MPA-RT-2

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

POSTAL RATE AND FEE CHANGES, 1997

DOCKET NO. R97-1

REBUTTAL TESTIMONY
OF
PAUL HIGGINS

ON BEHALF OF
ALLIANCE OF NONPROFIT MAILERS,
AMERICAN BUSINESS PRESS,
COALITION OF RELIGIOUS PRESS ASSOCIATIONS,
DOW JONES & COMPANY, INC.,
MAGAZINE PUBLISHERS OF AMERICA,
THE McGRAW-HILL COMPANIES, INC.,
NATIONAL NEWSPAPER ASSOCIATION,
AND
TIME WARNER INC.

Communications with respect to this document may be sent to:

James R. Cregan
Counsel
Magazine Publishers of America
Suite 610
1211 Connecticut Avenue, N.W.
Washington, D.C. 20036
(202) 296-7277

March 9, 1998

TABLE OF CONTENTS

AUTOBIOGRAPHICAL SKETCH

My name is Paul Higgins. I am a Senior Analyst with Project Performance Corporation. I filed testimony in this proceeding responding to Notice of Inquiry No. 4 on behalf of the Magazine Publishers of America. A full description of my background and qualifications appears in that testimony, filed as MPA-NOI-1.

I. PURPOSE AND SCOPE OF TESTIMONY

For over 25 years spanning nine rate cases, the Postal Service (and the Postal Rate Commission) assumed that mail processing costs were almost totally volume variable, or, in econometric parlance, that the cost elasticity of mail processing was approximately 1. During this time and these cases the Service offered no quantitative or statistical analysis to support that assumption despite the fact that mail processing has always been the largest component in both the attributable and the total costs of the Postal Service, and despite the fact that this period has been marked by both rapid and profound improvements in statistical techniques and by substantial changes in the way mail is prepared and processed. Any analysis that it did offer relied on supposition rather than data.

The record in this case clearly contradicts this long-held assumption, and is devoid of any evidence to support it. Indeed, there is substantial evidence to the contrary. As I state in my conclusion:

- There is no doubt on this record that the cost elasticity of mail processing is substantially less than 1.
- The long-held unsubstantiated assumption that mail processing costs are almost totally volume variable has been shown to be invalid.
- There is ample evidence on this record to resolve the issue of the cost elasticity of mail processing econometrically.

In this case, much to its credit, the Service tested this previously unsubstantiated but extraordinarily important assumption. USPS witness Bradley presented testimony (USPS-T-14) on the cost elasticity of mail processing based on an econometric study. It showed that the cost elasticity of many mail processing operations is much less than 1. Thus, he found that there are large returns to scale in mail processing operations.

His testimony has prompted three Presiding Officer's Information Requests, one Notice of Inquiry from the Commission, testimony from UPS witness Neels (UPS-T-1), and testimony from OCA witness Smith (OCA-T-600). Professor Bradley, witness Neels, and I all filed testimony responding to Notice of Inquiry No. 4. (NOI 4).

In this testimony, I review the written direct testimony of witnesses Neels and Smith, witness Neels's oral testimony, as well as witnesses Bradley's and Neels's responses to NOI 4. To set a perspective for this task, and since the direct testimonies of witnesses Neels and Smith criticize witness Bradley's analysis, I also review his analysis.

Given the evidence in this case and the current state of the art in applied econometrics, I believe that the issue of the variability of mail processing is fully amenable to an econometric analysis. In fact, given the evidence and the state of the art, it would be unreasonable to claim that the evidence is insufficient to allow an econometric cost elasticity estimate. I believe that if a "science court" comprised of leading economists were convened to decide whether the issue of mail processing cost elasticity should be resolved through the use of econometrics, that "court" would rule unanimously that not only should it be, but that any other approach would be unreasonable.

The evidence in this case makes it equally clear that the elasticity of mail processing is less than 1. In fact, as I pointed out in my response to NOI 4, "the average mail processing variability is no higher than Professor Bradley's figure of 76.4 percent." Tr. 29/16125-26. As I also pointed out in my response to NOI 4, the results of the requested analysis rule out the pooled model and show that the fixed-effects and the unrestricted models produce reasonably consistent results.

In the remainder of this testimony, I explain the basis for my conclusions that witnesses Neels and Smith are wrong in their criticisms of witness Bradley's approach and results. Witnesses Neels and Smith have both tried to attack the theoretical underpinnings and analytical approach of witness Bradley. In Section II, I rebut their criticisms, showing that they lack merit. Witnesses Neels and Smith have also tried to attack witness Bradley's data scrubs. In Section III, I show that witness Bradley's use of data is appropriate - his data scrubs are standard practice and his use of panel data helps in the error-in-variables problem. Witnesses Neels and Smith have also raised a wide range of issues pertaining to witness Bradley's econometric methods. In Section IV, I show that witness Bradley has used proper econometric methods and that their criticisms are groundless. Section V presents my analysis of the oral cross-examination of witness Neels. Finally, Section VI presents my conclusions.

II. ATTACKS ON WITNESS BRADLEY'S THEORETICAL UNDERPINNINGS AND ANALYTICAL APPROACH LACK MERIT

Witnesses Smith and Neels have both claimed that there are defects in witness Bradley's theoretical underpinnings and in his analytical approach. In this section of my testimony, I show that neither economic nor econometric theory provide any basis for these claims.

In part A, I examine witness Smith's claim that visual inspections of witness Bradley's data refute his results. I explain why statistical analysis is superior to visual inspection.

In part B, I analyze witness Smith's claim that witness Bradley should have estimated a production function rather than a cost function. I explain that while either approach is acceptable from a theoretical position, witness Smith's preferred approach is intractable. In part C, I also show that witness Smith's claim that witness Bradley's *method excludes the impact of capital is wrong. In similar fashion, I show that contrary* to witness Smith's assertions, witness Bradley's model is sufficiently detailed for the task at hand.

Finally, in part D, I refute points that witnesses Neels and Smith have raised about "length of run" issues. I show that they are extremely confused on these issues, not just because their criticisms are mutually contradictory, but because they also have misstated what determines "length of run" and how it influences ratemaking.

A. Visual Inspection of Data is Vastly Inferior to Statistical Analysis

Visual inspection of selective two-dimensional "slices" through multivariate data may sometimes be useful for suggesting relationships. It is, however, inadequate and misleading as a means of analyzing them. It is <u>inadequate</u> because it is entirely subjective – the human eye is simply incapable of discerning the curve or surface that best describes a complex cloud of data points, particularly if it has more than two dimensions; if the points are numerous, bunched up, or overlap each other; if the points are dispersed in irregular patterns; or if the points are not precisely indicated on the graph. It is <u>misleading</u> because a two-dimensional plot restricts the viewer to looking at partial relationships in the data, excluding from view other variables that may affect the dependent variable, or the relationship between the plotted variables. In effect, it invites the viewer to assume that all relevant information is either summarized by the graph or held constant, when in fact other confounding variables are merely hidden from view.

Witness Smith makes clear that his case for rejecting Professor Bradley's fixed-effects model, and for preferring the pooled model, is based largely on his examination of numerous plots of witness Bradley's data for various mail processing operations. Tr. 28/15841-49. In this section of my testimony, I show that witness Smith's analysis of these plots (exhibits OCA 602 and 603, Tr. 28/15870-77 and 15878-96, respectively) is without merit, as it contains numerous unwarranted assertions, unfounded generalizations, and errors in judgment.

Witness Smith's graphs consist of two basic types. The first set, contained in exhibit OCA 602, consists of plots of the logarithm of total hours against the logarithm of total piece-handlings (TPH) for each of seven direct mail processing activities at MODS facilities – OCR, BCS, LSM, Manual Letters, Manual Flats, SPBS Non-Priority, and Manual Priority – in which data from all of the facilities are combined. The second set, contained in exhibit OCA 603, consists of plots of hours against TPH for selected individual facilities in each of four activities: Manual Letters, Manual Flats, OCR, and LSM.

Regarding the first exhibit, witness Smith states:

> *The data presented in exhibit OCA 602 are visually compelling in demonstrating a variability approaching 100 percent between labor hours and mail volume [Tr. 28/15847.]*

This is, quite simply, indefensible. There is no conceivable way that witness Smith or anyone else could tell, by looking at these plots, whether the variability of hours with respect to TPH in each activity is "approaching 100 percent" or any other particular value. One can certainly form general impressions of what the elasticity of one variable might be with respect to another by looking at data plots, but no more than that. Such impressions, however, as witness Bradley has testified, can be misleading. Tr 11/5581-82. This is true for three major reasons.

First, the data that witness Smith plotted, and that witness Bradley and others statistically analyzed, are not simple cross-sections or time-series, but rather constitute a panel containing repeated observations on a cross-section of facilities over time. As such, they contain variation in the hours-TPH relationship in two different dimensions –

temporal (within any given facility over time) and cross-sectional (across different facilities in any given time period) – rather than just one. Witness Smith's graphs ignore this distinction.

To illustrate the implications of this error, consider the following simple, two-part thought experiment: (1) Turn to exhibit OCA 602 and select any two points at random from any one of the plots. (2) Now try to determine whether the two points are (a) observations from different sites in the same time period, (b) observations from the same site in different time periods, or (c) observations from different sites in different time periods. The point is, of course, that it can't be done using these plots. Yet according to witness Smith, making such distinctions is crucial:

> [T]he measurement of changes in labor with short-run changes in output is irrelevant for the purposes of this proceeding. The relevant measurement of cost incidence should focus on the expansion path reflecting expansion or contraction of the scale of the facility in the foreseeable future.... [Tr. 28/15841.]

Thus, witness Smith contradicts himself on this point: on one hand he claims to believe that short-run variations in costs should be ignored for purposes of analyzing cost variability, and that long-run variations are all that matter; on the other hand, his graphical analysis, which is the centerpiece of his testimony[1], conflates short-run and long-run variations in cost to the point of permitting no distinction between them whatsoever. Witness Smith's fascination with his data plots, which show disparate sets of points on a page in a manner that obscures their complex interrelationships, illustrates the power of such incorrect analysis to mislead.

