplished in manual labor operations or automated equipment/labor operations. Professor Roberts views these as parts of the whole production process. That is, a combination of manual and automated operations for letter sorting (or flat sorting) ultimately yields sorted letters and sorted flats. His econometric models reflect that structural view. His measure of output or volume for a plant is the number of "first-handled pieces" (FHP) rather than "total pieces fed" (TPF, the volume measure used by the Postal Service). FHP is essentially a count of the unsorted pieces coming into the plant, and it is the same number as the sorted pieces coming out of the plant. TPF, on the other hand (the measure used by the Postal Service), is a measure of the amount of processing done in a specific, discrete operation.

TPF in an operation may reflect many factors that are not related to the total volume of sorted letters or flats produced in the plant. These include the plant manager's decision to flow certain pieces to manual or automated operations, how many separations a particular plant is able to make based on the amount of automated equipment it has in place, and the sophistication of its equipment. In Professor Roberts' view, the variability in labor use in all sorting operations should be measured with respect to the number of unique pieces of mail that are processed in the plant, not with respect to the number of pieces fed into a particular operation.

Analyses performed by the Postal Service have missed that fundamental principle. Instead, the Postal Service has focused on discrete operations within plants

and performed a volume variability analysis for each discrete operation in isolation from other substitutable or complementary operations within the plant. The results of such an exercise do not measure the essential relationship being studied, i.e., the variability of labor hours with respect to the number of pieces of sorted mail that are produced in the plant.

Control Variables in the Econometric Models. Professor Roberts observed a number of factors that caused labor hours to change, but which were not related to volume. For example, different technology mixes evolve over the span of time observed. (Letter Sorting Machines (LSMs) were phased out in the time periods for which data were furnished, while BCSes gradually replaced them.) The amount of capital used in different sorting operations varies over time and across plants and explains changes in the hours used in specific sorting operations. Different wage rates for workers in manual and automated operations also explain adjustment in hours used in each sorting operation. National conversion factors used to calculate FHP figures were changed in 1999. This complicates an analysis that spans the pre-conversion and post-conversion periods. To account for all of these non-volume-related phenomena, Professor Roberts employs a set of control variables that isolate the non-volume effects.

Econometric Methodology. Professor Roberts performs the empirical analysis using several different approaches. He begins with Ordinary Least Squares (OLS) which is a common technique used by economists. OLS produces a line which best fits all of the data points being analyzed, minimizing the distance of all data points to the line finally produced. OLS can generate an accurate estimate of the output variability if

the error term in the equation (the one that covers all of the unspecified nonvolume factors) is not correlated with any of the explanatory variables in the model (i.e., volume as well as all of the control variables). This condition – no correlation between the error term and the specified explanatory variables – is often stated as there being no covariance between the error term and the explanatory variables. An equivalent statement is that the included explanatory variables are exogenous with respect to the error term. The condition of non-correlation between the error term and the model's explanatory variables is necessary for reliable results. When this condition does not exist, the explanatory variables are endogenous with respect to the error term, and OLS will produce misleading and inaccurate estimates of the variability of hours with volumes. This is true particularly if the output (volume) variable is endogenous with respect to the error term.

Three General Ways for the Output Variable (Volume) to Be Endogenous.

(1) The first of the ways that endogeneity can arise is the omission of necessary variables in the regression equation. Omitted variables are captured in the error term. If these variables are correlated with the output variable (volume) then the error term will also necessarily be correlated with the output variable. Depending on the nature of the correlation, this results in estimates of the volume variability that are either too high or too low.

(2) A second way for the output variable to be endogenous in the regression equation arises when there is measurement error in the output variable. Ideally, FHP should be a pure count of sorted letters and flats, but the data furnished by the Postal Service involve a conversion factor discussed on the previous page. The

differences between a pure FHP count and the derived FHP count (following application of conversion factors) are captured in the error term. This can cause endogeneity between the output variable and the error term and will cause the estimate of the volume variability to be too low.