Second, the data witness Smith attempts to represent in his bivariate plots are, in fact, *multivariate*. Even if, against all the evidence on this record, we were to assume away any individual facility effects – as one would have to do in order to prefer the pooled

[1] Tr. 28/15841.

model[2] – witness Smith admits that other explanatory variables besides TPH belong in the model.[3] Why, then, should he believe that it is possible to ascertain the value of the elasticity of hours with respect to TPH while restricting his attention to just this one explanatory variable? In fact, this type of partial graphical analysis suffers from the same infirmities as the analogous regression analysis in which all but one of the relevant explanatory variables have been excluded: any inferences about the elasticity of the dependent variable with respect to this one included variable will be biased and inconsistent.[4] The *only* conditions under which the relationship between hours and TPH could be assessed in this fashion without an automatic presumption of bias would be (i) if the other variables in witness Bradley's model had no effect on hours – which is tantamount to saying that they don't belong in the model – or (ii) if each variable witness Smith excludes from consideration in his bivariate graphs is completely uncorrelated with TPH.[5] Neither condition holds in this case: the other regressors in the model are highly significant, and clearly belong in the model; and not one is uncorrelated with TPH in the sample. Hence, the inferences witness Smith draws from his bivariate data plots concerning the volume variabilities of mail processing costs fail the test of reasonableness, since they are presumptively biased.

Finally, even if witness Smith's analysis of his graphs were not fatally flawed for the reasons already stated, his analysis would still fail because it is based on plots in which most of the data he claims to have examined do not actually appear. In all but one plot, in fact, the majority of the data points are hidden from view (in all but two, over 70 percent are hidden), as Table 1 shows.

[2] This assumption would, of course, be groundless since it has been rejected in multiple statistical hypothesis tests.
[3] This is implicit in Dr. Smith's recommendation that the Commission accept Bradley's pooled model, which he made at numerous points throughout his direct testimony, *e.g.*, Tr. 28/15839, 15841, 15843-4, 15846-7.
[4] For a discussion of the bias and inconsistency caused by omitting relevant variables, see Jan Kmenta, *Elements of Econometrics,* Macmillan, 2nd edition 1986, at 443-44.
[5] *Ibid.* See also William H. Greene, *Econometric Analysis*, Macmillan 1990, at 259-61, and Arthur S. Goldberger, *A Course in Econometrics,* Harvard University Press 1991, at 183-85.

Table 1 Hidden Observations in Smith's Exhibit 602			
MODS Operation	Usable Observations	Hidden Observations	Percent Hidden
OCR	18,497	15,131	81.8
BCS	22,737	18,818	82.8
LSM	23,919	17,382	72.7
Manual Letters	24,781	20,872	84.2
Manual Flats	23,989	19,918	83.0
SPBS Non-Priority	4,569	2,522	55.2
Manual Priority	15,736	7,604	48.3

The reason that the majority of the data points in witness Smith's graphs are hidden from view is that they are obscured by other points – piled one atop the other, as it were. SAS®, the statistical software he used to produce these plots, warns the user when this occurs by using different letters to represent different numbers of coincident points in the plot ("A" for one point, "B" for two points, and so on), and by printing a warning beneath each plot listing the number of hidden points.[6] Witness Smith apparently believes that this phenomenon of data "bunching" has no bearing on his ability to interpret visually the patterns observed in his data. In response to a written interrogatory asking him to confirm that he "could not visually inspect the pattern of 18,818 data points [on page 3 of 8 in OCA 602] because they are hidden and do not appear on the plot," witness Smith wrote:

> I confirm that I inspected the pattern for the points plotted. Some of the 'Z' data plot a large number of data points located at the same point, and, accordingly, data points which are plotted on a combined basis do not plot individually. [Tr. 28/15919.]

[6] "By default, PROC PLOT uses different plotting symbols (A, B, C, and so on) to represent observations whose values coincide on a plot....[T]he output [also] includes a message telling you how many observations are hidden." *SAS® Procedures Guide Version 6,* Third Edition, SAS Institute, Inc., Cary, NC, 1990 at 416.

Further, he admitted that each "Z" in the plot "indicates that 26 or more points are present in the vicinity of the letter" and that "the letter 'Z' appears in [this one] plot about 100 times."[7] *Ibid.*

Witness Smith seems unaware that the position and orientation of the least-squares regression surface describing these data (and hence its slope with respect to TPH) is *influenced not merely by the location of the points, but also by the number in any one* locale. A gravitational metaphor is apt here: as with gravity, the force of attraction exerted on the line, pulling it in a given direction is proportional to the "mass" of the data located near that spot. Thus, other factors being equal, the greater the number of hidden points at a spot on the graph, the greater the influence of that spot.
The graphs included in exhibit OCA 603 are plots of hours against TPH (in levels rather than logarithms this time) for selected facilities in each of four direct mail processing activities at MODS facilities – Manual Letters, Manual Flats, OCR, and LSM. Witness Smith claims to be able to determine the model specification, if any, with which the data *for a particular operation at a given site are most consistent:*

> A plotting of data points which ultimately has a positive intercept on the dependent variable, the hours-axis, is consistent with witness Bradley's fixed effects conclusions. A plotting of data points which result in a blob of data is not indicative that the fixed effects (or any other approach) is consistent with witness Bradley's conclusions. Finally, a plotting of data points essentially through the origin is consistent with the pooled case. [Tr. 28/15878.]

Further:

> For each of the four types of activities presented....I selected representative graphs....The three types of plots by location include, a plot that is in good agreement with a fixed effects regression; a 'blob' type of plot, indicating that for the location under consideration there does not appear to be a clear data relationship; and a plot that is in good

[7] In fact, since 18,818 observations are hidden from view in this plot, each "Z" would have to represent approximately 188 hidden points, on average.

agreement with a pooled effects regression such as presented, but not endorsed, by witness Bradley in response to Presiding Officer's Information Request #4. [Tr. 28/15879.]

I do not believe this is correct. Panel data need not have any particular implied intercept values (or signs) in order to reject the restriction of the pooled (or for that matter, the "between") model: they need only exhibit intercepts that vary by site, which is best determined by performing a statistical test of the restriction. Nor does a plot in which most of the points seem to lie along a well-defined curve or line emanating from the origin necessarily imply that the pooled model is appropriate: this is an empirical question, and depends on whether or not the data support the restriction of a common intercept. If they do not, then the plot is not "compelling," but merely misleading.[8] Finally, a plot that looks like a "blob" doesn't necessarily indicate a lack of consistency with either model: again, without taking into account the nature of the cross-section and time-series components in the data, a "blob" could be represented best by a single regression surface or by several regression surfaces, a question best left to a formal statistical test. That witness Smith thinks otherwise is an indication of his misunderstanding of panel data. By plotting his data without regard for the obvious fact that each point can be classified simultaneously in two dimensions – by its time period and its facility – he lost whatever ability he might have had to contribute a meaningful graphical analysis to these proceedings.[9]

---

[8] This point is aptly made in Figure 1.1 of Cheng Hsiao, *Analysis of Panel Data.* Econometric Society Monograph No. 11. Cambridge University press, 1986, p. 7.
[9] In principle, it is conceivable that a plot of panel data could be constructed that would allow a researcher to perform an analysis similar to the one Smith tried to do – for instance, one could imagine identifying cross-section units by color and time-period by using distinct symbols in the plots. In practice, however, this would be unlikely to be useful, because of the large number of facilities and time periods in witness Bradley's data set.

B. Witness Smith Confuses Theoretical with Applied Statistics

I believe that Witness Smith believes that Professor Bradley's analysis of mail processing labor costs is flawed because it did not begin by specifying the production function of the Postal Service. According to witness Smith,

> Economic theory uses production functions in specifying cost functions. Economists specify production functions as representing the relationship between the inputs to the production process (*i.e.*, labor, capital, etc.) and the outputs (*i.e.*, the product)....Cost functions are derived from the theory of production functions. [Tr. 28/15826.]

Further,

> [s]ince witness Bradley's cost equations for each activity are not fully derived and justified in terms of economic theory, the cost equations may provide a good data fit on an operational basis at a given facility. Nevertheless, the equations generally lack explanatory power for the purpose of cost allocation. [Tr. 28/15828.]

I believe that witness Smith is incorrect in this assertion: I am aware of no requirement that an empirical analysis of cost variability need begin by explicitly specifying or estimating the production function.

It is true that the theoretical development of the cost function entails minimizing the costs of producing a quantity of output subject to the limitations imposed by the current technological possibilities, as represented by a production function. But this is largely a textbook exercise[10], used to illustrate the relationship between the firm's cost function and the underlying production function, and has very little bearing on applied studies such as that undertaken by Professor Bradley.

[10] Indeed, one can find reference to this development in most graduate economics textbooks concerned with the theory of the firm. Two of the most commonly used are Hal R. Varian, *Microeconomic Analysis* (3rd edition, W. W. Norton & Co. 1992), and David M. Kreps, *A Course in Microeconomic Theory* (Princeton University Press 1990).

It is, however, worth noting one of the results of this theoretical development in light of witness Smith's criticism. One of the key elements of the theory of the firm is that there is, in general, no reason to prefer a production function approach to one that begins directly from the firm's cost function, because the latter "summarizes all of the economically relevant aspects of its technology."[11] In other words, the production function cannot provide any information above and beyond what can be obtained from the cost function.

Conversely, there are a number of reasons why a cost function approach is preferable. The most obvious, in this case, is a practical issue: a production function assumes the firm has just one output.[12] This clearly is not true of the Postal Service, which offers dozens of separate subclasses and rate categories of mail delivery service, as well as numerous other products and services, for sale to the public.

Even ignoring this obvious impediment, witness Bradley was wise to estimate a cost function directly for two additional reasons. First, estimating a production function would have taken him far afield of his task. Witness Smith apparently believes that encyclopedic knowledge of "the relationship between the inputs to the production process (*i.e.*, labor, capital, *etc.*) and the output... including capital/labor tradeoffs, expansion paths, and economies of scale" – in short, of the precise form and parameter values of the *production function* – *is a prerequisite for obtaining reliable estimates of* volume variability. In fact, volume variability entails knowledge only of the much more limited concept of scale economies, which can be adequately estimated without knowing the precise specification of the production function.

Second, production function estimation is much more burdensome than cost function estimation. This is largely because it requires the analyst to assemble accurate observations not only on labor and output, but on capital services as well. This

[11] Varian, *op. cit.*, at 84.
[12] Kreps, *op. cit.*, at 238.

requirement raises a host of intractable measurement problems, all of which revolve about the fact that capital goods consist not of flows of (relatively) homogeneous factor services, but of a heterogeneous stock of widely disparate real assets – buildings, machines and tools, vehicles, storage areas – possessing different prices and vintages, different uses, and different expected lifetimes. There is neither an easily obtainable "flow" measure of capital services corresponding to the hours of labor input, nor an easily obtainable marginal value measure for capital services corresponding to the wage.[13] In summary, this is a swamp that is best avoided:

> There are basically two ways to approach [the estimation of factor-demand and cost functions]. One...is to estimate, by some procedure, the underlying production function for some activity and to then calculate, by inverting the implied first-order relations, the factor-demand curves (holding output constant). The cost function can then be calculated also. <u>This, however, is a very arduous procedure</u>....It would seem to make more sense to start with estimating the cost function or the factor-demand curves directly.[14] [Emphasis added.]

Finally, witness Smith does not seem to understand the advantages of using a flexible functional form when estimating the cost function. Unless the researcher knows *a priori* the precise form of the production function – and I know of no one who would claim to possess this information in the case of the Postal Service – specifying the production function in advance of performing a cost analysis at best would serve no useful purpose, and at worst could tempt the researcher to impose unwarranted restrictions on the cost function via his choice of model specification for the production function. This is due to the "primal-dual" relationship between the cost and production functions.[15] Each specific form of the production function implies a specific functional form for the corresponding cost function that is its dual. Clearly, if one knew the specific form of the

---

[13] See, *e.g.*, William J. Baumol, *Economic Theory and Operations Analysis* (4th edition, Prentice-Hall 1977), chapters 25 and 26.

[14] Eugene Silberberg, *The Structure of Economics: A Mathematical Analysis* (2nd edition, 1990) at 281-5.

[15] For a discussion of duality in economics, see Silberberg, *op. cit.*, chapter 7, and Baumol, *op. cit.*, chapter 14.

primal (*i.e.*, production) function, then imposing this prior information when estimating the cost function would improve the resulting estimates. For example, if we somehow knew that the production function were of the Cobb-Douglas form with constant returns to scale, then one can show that the cost function obtained by minimizing costs subject to this production function would be of the same form – a useful piece of information.[16] In the absence of such specific information, however, there is little to be gained by going through this exercise. Either one could specify the production function as a flexible form, in which case the information obtained would be equivalent to what one would obtain by fitting a flexible form of the cost function directly, or one could impose a more restrictive form onto the production function arbitrarily, which would bias the resulting estimates.

C. Witness Smith is Wrong to Claim that Witness Bradley's Analysis Excludes Capital

Witness Smith alleges that total pieces handled is not "the <u>only</u> or even the <u>major</u> driver" of labor costs in mail processing (emphasis in original). Tr. 28/15825. He goes on to list "the types and age of equipment, arrangement of the production process, product demand, and types of processing activities" as the additional cost drivers that, he believes, witness Bradley should have used in his model. *Ibid.*

This is an odd assertion for someone who criticizes the Postal Service's variability witness in this case for failing to adhere sufficiently to the orthodox economic theory of the firm.[17] According to this theory, a fully-specified cost function contains two general types of explanatory variables: input prices, and the level of output.[18] If input prices are

[16] Varian, *op. cit.*, at 87. Of course, if one knew this it would still be unnecessary to begin one's analysis with the production function as witness Smith claims: one could simply impose the restriction directly when specifying the cost function.
[17] *E.g.*, Tr. 28/15822, 15823, 15825, 15826.
[18] Varian, *op. cit.*, at 49-77; Kreps, *op. cit.*, at 250-58.

assumed to be constant, then the cost function has but one explanatory variable – the level of output.

But if a theoretically correct cost function need not include direct measures of the stocks and vintages of various types of capital goods, it is nonetheless important that it include indicators of the effects of technological change on costs. This is typically done by including time trends in the model, on the assumption that technical advances occur over time.[19] Witness Bradley's model goes beyond this simple approach in two respects: he addresses the issue of automation directly by his inclusion of the manual ratio in the letter and flat sorting models, and he accommodates the uneven spatial distribution of technical advances by allowing for site-specific intercepts.

D. Witnesses Neels and Smith are Confused And At Odds with One Another on the Issue of Length-of-Run. Each Has Ignored Results Reported by Bradley that Answer Their Criticisms in This Regard.

Neither witness Neels nor witness Smith is precise in defining "long-run" and "short-run", either in theoretical or econometric terms. As it happens, these witnesses also disagree with one another on this important point. Worse still, each of them is internally inconsistent: as I show in this section, their theoretical definitions are at odds with their respective econometric specifications. Finally, each of their criticisms

[19] This is not to say that time trends necessarily capture <u>only</u> the effects of changes in technology, as witnesses Neels and Smith appear, mistakenly, to believe (*e.g.*, Tr. 28/15620, Tr. 28/15831). As witness Bradley stated in his response to DMA/USPS-T14-24, trend variables capture not only the effects of technological changes over time, but also the effects of any other covariate of mail processing labor hours that is correlated with time but not represented elsewhere in the model. It is therefore futile to attack Bradley's results on the ground that the trend coefficients don't conform to one's prior expectations concerning the impact of technological change, as Neels does (Tr. 28/15621-25), since the trends may be capturing other (possibly unknown) effects that vary with time. Virtually all cost models estimated with time-series or panel data include trends, for the very good reason that whatever is causing costs to vary over time should not be allowed to contaminate the parameter estimates of interest, in this case those associated with total piece-handlings.

concerning length of run has already been answered effectively by witness Bradley in his direct testimony. Thus, their testimony on the subject should be rejected.

Witness Neels claims that:

> [t]he fixed effects models that Bradley relies upon for his variability estimates do not appear to be capable of providing reliable estimates of the long-run variability of mail processing labor costs. These models relate mail processing labor hours in a four-week accounting period to the number of piece handlings in that same period and in the previous period. <u>Because these models look back only a single accounting period</u>, they are not capable of detecting or accounting for changes that take place over longer periods of time. Their short-run view of labor cost variability calls into question their relevance to this proceeding. [Tr. 28/15625] (emphasis added).

Continuing in this vein, he states that

> [t]he extent to which mail processing labor costs vary with volume will depend upon the time horizon over which volumes and costs change....Thus, the estimate one gets for the volume variability of mail processing labor costs may differ, depending upon how long a time is allowed for costs to respond to changes in volume. [Tr. 28/15625-26.]

Witness Neels is arguing that it is the <u>time horizon</u> of the model that determines whether it is "short-run" or "long-run." Note that he does not, in either excerpt, provide an indication of the length of the time horizon that he has in mind. He goes on to note, however, with apparent approval, that

> *"[i]n past proceedings, the Commission has relied upon evidence of the long-run variability of costs in its findings regarding the attribution of costs. 'Long-run' in this context, has been interpreted as changes that occur over periods longer than a year."* *Ibid.* (footnote omitted).

Thus, it would seem that we have our answer: witness Neels believes that witness Bradley's model is flawed because the latter includes only a single-period lag in the set of explanatory variables, rather than including more lags, or lags at a distance of more than one accounting period from the one in which current labor hours are measured.

The remedy for this supposed infirmity, were we to accept it at face value, is clear: include a lag (or lags) of duration longer than a single accounting period. It is therefore surprising that this is not what he recommends. He instead proposes an altogether different model, the "between" model estimated by witness Bradley as a by-product of one of his statistical tests.[20] *Ibid.* Witness Neels claims, without substantiation, that the between model "de-emphasizes the effects of short-term increases and decreases in volume" (Tr. 28/15627), "emphasizes the contrast between facilities that differ systematically in the volume of mail they process" (Tr. 28/15628), and is "less subject to attenuation due to errors-in-variables bias than [the] fixed effects model" (Tr. 28/15629). In fact, the between model does not merely "de-emphasize" time-varying effects and "emphasize" cross-section effects – it ignores all information in the data <u>except</u> "the contrast[s] between facilities". This is one of the reasons this estimator is biased: it excludes the (very significant) non-volume time-varying effects that are clearly expressed in the data when a model accommodating such effects is used. Whether cross-section data are less subject to the attenuation problem caused by measurement error, as witness Neels asserts, is an untested hypothesis. He offers an explanation for why this might be true, based on the metaphor of measurement errors "averaging out" in going from the full data set to facility-level means, but no other evidence in support of his assertion. In any case, the relevance of this argument depends in large measure on the seriousness one ascribes to the errors-in-variables problem in total piece-handlings. As I show in section III of my testimony, witness Neels has greatly exaggerated this problem.

Witness Neels then observes that

> [t]he volume variabilities implied by the cross-sectional models are often higher than those reported by Bradley and are generally very close to 100

---

[20] The between model is a cross-section model estimated on a single data point for each facility consisting of the arithmetic mean of each variable for each facility.

> percent (or greater than 100 percent, implying diseconomies of scale). The differences between the...results [of the between model] and the fixed effect results can be attributed to the fact that the cross-sectional results are closer to the long-run volume variabilities and are less subject to attenuation effects caused by measurement error in the piece-handlings variables. [Tr. 28/15629.]

Witness Neels is incorrect. The between model, as well as the pooled model, is biased and inconsistent. This is so because, like the pooled model of which witness Smith is so enamored, the between model imposes the unrealistic restriction of common slopes and common intercepts across all mail processing facilities. This restriction was, of course, tested and thoroughly discredited in the statistical tests performed by witnesses Bradley and myself in response to Notice of Inquiry No. 4. Tr. 28/16070-101 and T5. 29/16121-140. It is this bias, and not any supposed "long-run" qualities, that explains the exceptionally high variability estimates produced by the between model.

While witness Neels did not recommend, or provide for the record, results from a model that comported with his suggestion of a model with a longer time horizon, witness Bradley did so in his direct testimony. Although it would not have been practical to include 12 or more separate lagged piece-handling terms in order to comply with witness Neels's demand for a model with a time horizon "longer than a year"[21],. Professor Bradley did re-estimate his fixed-effects model using "same-period-last-year" (SPLY) data. This model tests the hypothesis "that the determinant of staffing for mail processing activity in a given accounting period is the amount of volume growth over the same period in the previous year." He found that:

> the results from estimation on the SPLY data confirm the general result [that] the variabilities are less than one and repeat the pattern that the variabilities for manual activities are below variabilities for mechanized and automated activities. The estimated variabilities are quite low, however. [USPS-T-14 at 77-78.]

[21] Since piece-handlings are highly correlated from one AP to the next, including more than one or two would lead to intractable multicollinearity problems.

Thus, it would seem that witness Neels's concern over the supposed shortness of the *time horizon was for naught: if anything, including lags at greater distance from the* current accounting period seems to lower, not raise, variabilities.

In contrast to witness Neels, witness Smith takes a different view on the length-of-run question. He complains that

> [t]he time period under analysis for the cost function estimation is not adequately defined for [witness Bradley's] cost equation. The data span at least 39 time periods; however, most of witness Bradley's comments and analysis suggest that he is looking at essentially "monthly" or, more precisely, four-week periods. Given the short-run four week time frames he nevertheless intermingles short-run and longer-run considerations. [Tr. 28/15835-36.]