(3) A third source of endogeneity arises if labor hours in a sorting operation and the output (volume) of that operation are chosen simultaneously in the plant. This is likely to be true in postal plants, with postal managers frequently deciding the direction of pieces to given operations and the amount of labor to be used in each operation, given the amount of automated equipment available and the wages for the laborers in various operations. It is likely, therefore, that the Postal Service's econometric analysis, which relies on a volume measure, TPF, that is specific to an operation, will violate the condition of exogeneity – i.e., output will be endogenous. Professor Roberts' model, on the other hand, does not suffer endogeneity of this type since the plant manager likely will be expected to sort all of the mail arriving at the plant and will not be able to control the amount of mail or the shapes of mail that must be processed. Professor Roberts applies a statistical test for endogeneity, finds that the output measure is endogenous, and then utilizes statistical methods that control for endogeneity.

Correcting for Endogeneity. Professor Roberts is concerned that there may be influences on output in each plant that are specific to the plant, that do not vary with time, that cannot or have not been measured (else they would be specified as explanatory variables), that would be part of the error term as omitted variables, and

that are correlated with output.[3] This endogenous influence can be eliminated or "neutralized" in two ways. First, one may look solely at the *differences* in variable values between time periods, rather than the plant values themselves. Any plant-specific effect that is not variant with time will drop out of the estimating equation. Second, each plant's data (hours, volume, and other explanatory variables) can be expressed as deviations (or differences) from the plant-specific mean of the variable. The plant-specific mean (which has averaged identical values over all time periods) never differs from any specified observation. Therefore, the differences between a plant-level observation and the plant-specific mean of that variable are zero and thus have eliminated the plant-specific effect. Either of these methods can correct for the bias that arises from the omitted variable.

Presence of Measurement Error in Output Is Identified. While useful in removing one source of output endogeneity, differencing the data does not correct, and can actually increase, the bias (understatement of output variability) caused by measurement error. Professor Roberts finds evidence that measurement error in output is an important problem. Thus, estimates based on differencing the data, such as employed by the Postal Service, will tend to produce systematically lower variability results than OLS without such restrictions. To correct for this understatement, Professor Roberts employs an econometric methodology known as instrumental variables (IV). In this method, he relies on a variable that is highly correlated with

---

[3] Professor Roberts explains that this is sometimes referred to as unobserved heterogeneity. This means that there are characteristics that differ across plants (the plants are heterogeneous for these characteristics) and have an influence on labor use that has nothing to do with volume. As with other omitted variables that can mask the volume-hours relationship, Professor Roberts must try to isolate or remove the influence.

output but is not correlated with the error term in the regression equation. Professor Roberts finds that volume variability estimates are substantially higher when the IV estimator is used.[4]

Results of the Roberts Study. Professor Roberts reports several estimates of the volume variability for flats. His estimator for flats using different aggregations of sorting operations results in variability figures between .838 and .956. For letters, the different aggregations result in estimates between .951 and 1.026. Looking at individual sorting operations he finds variabilities that are not statistically different than 1.0 for manual flat and letter sorting and automated letter sorting operations. Estimates less than one are found for mechanized flats. He reports estimates of a composite variability for the whole plant that ranges from .952 to .992. That is substantially higher than the composite estimate of .71 reported by the Postal Service.

Priority Mail. Professor Roberts' model is well suited to determine labor use variability for any properly defined mailstream. The analysis of Priority Mail presents some additional complications because not all plants are engaged in sorting this category of mail. In analyzing labor adjustment, it becomes necessary to control for the

[4] Professor Roberts explores several differences in specifications between his model and that employed by the Postal Service and finds that they are not significant. The main source of difference between his model and the Postal Service's is the use of the instrumental variable.

entry and exit of plants to the sorting operation. He undertakes a preliminary analysis of the Priority Mail sorting operations that focuses exclusively on the subset of plants that always sort this mail. The preliminary Priority Mail results indicate variabilities between .89 and 1.105 for manual operations and greater than one for mechanized operations.

Professor Roberts' paper is strongly endorsed by the OCA as a carefully developed model of volume-labor variability that has been thoroughly tested and whose first principles have been fully explained. All of the reasoning steps employed by Professor Roberts have been laid out so that readers can follow his choices and understand the results generated by his model. The Roberts' analysis is a work product of the OCA, not the Postal Rate Commission, and reflects only the policies and views of the OCA.

Respectfully submitted,

Shelley S. Dreifuss

Shelley S. Dreifuss, Director
Office of the Consumer Advocate

1333 H Street, N.W.
Washington, D.C. 20268-0001
(202) 789-6830; Fax (202) 789-6819
e-mail: dreifusss@prc.gov