Not content with just the two options of "short" and "long" runs, witness Smith has introduced yet a third concept – "longer-run" – without giving a definition of what he means. With some diligence, however, we can infer that what troubles witness Smith is the high frequency of witness Bradley's data – the fact that the data come to us in four-week "frames" or periods, rather than in longer increments. I therefore am puzzled that witness Smith appears to have ignored witness Bradley's inclusion of results from the re-estimation of his regression equations using annual data. The results are instructive:

> The results based upon the annual data generally support the results from the AP data in the sense of replicating the pattern and magnitude of the estimated variabilities. The annual results are not preferred, however, because they are based upon substantially less data than the accounting period data and thus do not embody an effective way to capture non-volume time-related effects. [USPS-T-14 at 75-77.]

As we found with witness Neels's concern over the time horizon, it appears witness Smith was worried for naught over the problem of the 4-week time-frame: it makes little real difference to the estimated variabilities.

In concluding this section, I should like to be clear on two points, one statistical, the other theoretical. From a statistical standpoint, neither witness Neels nor witness Smith has proposed an alternative model that comports with his putative criticism of the length of run of the witness Bradley's model. Both sought to portray witness Bradley's reliance on a fixed-effects model as somehow violating what, they believed, was the proper length of run for mail processing cost models. In fact, the fixed-effects specification does not preclude a model from being either "short-run" or "long-run." A fixed-effects model merely affords each site its own intercept, which can be separately identified so long as at least two distinct data points are available on each facility. The separate-intercepts model is thus one of the models to consider when estimating a cost relationship with panel data. One may decide to reject it, upon application of the appropriate statistical tests, if the results indicates either that time-invariant facility-specific effects are not significant, or that a less restrictive model is justified.[22] It is, however, perfectly consistent with either a short-run or a long-run cost function. Failing to account for inter-facility heterogeneity in the presence of significant facility-specific effects, on the other hand, is clearly unreasonable, since doing so yields biased, inconsistent variability estimates.[23]

From a theoretical standpoint, there is no reason for confusion concerning long-run and short-run costs. Both concepts are well understood, and are included as part of the common curriculum in economics at the introductory undergraduate level. In economics, calendar time is <u>not</u> what determines length of run. Rather, length of run has to do with which inputs are variable and which are fixed. "Long-run" refers to a period of time that is sufficiently long that all factors of production – including structures as well as machinery – are freely variable. "Short-run" refers to any length of time shorter than that, so that at least one factor is fixed.[24]

---

[22] As I show in response to NOI-4 Tr. 29/16122-7, the first hypothesis is clearly false, whereas the latter has some validity.
[23] See, *e.g.*, Hsiao, *op. cit.*, at 5-8.
[24] *Cf.* Tr. 11/5523 *et seq.*

Thus, while there is only one long-run in any given context, there is no unique short-run, since different types of capital require widely differing periods of time to be purchased and installed, or altered in any significant way. (In the Postal Service context it might require, say, up to a year to upgrade or replace an OCR in a given facility, including the time required for planning and budgeting, installation, and troubleshooting, while replacing or significantly altering a building would presumably take far longer.) Since there is no unique definition of short-run, it is therefore all the more important to be precise when using these terms to stipulate what one means.

In the context of Postal rate-making, the appropriate length of run to consider is not a mystery: it is the period of time during which the proposed rates are expected to be in effect. This point was made quite succinctly by Professor Panzar on this record when he said that rates should be based "on the marginal costs that will actually be incurred...to serve a sustained increase in volume over the time period during which the prices will be in effect." Tr. 11/5417-8.[25]

Witness Bradley's variability estimates meet this criterion. Empirically estimable cost functions embody length of run by the manner in which they are specified. As the dual to the production function, a fully-specified long-run cost function includes the relative prices (or "rental rates" in the case of capital) of all productive factors -- it is the inclusion in the model of factor <u>prices</u>, not factor <u>levels</u>, witness Smith's arguments to the contrary notwithstanding, that accommodates variations in factors over time.[26] Alternatively, if detailed price data are unavailable or unobservable -- as is usually the case for capital goods -- then proxies can be used to capture the impact of such changes.

---

[25] See also the testimony of William J. Baumol (USPS-T-3) on behalf of the Postal Service in the R87-1 Docket at 12.
[26] See, *e.g.*, Robert G. Chambers, *Applied Production Analysis: A Dual Approach,* Cambridge University Press, 1988.

Professor Bradley's model contains such proxies for capital "prices." Consider a general specification of a long-run cost function of the form

$$(1) \quad C_{it} = f(\alpha_i, w_{it}, y_{it})$$

where C is a measure of cost, f(.) is a general function, w is a vector of input prices, y is the output level, and α is the facility fixed effect. Assuming that the vector of input prices can be decomposed into time-varying and cross-sectional components, then we can write:

$$(2) \quad w_{it} = \mu_i + \lambda_t + \varepsilon_{it}$$

where μ_i is a time-invariant component that does not vary within facility, λ_t is the time-varying component, and ε_{it} is a white-noise disturbance term. Substituting this specification into equation (1) yields:

$$(3) \quad C = f(\alpha_i, \mu_i, \lambda_t, \varepsilon_{it}, y_{it})$$

This expression is, in effect, witness Bradley's cost equation. Note that α_i and μ_i cannot be separately identified (nor need they be), although the facility fixed-effects specification captures their joint impact, and λ_t it is witness Bradley's trend variables.

III. GIVEN THE NATURE OF THE DATA, WITNESS BRADLEY'S APPROACH TO DATA CLEANING IS APPROPRIATE. FURTHERMORE, HIS ERRORS-IN-VARIABLES ANALYSIS IS BOTH CORRECT AND INSIGHTFUL. WITNESS NEELS'S TREATMENT OF THESE ISSUES, BY CONTRAST, IS FLAWED.

Witness Neels exaggerates the severity and extent of the measurement error problem in Professor Bradley's data. He then objects to the data cleaning witness Bradley undertakes prior to his regression analysis, claiming that no data scrubbing should be performed. When viewed in light of his expressions of concern over the impact of measurement errors on the variability estimates, witness Neels's opposition to any data cleaning, no matter how careful and reasonable, is inexplicable.

Worse, however, is witness Neels's apparent misunderstanding of the errors-in-variables problem in the context of panel data. He does not seem to understand the power that panel data bring to this analysis. Witness Neels's claims – that witness Bradley's errors-in-variables analysis contains (unspecified) mathematical flaws, that measurement errors in TPH necessarily bias witness Bradley's variabilities downward, that comparing the relative magnitudes of automatic versus manual variabilities provides insight into the errors-in-variables problem – are all groundless.

In part A of this section, I demonstrate that witness Neels has exaggerated the extent of the measurement error problem. In part B, I criticize his extreme position on data cleaning. Finally, in part C, I illustrate how witness Neels has failed to grasp the errors-in-variables problem as it applies in the case of panel data.

A. Witness Neels Exaggerates the Extent of the Measurement Error Problem

Witness Neels exaggerates the extent of the measurement error problem in the MODS data. He uses the presumed presence of measurement errors in the total piece-handlings (TPH) variable for certain operations as though it infects all of witness Bradley's results equally, stating:

> The MODS piece handlings data that Bradley relies on for major portions of his analysis have been the target of considerable criticism. A recent review of measurement systems conducted by the U.S. Postal Inspection Service found large variances between the piece handlings figures contained in the MODS system and actual piece counts. These variances were attributed to a variety of different causes, including inadequate conversion factors, improper data input, and out-of-tolerance scales. The magnitudes of these variances could be substantial. [Tr. 28/15601 (footnotes omitted).]

The presence of such problems is not in dispute on this record. Indeed, witness Bradley has acknowledged their existence. Tr. 11/5369-70. What is at issue is, first, the prevalence and distribution of such errors in the data, and second, their likely impact on the variability estimates derived from these data.

Regarding the first issue, all data sets used to conduct applied statistical research have some likelihood of containing keypunch errors and similar mistakes that accumulate due, in large part, to simple human fallibility. Witness Shew is instructive on this point:

> Errors can creep into each stage of a data collection process, from observing an activity (*e.g.*, mail handling) to recording the observations, to compiling them in summary records. Once that process is complete, it is usually impossible, in effect, to reach back in time to spot mistakes that were made. That leads many researchers, myself included, to assume that any data set is likely to contain errors, some perhaps quite serious, that will remain invisible. [Tr. 28/15548.]

Despite the ubiquity of random errors, applied statistical analysis remains a useful analytical and management tool, with good reason: regression analysis, as with most of the tools employed by applied statisticians, is surprisingly robust to most of the commonly-encountered violations of assumptions about how our data are generated. In the case of measurement error, where a possibility exists of systematic bias or inconsistency, the likely consequences are generally well understood.

The specific sources of error cited by witness Neels provide clues as to their likely prevalence and distribution in the MODS data set. In the testimony cited above, witness Neels is highlighting the findings of the 1996 "National Coordination Audit of Mail Volume Measurement and Reporting System", which found an unusual prevalence of errors resulting from "inadequate conversion factors, improper data input[ting], and out-of-tolerance scales" (*ibid.*). This has mixed implications for the distribution of measurement errors in the MODS data set, as witness Bradley has noted:

> [S]everal of the report's findings are irrelevant for my analysis because much of the data set used in my analysis is not based upon FHPs [First Handling Pieces], but rather on the end-of-run data and machine counts. This is true for all automated and mechanized activities. The issues of measurement error due to inaccurate weighing and/or conversion factors is an issue only in the manual activities. [Tr.11/5369.]

Witness Neels attempts to rebut witness Bradley on this point, arguing that:

> First Handling Pieces is a part of the piece handling variable used by Professor Bradley; the MODS Manual states clearly in Section 212.2 that Total Piece Handlings is the sum of First Handling Pieces and Subsequent Handling Pieces. Even if the MODS counts of downstream handlings are totally free from the measurement problems that infect estimates of First Handling Pieces, all of the problems surrounding the measurement of First Handling Pieces are still passed forward into Bradley's analysis. [Tr. 28/15602.]

But witness Neels merely begs the question of the accuracy of the FHP counts. He appears to have ignored witness Bradley's main point in the response quoted above, namely that *both First Handling Pieces and Subsequent Handling Pieces are the result of machine counts in all activities other than the manual operations*, and are therefore substantially free of errors. Moreover, an increasing fraction of mail volume processed in the manual operations consists of rejected pieces from mechanized and automated operations, for which machine counts also exist. For these portions of the mail processed in manual operations, as well, there is no presumption of error. In sum, witness Neels has stirred up a tempest in a tea cup. While inaccurate scales and conversion factors remain a concern, they are a problem only in the manual operations, which account for only a small, and declining, fraction of the total mail volume processed by the Postal Service, and increasingly are not problematical there, either.

B. Witness Neels Is Incorrect on Data Cleaning Issues

In view of his expressed concern over the errors-in-variables problem, witness Neels's attitude towards data cleaning is inexplicable. On one hand, he expresses concern over the quality of the data, as I have already shown. On the other hand, witness Neels argues that no data cleaning is permissible, even when independent information is available that could improve the quality of the data, and hence of the estimates based on that data:

> In the absence of any external validity checks, it is hard to find a clear and objective basis for deciding which data to use and which data to discard. For this reason, as described above, the best approach is to dispense with all of the "scrubbing" and run the analyses on the full set of data. [Tr. 28/15632.]

This position does not comport with current practice in applied statistics. Data scrubs that improve the quality of the data by eliminating influential outliers that are believed to be contaminated with gross measurement error improve the properties of the estimates derived therefrom, rather than biasing them.

Witness Neels is correct that outlier observations in the extreme tails of the distribution are not necessarily caused by measurement error. Tr. 28/15612. One could, in fact, go further, and stipulate that there is no reason to suppose that most data points with measurement error necessarily reside in the extremities of the distribution, making detection of all measurement errors in the data an impossible task. It is the measurement errors in the tails of the data distribution ("outliers"), however, that tend to cause the greatest mischief when the goal of the research is to obtain reliable estimates of slope parameters.[27] Therefore, it is incumbent upon the researcher to focus attention on possible data errors in the outliers, and to correct or eliminate them where possible.

Witness Bradley's data scrubs were not simple-minded, as witness Neels appears to believe. His first "scrub" merely eliminates observations with missing values from the

---

[27] See, *e.g.*, David A. Belsley, Edwin Kuh, and Roy E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* Wiley Series in Probability and Mathematical Statistics, 1980, especially pp. 6-9.

data set. This is not properly termed a scrub at all, but is rather a computational necessity if econometric estimates are to be obtained. His continuity scrub, while not strictly necessary, is appropriate: it makes sense to restrict the estimates to facilities where a long run of data are available in order to obtain a data set with adequate variability to estimate the time-varying (i.e., either trending or seasonal) elements of the model. The productivity scrub is eminently reasonable, since it eliminates values that are physically impossible.[28]

A major concern when cleaning data is that the procedure not lead to greater bias than that occurring as a result of using error-ridden data. Witness Neels claims that he prefers to examine each possible data point measured with error in order to try to "understand" the process of how the error crept in (Tr. 28/15800). While this approach appears reasonable, in fact it depends on subjective judgment, inviting a results-driven data cleaning. Witness Bradley, on the other hand, uses a series of impersonal data screens so as to minimize the likelihood of introducing, perhaps unconsciously, experimenter bias into the results in this manner.

C. Witness Neels Does Not Appear to Understand the Nature of the Errors-In-Variables Problem in Panel Data

Witness Neels does not appear to appreciate the distinction between the simple errors-in-variables analysis derived from a simple model and the analysis that is applicable with panel data. In discussing his understanding of the errors-in-variables problem, witness Neels states:

> Econometric studies are especially sensitive to data errors. It is a well-established econometric principle that measurement error in an independent variable causes downward bias in coefficient estimates. This result is stated clearly in a recent text:

[28] Tr. 11/5285.

> As long as σ_u^2 [the variance of the measurement error in the independent variable] is positive, b [its estimated coefficient] is inconsistent, with a persistent bias toward zero.... The effect of biasing the coefficient toward zero is called attenuation.[29] [Tr. 28/15604-5.]

In fact, the model of measurement error being discussed in the particular passage witness Neels excerpted from Professor Greene's textbook concerns "that of a regression model with a single [badly measured] regressor and no constant term"[30], rather than a fixed-effects model with multiple regressors and multiple site-specific intercepts of the sort that witness Bradley estimated.

Professor Greene does go on to discuss the effects of measurement error in a fixed-effects model in a later section of his textbook that covers models for panel data. Here is how he frames the issue of measurement error in the introduction to that section:

> A recurrent problem in using microeconomic data is errors of measurement. As we saw [in a previous chapter], this is a thorny problem, and our conclusion there was a pessimistic one. Unless we can use some otherwise unknown parameters, the least squares estimates will be inconsistent, and little can be done to remedy the problem. Griliches and Hausman (1986) show that <u>the picture brightens considerably when panel data are used</u>. [31] [Emphasis added.]

The reason that "the picture brightens considerably" when using panel data, according to Professor Greene, is that several alternative estimators are available, notably the fixed-effects (or "within") estimator and various differenced estimators. While each of these alternative estimators, considered separately, is inconsistent in the presence of measurement errors in one or more of the explanatory variables, each one results in a

[29] The excerpted passage is from William H. Greene, *Econometric Analysis,* Third Edition, Prentice Hall, 1997, p. 437.
[30] *Ibid.*
[31] The citation in this passage is to Zvi Griliches and Jerry Hausman, "Errors in Variables in Panel Data," *Journal of Econometrics* 31 (1986), pp. 93-118. See also Cheng Hsiao, *Analysis of Panel Data*. Econometric Society Monograph No. 11. Cambridge University Press, 1986, pp. 63-65.

different sort of inconsistency. By exploiting the additional information gleaned from the differences between these estimators, its possible to identify the seriousness of the errors-in-variables problem, and recover consistent estimates of the parameters. This approach is known as the "method of moments", and is what witness Bradley implemented in his direct testimony.[32]

Since this method has already been adequately explained in witness Bradley's direct testimony I will not belabor it here, other than to point out errors witness Neels made. Witness Neels claims to have proven that witness Bradley's errors-in-variables analysis contains mathematical errors. Rather than finding them, however, witness Neels merely provides what he believes is an indirect proof by contradiction:

> In his direct testimony Professor Bradley presents the results of an analysis that, he claims, quantifies the effects of measurement error in his piece handlings variable....However, there are problems in this analysis that call into question its ability to support these claims. Bradley claims to have found upward bias in his estimate of the volume variability of the manual letter sorting activity rather than the downward bias that Greene states is the result of measurement error. As shown in Appendix A to my testimony, the formulas that Bradley himself presents in his direct testimony show clearly that upward bias is a mathematically impossible result. Bradley's finding of upward bias is therefore a sign of serious and fundamental flaws in his analysis. [Tr. 28/15608.]

There are, I believe, a number of factual errors in this statement which I will address in turn. First, the passage that witness Neels excerpted from Professor Greene's textbook discusses attenuation in an estimated <u>slope parameter</u>, not in the volume variability estimate derived from parameter estimates. Witness Neels has made an incorrect leap when he infers from Greene's discussion that attenuation in a parameter estimate implies attenuation in the variability estimate. To see why this is the case, recall that witness Bradley used the translog functional form, which includes both linear (in logs) and quadratic (in logs) terms in TPH. If there are errors in the TPH, then both the linear and quadratic regressors also contain error, and the attenuation phenomenon

[32] USPS-T-14 at 80-84.

that witness Neels discusses applies equally to both. Since the estimated slopes on these two regressors are of opposite sign, the sign of the *net* effect of any attenuation bias on the variability estimate is indeterminate. Thus, witness Neels should not claim, as he does in the above-cited passage, that witness Bradley's finding of evidence of upward bias in the manual letters variability estimate is *prima facie* evidence that witness Bradley's analysis is flawed. In fact, it is entirely consistent with the standard errors-in-variables analysis.

Second, witness Neels's claim that the analysis in his Appendix A (Tr. 28/15635-39) provides proof that witness Bradley's errors-in-variables analysis is flawed suffers from a fundamental misperception. Witness Neels claims to have shown that "the only way to arrive at such a conclusion [namely that the errors-in-variables estimator of the volume variability for manual letters is lower than either the fixed-effects or the first-difference estimator] would be for the variance of the measurement error to be negative, a mathematically impossible result."[33] In fact, Professor Bradley's result is not anomalous at all. Witness Neels's mistake was to assume that a variance estimate obtained by substituting regression point estimates derived from a finite sample into an equation that only holds exactly in the limit as the number of facilities becomes arbitrarily large necessarily would yield values that comport with theoretical variances in all cases. As is true of any finite-sample estimator, there is always a small chance that low-probability events can occur in the sample. In this case, the relative magnitudes of the within and first-difference estimators are such that a negative variance is implied. Clearly, this is not a "mathematically impossible result," since witness Neels, himself, has derived it mathematically.

Finally, it is possible to show mathematically that the degree of attenuation resulting from errors in an explanatory variable in a panel data set is inversely related to the degree of variation between cross-section units under fairly general conditions. Since the MODS data display very large between-facility variances, this could well explain the lack of evidence of attenuation in the manual sorting activities that apparently bothers witness Neels so much.

But for the sake of argument, let us examine witness Neels's point, in order to determine the maximum impact it may have on witness Bradley's volume variability estimates. Witness Neels states that

> Bradley's volume variability estimates are derived from a dataset that is the end product of an extensive editing process in which enormous amounts of data are eliminated.... [Tr. 28/15609]

Witness Neels claims that Bradley's data scrubs altered the statistical characteristics of the data so completely that his variability estimates were fundamentally altered:

> The volume variability estimates derived from this reduced dataset are <u>substantially altered from those derived from the initial dataset</u>. [*Ibid.*, emphasis added.]

I examined the evidence to evaluate the validity of this claim. Table 2 compares witness Neels's estimated volume variabilities based on witness Bradley's scrubbed data with those he derived using Bradley's methodology applied to the unscrubbed data. Witness Neels's recommended ("between") model estimates are also included for the sake of comparison. (All three sets of estimates were take from PRC/UPS XE 2, Tr. 28/15781.)

[33] Tr. 28/15637. Neels goes on to admit that he does not know why Bradley has obtained the results that he did – "The reasons for these anomalous results are not completely clear...." *Ibid.*

Table 2 Volume Variability by Operation: Comparison of Effects of Scrubbing and Misspecification					
Operation	Neels Scrubbed (1)	Neels Unscrubbed (2)	Neels Between (3)	Scrubbing Effect (2) - (1)	Specification Error (3) - (1)
Manual Letters	80	84	125	4	45
Manual Flats	87	90	131	3	44
OCR	79	83	121	4	42
BCS	95	106	132	11	37
LSM	91	97	121	6	30
FSM	92	102	116	10	24

Source: PRC/UPS XE 2 (Tr. 28/15785).

As the numbers in Table 2 clearly show, witness Neels's claim is without merit. Despite his inclusion of observations that "clearly contain some cases beyond what is considered to be physically possible" (Tr. 11/5285), the impact of witness Neels's elimination of the Bradley data scrubs on the estimated variabilities is relatively small – in all but two cases in the range of 3 to 6 percentage points, and never by more than 11. The average change was a bit over 6 percentage points.

By contrast, the estimates produced by witness Neels's preferred ("between") model cause much more dramatic shifts in the variability, ranging from 24 to 45 percentage points difference. The average change was 37 percentage points. Comparing the two Neels estimates to witness Bradley's results provides perspective on the relative importance of measurement errors and specification error in causing potential bias in the variability estimates. It is clear that the specification errors caused by imposing the common slopes/common intercept assumption on the data is a far more serious problem.

IV. WITNESS BRADLEY APPLIED STATE-OF-THE-ART ECONOMETRIC METHODS IN DERIVING HIS VARIABILITY ESTIMATES

Contrary to the opinions of witnesses Neels and Smith, Professor Bradley applied a state-of-the-art econometric analysis to the problem of empirically estimating the volume variability of mail processing costs. Notable features of his analysis include his use of an unusually rich panel data set that captures both the cross-sectional variation in the productivity relationship among individual facilities, as well as the time-varying components; using a flexible functional form that allows the estimated regression to approximate any functional form indicated by observed patterns in the data, rather than imposing one arbitrarily; correcting for the effects of serial correlation; and allowing for time-varying effects through the use of seasonal dummies, trend variables, and a dynamic structure.

Witness Bradley also tested the major assumptions underlying his model. Rather than arbitrarily selecting a model and assuming that it met the criteria of the particular problem at hand, he used the data to test for (and confirm the presence of) individual facility effects, serially correlated residuals, and lagged adjustment of a facility's labor force to changes in volume. He also provided results from a number of alternative models to indicate the robustness of his estimates to alternative assumptions and measurement errors.

Because Professor Bradley did a good job of explicating his econometric methodology in his direct testimony, there is relatively little that needs to be added. However, I will address three areas in this section of my testimony that, I feel, should be emphasized in this record. In part A of this section, I discuss the added power that panel data can bring to an empirical cost analysis, and point out how witness Bradley took advantage of this power to enhance his analysis. In part B, I discuss the specification testing that Bradley undertook during the model design phase of his analysis. In part C, I discuss his choice of functional form.

A. Witness Bradley Exploited the Added Power Afforded by Panel Data

The availability of panel data from the MODS and PIRS data sets on this record has meant that detailed information was available on both the cross-section variation among different mail processing facilities at a point in time, and the changes over time within individual facilities. It is thus much more informative than either pure time series or pure cross-section data, permitting the analyst to distinguish between purely localized factors affecting the relationship between costs and volume in a facility and those that characterize the Postal mail processing system as a whole. It also provides a powerful antidote to the often intractable problem of measurement error in plant-level data. Witness Bradley's approach took advantage of this power, as I have already shown in section III of this testimony.

B. Witness Bradley Performed Numerous Tests of his Model's Specification

Witness Bradley performed numerous statistical tests during the model specification stage of his research to guide his choices of model and estimation technique. The first question he addressed was whether there was evidence of significant time-invariant individual facility effects. These would be substantial differences among sites in the average labor productivity of a given operation due to intrinsic differences among facilities, including "the age of the facility, the quality of the local labor force, and the quality of the mail that the facility must process." USPS-T-14 at 39-40. Bradley performed what he termed a "Gauss-Newton regression (GNR)" test for individual facility effects. *Ibid.* at 41-43. More commonly termed a Lagrange multiplier (or "LM") test in the econometrics literature, it involves the estimation of the restricted (in this case, pooled) model to obtain the residuals, which are subsequently analyzed for evidence of misspecification.[34] The hypothesis of no individual facility effects was strongly rejected by this test.[35]

Having identified the presence of significant facility effects, Professor Bradley next faced the problem of how best to accommodate them in his model. As he noted, there are two basic choices: the random-effects and fixed-effects specifications. Because he was working with data that were not a random sample, and the intended use of the model was to make inferences that would be applied primarily to within-sample facilities, he noted that there was some *a priori* justification for using the fixed-effects specification. USPS-T-14 at 44. However, the random-effects model has the advantage of greater efficiency because fewer parameters are estimated.[36] The main danger of using random-effects is that the individual effects may be correlated with the included explanatory variables, which, if true, would imply that the random-effects estimator is inconsistent.[37] Witness Bradley performed a Hausman test – the standard statistical method for detecting the presence of correlation between the individual effects and the included RHS variables.[38] The random-effects model was decisively rejected. USPS-T-14 at 45-6.

Witness Bradley next considered the possibility that, because of the high frequency and long duration of his data, his model might need to accommodate serial correlation of the disturbances. He performed a Durbin-Watson test, modified to allow for the fixed-effects specification, and found strong evidence of autocorrelation. USPS-T-14 at 48-9. Failure to account for this in his model, while implying no bias, would have

---

[34] Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, 1993, ch. 3.
[35] This finding was later confirmed by *F* tests conducted by witness Bradley and myself in response to NOI No. 4. Tr. 28/16071-94 and Tr. 29/16121-40.
[36] Greene, *op. cit.*, p. 495.
[37] *Ibid.*
[38] J. Hausman, "Specification Tests in Econometrics," *Econometrica* 46 (1978), pp. 69-85; J. Hausman and W. Taylor, "Panel Data and Unobservable Individual Effects," *Econometrica* 49 (1981), pp. 1377-98.

implied a relatively inefficient estimator and strictly invalid inferences.[39] Thus, witness Bradley chose to correct for autocorrelated disturbances. USPS-T-14 at 49-51.

Finally, while not part of witness Bradley's formal specification testing, I would note that he performed a number of informal sensitivity tests to assess the robustness of his chosen specification. In this category I would include the fixed-effects model without serial correlation correction (USPS-T-14 at 70-1), which indicated that the results of his preferred model did not depend on this correction; the two-way classification model (*ibid.* at 72-4), an alternative specification of time-varying effects to illustrate that the inclusion of time trends was not driving his main results; the model estimated using annual data, which showed that his results did not depend on the use of high frequency data (*ibid.* at 75-7); and the model estimated using SPLY data, which illustrated that his results did not change dramatically with the inclusion of a lagged TPH variable at greater remove from the current period (*ibid.* at 77-9). These were informal assessments involving judgment, rather than formal statistical test procedures. Taken as a whole, they indicate that witness Bradley's preferred model is reasonably robust, and confirm the general conclusion that volume variabilities in mail processing are generally well under 100 percent.

C. Witness Bradley's Functional Form Is Appropriate

Witness Bradley chose to specify a transcendental logarithmic ("translog") functional form for his cost functions. USPS-T-14 at 35-38. He states that he did so because he had no "prior operational knowledge" to guide him to a specific functional form for the cost or production function. *Ibid.* This admission of seeming ignorance actually represents the current state of the econometric art for empirical cost functions. A flexible function form avoids imposing unjustified restrictions on the parameters of the underlying technology through the choice of functional form, by instead approximating the true, but unknown, cost function with a specification containing enough parameters

---

[39] Greene, op. cit., pp. 436-9.

to provide a reasonable approximation to whatever the true function might be.[40] The translog, in particular, is a generalization of the Cobb-Douglas and similar functions that restrict factor substitution elasticities to be equal everywhere.[41]

In spite of its common use in the econometrics literature, Professor Bradley's use of the translog form caused confusion on the part of some parties. Witness Smith, as I have already mentioned, appears to believe that intimate knowledge of the specific "capital/labor tradeoffs, expansion paths, and economies of scale" (Tr. 28/15826) is readily available. On the contrary, it rarely if ever is and, when prior knowledge is not available, specifying and estimating a translog cost function provides the analyst no less information about these characteristics than would a similarly specified production function. Witness Neels apparently misinterpreted the additional parameters contained in the translog functional form and, as a result, in his discussion of attenuation due to errors-in-variables confused the notion of attenuation in a parameter estimate and attenuation in a statistic calculated from multiple parameter estimates.

V. WITNESS NEELS ERRED ON ORAL CROSS-EXAMINATION

In his appearance before the Commission, witness Neels responded to a number of questions. I believe that some of his answers were incorrect. In part A of this section, I discuss his responses to questions concerning Cross-Examination Exhibit PRC/UPS-XE-1 entitled "Nested Sequence of Models". Tr. 28/15776. In part B, I discuss his responses to questions concerning Cross-Examination Exhibit PRC/UPS-XE-2 entitled "Comparison of Bradley and Neels Econometric Results". Tr. 28/15785.

---

[40] E. Diewert, "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *Journal of Political Economy* 79 (1971), pp. 481-507; and E. Berndt and L. Christensen, "The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing, 1929-1968" *Journal of Econometrics* 2 (1973), pp. 81-114.
[41] Greene, *op. cit.*, p. 526.

A. Answers Regarding "Nested Sequence of Models"

Witness Neels was asked a series of questions based upon a Cross-Examination Exhibit entitled "Nested Sequence of Models". He was asked about the manner in which witness Bradley had tested the random-effects model.[42] In fact, witness Bradley performed the standard statistical test – Hausman's test – to ascertain whether a random-effects model could be used; the random-effects model was strongly rejected.[43] USPS-T-14 at 43-46. Witness Neels was correct in pointing out that witness Bradley <u>also</u> made an *a priori* argument in favor of the fixed-effects model, based on the fact that the data are not a random sample, and inferences from the model were to be applied primarily to facilities within the sample. *Ibid.*

Witness Neels was then asked whether the model with facility-specific slope and intercept parameters was tested.[44] His answer contains both a factual error and a statement which I consider at best confusing. Witness Neels erred when he said that

---

[42] CHAIRMAN GLEIMAN: Did he then test and reject the random effects model against the next most restrictive model that lacks time-indexed coefficient, the fixed index, the fixed effects model?

THE WITNESS: I hesitate – what I recall of Dr. Bradley's, and I may have misspoken before, but my recollection of Dr. Bradley's testimony was that he discussed the random effects and the fixed effects as alternatives and I recall he had somewhat of an a priori argument, not a statistical argument, in favor of the fixed effects model. I don't recall what his test was between those two, whether it was statistical or whether it was theoretical. [Tr. 28/15777-78.]

[43] Neels later admitted this on oral cross-examination by the Postal Service. Tr. 28/15806.

[44] CHAIRMAN GLEIMAN: Was the next most restrictive model that lacks time-indexed coefficients, the model that allows both the slope and the intercept to vary by facility, tested to see if it is consistent with the data?

THE WITNESS: It was tested relative to the fixed effects model and I think by Higgins and Bradley against the pooled model. I don't believe it was tested against the more general model where both the slope coefficients and the intercept coefficients vary both across facilities and across time. That would be the model shown in the topmost box [in the Cross-Examination Exhibit, "Nested Sequence of Models"]. [Tr. 28/15779.]

the general model with intercept and slope parameters that vary by facility "was tested relative to the fixed effects model and I think by Higgins and Bradley against the pooled model." Tr. 28/15779. The *F* test used in NOI 4 is a test of restrictions imposed on a more general model, not a test of the more general model *per se*. In the case of the mentioned test, it was the set of restrictions embodied in the fixed-effects model that was tested, not the more general model; it would have been more correct to say that the fixed-effects and pooled models were tested relative to the more general model with all parameters varying across facilities.

Witness Neels's answer was, again, at best confusing. He implied that the model with facility-specific slope and intercept parameters could be tested against the model at the top of the Cross-Examination Exhibit, enclosed in the box labeled "Most General Model". No such test can be performed. The so-called "Most General Model" in the exhibit is not estimable, because it has a far greater number of unknown parameters than there are observations in the data set. It is therefore not likely of any practical relevance to this proceeding.

More generally, none of the models with accounting period-specific effects that appear on the right-hand side of the Cross-Examination Exhibit, "Nested Sequence of Models", is strictly relevant, either. The time and facility indexes shown in the exhibit suggest that a logical symmetry exists between these two types of effects. No such symmetry exists, for the simple reason that specifications with separate intercepts for each accounting period do not make a great deal of sense. By contrast, it is reasonable to assume the possibility of separate intercepts for each site.

Given the extremely wide range of sizes and the geographic dispersion apparent in the MODS data, there is a strong presumption that the mean level of labor hours will vary discontinuously from site to site. There is no such presumption in the case of accounting period effects. If anything it is, rather, the reverse. Recall that these effects consist of the average system-wide change in mean hours from one four-week accounting period to the next. The entire mail processing system's labor costs for mail processing would probably not move around discontinuously between one accounting period and another, apart from the seasonal effects that are already included in the model. This is simply not how non-volume time-varying effects occur, in mail processing or any other industrial process. This is why applied econometricians generally include site (or plant, or firm) effects in their cost functions, whereas they do not generally include time-varying parameters.[45]

This is not to say that such a specification is not of interest, or that no such model was considered on the record. Witness Bradley reported the results of a two-way model with time-period and facility intercepts in his direct testimony. He found that "the two-way variabilities are lower than the [fixed-effects] model and in some cases the two-way variabilities are materially lower. Nevertheless, the general patterns found in the [fixed-effects] model are confirmed." USPS-T-14 at 72-74.

Witness Neels then made a statement about the preferred order of testing.[46] On that issue, witness Neels is mistaken: it is generally accepted good practice, when testing

[45] See, *e.g.*, J. R. Norsworthy and S. L. Jang, *Empirical Measurement and Analysis of Productivity and Technological Change: Applications in High-Technology and Service Industries*. Elsevier, 1992; and Robert G. Chambers, An exception to this generalization might arise in cases where the frequency of the data were much lower than the four weeks that Bradley worked with, or if there were only a few time-periods worth of data per cross-section unit.

[46] CHAIRMAN GLEIMAN: Now, is it standard econometric practice to search for an estimation method by sequentially testing more restrictive models against less restrictive models, in other words, to go from the specific to the general?

THE WITNESS: This is an area of sort of what is considered to be good practice. I think – my sense is that one, generally, should begin with the most general and ask whether you can move to the more restrictive, because if you start with the more general, you are less likely to make a wrong turn. There are some technical reasons for starting with a more general model. You are less likely to run into a model which is subject to misspecification. So, I think – I think the counsel of perfection is probably to start with the more general and work your way in the other direction to see, you know, *nested linear restrictions of the sort that witness Bradley and I performed in response to* NOI No. 4, to begin from the most parsimonious model and work from there toward more general specifications.[47] This is consistent with generally accepted scientific methods wherein, on one hand, we have our empirical data, and on the other we have our hypotheses – some of which may be theoretical in nature, and others of which may be more speculative. The general goal is to reconcile our hypotheses with the available evidence, if possible, or if not, then to call into question one or more of our hypotheses. In so doing, we are guided by the objective of finding the simplest model that is consistent with the data.

B. Answers Regarding "Comparison of Bradley and Neels Econometric Results"

Witness Neels was also asked a series of questions concerning Cross-Examination Exhibit PRC/UPS-XE-2 entitled "Comparison of Bradley and Neels Econometric Results" (Tr. 28/15785). These questions concerned the state of the record with respect to the various econometric estimates of volume variability of mail processing labor costs. Witness Neels was first asked about the robustness of the volume variability estimates.[48] His response is marred by a significant mistake. Witness Neels

---

see whether imposing restrictions to get – to achieve a more parsimonious model leaves you with something that is statistically defensible. [Tr. 28/15780.]

[47] This is a reflection of a general preference for the simplest model that is consistent with the available evidence – an application of Occam's Razor. For an example which embodies this ordering of tests from most restrictive to less restrictive, see Hsiao, *op. cit.* at 12-18.

[48] CHAIRMAN GLEIMAN: Now the question is, in your opinion, are any of the econometric results shown in the table robust and stable?

THE WITNESS: Not in my opinion. I've actually said to my associates that work with me on econometric studies that a good study should be like shooting elephants. It should be a really big target and easy to hit no matter how you do it. And if differences in methodology give you pretty drastic differences in results, that is always to me a warning sign that we don't fully understand what's going on, and it's really – that's the basis for my unease with this line of analysis, and I think, you know, the information that's presented in this table to me amply demonstrates the fact that, you know, we haven't yet figured out what the relationship is between labor – mail handling labor costs and volume. [Tr. 28/15786-87.]

made a fundamental error by placing all of the specifications listed in the exhibit on the same plane: those that were shown to be completely inconsistent with the available evidence in the various statistical tests that have been performed in this proceeding should not have been included in his answer.

I am not arguing that stability and robustness should not be considered when selecting a preferred model – indeed, they are important criteria for evaluating alternative specifications of econometric models. But they are by no means the only criteria. Others that are equally important include data coherency and admissibility (including absence of autocorrelated disturbances and misspecification), parsimonious parameterization, and encompassing (or the ability to explain the characteristics of rival models).[49] The coherency and admissibility criteria clearly rule out at least two of the models specified in the exhibit – witness Bradley's estimates without correction for serial correlation, and witness Neels's "modified version of Bradley's cross-sectional (i.e., between) model.[50]

If these inadmissible models are excluded from the exhibit, then the picture changes considerably from the one witness Neels painted. Rather than arrows scattered around the target, there are instead tight patterns of arrows clustered about the bulls-eyes. I would also observe that the exhibit did not show the variability estimates for the one statistical model that was not rejected by a statistical test in this proceeding, namely the model with intercept and slope parameters that vary by facility. In the concluding section of my testimony, I include a table that shows the variability estimates from this unrestricted model using both the simple arithmetic means and the TPH-weighted

---

[49] See, *e.g.*, David F. Hendry and Jean-Francois Richard, "On the Formulation of Empirical Models in Dynamic Econometrics," *Journal of Econometrics* 20 (October 1982), pp. 3-33; and Edward E. Leamer, "Model Choice and Specification Analysis," Ch. 5 in *Handbook of Econometrics*, Vol. I, Zvi Griliches and Michael D. Intriligator (eds.), North-Holland, 1983.

means of the facility-specific estimates for each of the direct MODS mail processing activities. Note that these estimates, while in most cases somewhat lower than those produced by witness Bradley's preferred model, are nonetheless reasonably close and certainly provide additional evidence in favor of the hypothesis that mail processing labors costs are less than 100 percent variable.

Witness Neels and I definitely differ regarding econometric practice. He seems, by his "shooting at elephants" comment, to imply that econometric analysis is only valid when one obtains the same results regardless of the specification one chooses. I do not agree. Not all models are created equal – some models are "more equal than others." In particular, specifications that are clearly at odds with the available evidence – and at the risk of sounding redundant, let me emphasize that by this I mean these models that fail to account for the obvious individual facility effects that are present in the data are not relevant. They are "off the table", not part of the conversation, unworthy of consideration. From a statistical standpoint, the only models that remain standing after the responses to NOI No. 4 by witness Bradley and myself are the fixed-effects model of witness Bradley, and some form of the unrestricted model with facility-specific slopes and intercepts.

Witness Neels was then asked to comment further on the alleged "instability" of the results presented in the exhibit. He uses this opportunity to argue, once again, for the discredited cross-sectional model:

> THE WITNESS: The one that I will share my thoughts about, the distinction that I thought about the most and that's the one between Dr. Bradley's recommended results and my own, I think I said in my direct testimony that there were two – it seemed to me that there were two aspects of the cross-sectional models that I had identified as the best of the bunch, which I thought helped to explain the difference in variabilities.

---

[50] Witness Neels's "All Usable Observations" results are inadmissible as well, but for another reason: he erred in his construction of the time trend variables, causing them to count consecutively across discontinuities (gaps) in his data.

> One is the fact that in the cross-sectional model, you know, the way it was implemented you average across all the observations associated with the site, so you're constructing in a sense a composite observation that summarized what we know about volume over an extended period of time. I think – as I said in my direct testimony, that has the effect of averaging out some of the measurement error that's associated with the MODS data, and as it reduces the relative importance of measurement error, I think it eliminates some of the downward bias and variability estimates that can be attributed to that cause. [Tr. 28/15787-88.]

I have already discussed these matters in section III of my testimony. I will merely highlight the conclusions I drew there that bear directly upon what witness Neels has said here: (1) witness Neels greatly exaggerates the measurement error problem; (2) even if he had not done so, his assertion that this would bias Dr. Bradley's variabilities downward is wrong: there is no automatic presumption that the direction of asymptotic bias is downward, given the functional form witness Bradley employed; (3) he ignores witness Bradley's errors-in-variables analysis, which showed that its effects are negligible for the manual operations; (4) while compressing all of the observations on each facility into a single average point may (or may not – he offers no proof) reduce the measurement error problem, it leaves witness Neels with a far more serious problem: he is recommending a model that has already been statistically tested and rejected out of hand, namely the specification with common slope and intercept parameters for all facilities. The *F* tests I performed in response to NOI No. 4, and especially the test of the pooled versus fixed-effects specifications performed by witness Bradley in his response to NOI No. 4, leave no room for doubt on this point. It is much more plausible that it was the imposition of this restriction, rather than any alleged errors-in-variables problem, that caused the variability estimates from the between model to differ so drastically from the others on the record.

Witness Neels continued:

> I think the other thing which partly explains it is the nature of a cross-sectional model[;] it's generally held that cross-sectional analysis comes closer to giving you long-run effects, because you're comparing different types of facilities with different levels of volume. I mean, as you know, my

> cross-examination earlier today indicated there are systematic differences in volume across facilities, and you get a chance to see what the operation looks like as its adopted to those different levels of volume.
>
> I think the numbers here suggest that if you look across from smaller facilities to larger facilities you find labor hours increasing more than proportionately, and I think that may be closer to the long-run effect, although, you know, I repeat my earlier reservations about this line of approach. [Tr. 28/15788.]

Again, I have already disposed of witness Neels's conflation of "cross-section" with "long-run" in section II of my testimony. To assert that a fixed-effects model is associated with a particular length of run is simply wrong: the fixed-effects specification does not preclude a model from being either "short-run" or "long-run." It merely permits the regression to reflect time-invariant facility effects to the extent that they are present in the data. Acceptance or rejection of the model is properly determined by application of the appropriate statistical test, not by simple references to "long-run" costs. Failing to allow for inter-facility heterogeneity, on the other hand, yields biased, inconsistent variability estimates.

There are also other, much more intuitive arguments against witness Neels on this point that also may be worth considering. No matter what one thinks of the U.S. Postal Service's ability to manage its mail processing operations, it is unlikely that they would ramp up the scale of their processing facilities to an extent that they find themselves operating well beyond the point at which all scale economies have been exhausted – in other words, well into the region where unit costs are rising rapidly. And yet this is precisely what witness Neels would have us believe when he says "if you look across from smaller facilities to larger facilities you find labor hours increasing more than proportionately." Tr. 28/15788.

This, of course, is completely at odds with what we would expect by the economics of the firm. But, more tellingly, it also directly contradicts a point witness Neels made in response to another question:

CHAIRMAN GLEIMAN: Now suppose the number of facilities were increased by ten percent while the average volume at those facilities remained unchanged. Would total processing labor cost for the system as a whole increase by ten percent regardless of the mail processing variability observed at the facility level?

THE WITNESS: That's what I would expect to see happen. It's – in assuming that the new facilities look overall like the old facilities, all you're doing is replicating an identical operation at a new site, and if that's true, you would expect cost to just increase linearly with the number of facilities, or in your example with the volume. [Tr. 28/15790-91.]

Thus, it is witness Neels's estimates of volume variability that seem to be unstable, rather than the other estimates on this record: on one hand he is recommending a model that yields variability estimates well in excess of 100 percent – 125 percent for Manual Letters, 131 percent for Manual Flats, 121 percent for OCR, 132 percent for BCS, and so forth – and yet, in response to a direct question, he states that mail processing is characterized by constant returns to scale on the basis of discussion about the Postal Service "replicating" its operations.[51]

Finally, witness Neels engages in a series of responses to questions concerning how elasticities are computed from an empirically estimated cost function.[52] His responses

[51] Neels later admitted on cross-examination that he did not believe his own replication story. Tr. 28/15808-9.

[52] CHAIRMAN GLEIMAN: [Referring to Table 7 on page 54 of witness Bradley's direct testimony] Coefficient estimates involving squares or cross products are omitted from table 7 as a consequence of using mean centered data. Some coefficient estimates are in table 7 but do not enter into the calculation of elasticities that appear on the bottom line. Among these are manual ratio at facility[,] time trends – time trend 1 and time trend 2.

Now, I have some questions I want to ask you. If an estimated coefficient is not used to calculate elasticity, does it constitute an assumption that the variable is not influenced by the volume directly or indirectly?

THE WITNESS: I believe that's correct.

CHAIRMAN GLEIMAN: Is this assumption plausible for a manual ratio?

suggest a lack of understanding about the econometric estimation of cost elasticities. He should have answered that every variable included in the model enters into the calculation of the estimated volume variabilities. Witness Neels can test whether the manual ratio affects the estimated volume variability by rerunning one of witness Bradley's letter or flat operation regressions with the manual ratio excluded. I am quite certain he will find that the estimated variability will change.

Witness Neels also let stand the impression that the squared and cross-product terms do not enter into the calculation of the variabilities "as a consequence of using mean centered data." Of course, they do enter into the computation of the elasticity. The elasticity of a dependent variable of a function with respect to a marginal change in one of the independent variables of the function is approximated by the partial derivative of the function with respect to that independent variable. Given witness Bradley's choice of the translog functional form, the squared and cross-product terms do indeed enter into the calculation. The stratagem of estimating the model in deviations from means is simply an expedient that allows the researcher to obtain the elasticity directly off of the regression printout, rather than having to compute it after the fact with a calculator or pencil and paper. It has no effect on the value of the elasticity, and will give precisely the same answer as if the model had been estimated on the untransformed data and the elasticity then computed at the mean values of the data.

THE WITNESS: I'm not sure that it is. I spoke earlier about a hypothetical situation in which increases in volume could lead to a change in the manual ratio which would have an indirect – establish an indirect relationship between volume and costs that would not be captured simply by focusing on the coefficients on pieces and lagged pieces shown in table 7.

CHAIRMAN GLEIMAN: Should the coefficient of manual ratio be used in elasticity calculation given that the TPH is a determinant of manual ratio?

THE WITNESS: If TPH across activities, which would have to be the case, is a determinant of the manual ratio, then that contribution to volume variability should be taken into account. Tr. 28/15794-95.

Witness Neels was then asked a related question concerning the facility fixed effects:

> CHAIRMAN GLEIMAN: If fixed effects coefficients in the Bradley model, alpha (i), reflect differences among facilities that are indirectly influenced by volume, should the fixed effects coefficients also enter into the elasticity calculation?
>
> THE WITNESS: I think the same argument holds there. If a relationship can be established between volume and the fixed effects coefficients, then I think that indirect effect should also be incorporated into the overall estimate of the relationship between volume and cost. [Tr. 28/15796]

Once again, witness Neels's answer is not correct. The fixed-effects coefficients clearly do "enter into the elasticity calculation" – they affect the estimated variabilities quite strongly, as evidenced by the dramatic manner in which they leap upward as a result of heterogeneity bias when the individual facility effects are suppressed.

There is also a more subtle argument at play here that should also be addressed. The question could be interpreted as asking whether, to the extent that the individual facility *fixed effects are correlated with volume changes, shouldn't they be added onto witness* Bradley's elasticity estimates? If this was witness Neels's interpretation of the question, then his answer was also incorrect, for the following reason. Mail processing volume variability is concerned with response of costs to a small added increment of mail of a given type to the overall mail processing system – not, with respect to an *increment of mail entering into a specific processing facility.* So while it is likely true that the individual effect for a given facility is positively correlated with total piece-handlings <u>at that facility</u>, the correlation of the latter with respect to volume changes at the national level is approximately zero. This, in turn, implies that the correct answer to the question is "no".

VI. CONCLUSION

I have reviewed all of the evidence filed in this case pertaining to the cost elasticity of mail processing. Based both on this review and my training in econometrics, I conclude that witnesses Neels's and Smith's attacks on witness Bradley's study are unwarranted. Witness Bradley has presented a textbook example of how to perform this kind of analysis. His theoretical underpinnings are correct, his data scrubs are reasonable, and his econometric methods are proper. In fact, given his training and his years of experience in postal economics, one would and should expect nothing less. In contrast, I showed that there is generally no merit to the criticisms of witnesses Neels and Smith.

There is no doubt on this record that the cost elasticity of mail processing is substantially less than 1. The long-held unsubstantiated assumption that mail processing costs are almost totally volume variable has been shown to be invalid. There is ample evidence on this record to resolve the issue of the cost elasticity of mail processing econometrically. In this case, much to its credit, the Service tested this *previously unsubstantiated It would be wrong to conclude, based on the evidence in* this record, that while witness Bradley had made a nice start, there still remains too much uncertainty or too many unresolved issues to estimate mail processing cost elasticities using an econometric analysis. This is simply not the case: not only did witness Bradley make a nice start, he also made a nice finish. While we may debate which is the better econometric approach to estimate numeric values for the cost elasticity, clearly such an estimate should be computed and used in this case. The fact that an econometric estimate is, and perhaps always will be, imperfect should not deter us. Use of a sophisticated, state-of-the-art estimate is far superior to reliance on an invalid assumption.

I also find that there is ample evidence on this record to estimate numeric values for the cost elasticity of mail processing operations. The evidence already in this record and the evidence I present in this testimony together clearly indicate that either witness

Bradley's fixed-effects model or the unrestricted model is appropriate. Witness Bradley's model has an advantage in that it can be used to estimate a single cost elasticity for each operation but the disadvantage that the F test shows it is inferior to the unrestricted model. The unrestricted model, while superior from the perspective of the F test, has the disadvantage that there is still a need to weight the individual results to produce the requisite national elasticity estimate. Fortunately, there is no need to choose between the two models; the results they produce, if one combines the facility-specific variability estimates from the unrestricted model by using the arithmetic (unweighted) mean or the mean weighted by piece-handling variability, are remarkably consistent, as I show in Table 3 of my NOI response, reproduced here. Tr. 29/16127 (MPA-NOI-1 at 6).

Table 3
Variabilities From Witness Bradley's and Unrestricted Models

Operation	Unweighted Mean Variability	Weighted Mean Variability	Witness Bradley's Variability
Manual Letters	.511	.462	.797
Manual Flats	.562	.491	.866
OCR	.670	.736	.786
BCS	.845	.795	.945
LSM	.805	.809	.905
FSM	.733	.733	.918
SPBS Priority	.681	.667	.802
SPBS Nonpriority	.492	.472	.469
Manual Priority	.307	.371	.448
Manual Parcels	.277	.295	.395
Cancel/Mail Prep.	.358	.348	.654

In the interest of taking a reasonably conservative position, I suggest that the variabilities from the fixed-effects model, which in all but one instance are larger than either the weighted or unweighted mean variabilities from the unrestricted model, are the appropriate choice. I recommend this even while understanding that, on econometric grounds, the variability estimates from the unrestricted model may be slightly preferred.

I should also point out that both of these models are far superior to every one of the alternative models. As I describe in my testimony, there are compelling reasons to reject each of these alternatives. The pooled model is decisively rejected by the F test requested in NOI 4, and its specification of a single slope and single intercept has nothing to recommend it from a theoretical, statistical, or operational perspective. Further, as I point out in my testimony, the family of models on the right hand side of PRC/UPS Cross-Examination Exhibit No. 1 (Tr. 28/15776) has no theoretical basis. *Finally, the most general model is not an alternative since it cannot be estimated.*

CERTIFICATE OF SERVICE

I hereby certify that I have this date served the foregoing document upon all participants of record in this proceeding in accordance with section 12 of the rules of practice.

James R. Cregan

Washington, D.C.
March 9, 1998