RECEIVED
AUG 14 5 18 PM '00
POSTAL RATE COMMISSION
OFFICE OF THE SECRETARY

USPS-RT-7

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

POSTAL RATE AND FEE CHANGES, 2000	Docket No. R2000–1

REBUTTAL TESTIMONY
OF
WILLIAM H. GREENE
ON BEHALF OF THE
UNITED STATES POSTAL SERVICE

Table of Contents

I. STATEMENT OF QUALIFICATIONS

My name is William H. Greene. I am a professor of Econometrics at the Stern School of Business at New York University and, since 1995, chairman of Stern's Economics Department. I have taught at Stern since 1981. Prior to that I taught Econometrics at Cornell University from 1976 to 1981. I received Masters and Ph.D. degrees from the University of Wisconsin at Madison in 1974 and 1976, respectively. I worked briefly as an economic consultant in the private sector in 1980–1981 at National Economic Research Associates and have also provided consultation to numerous corporations, including American Express Corp, Ortho Biotech, Inc., and The Reader's Digest. I have published numerous works in econometrics, including roughly 40 articles, one of the world's most widely used computer programs for econometric computation, *LIMDEP*, and, notably for this proceeding, the widely used textbook *Econometric Analysis*, which several of the witnesses in this and the prior omnibus rate proceeding, including Neels, Smith, Bradley, Higgins, and Bozzo have all cited in their testimonies.

I do note that this is my first appearance before the Postal Rate Commission. I have no knowledge of the details of Postal Service operations or data systems beyond that contained in the testimonies that I reviewed. The scope and nature of my testimony will be limited to econometric technique and methodology, about which I have written extensively. I will discuss this further in Section II.

II. PURPOSE AND SCOPE OF MY TESTIMONY

I have been asked by the United States Postal Service, the sponsor of Dr. Bozzo's testimony, to comment on the testimonies of Kevin Neels and J. Edward Smith, both of which seek to rebut Dr. Bozzo's testimony and its predecessor by Michael Bradley in the 1997 counterpart to this proceeding. In particular, a number of issues have been raised regarding the econometric techniques used by Drs. Bradley and Bozzo to estimate 'volume variability factors' for labor costs in mail processing. (Volume variability is a measure of the elasticity of mail processing costs with respect to volume.)

I have learned through my reading of the various testimonies that I have reviewed that the Postal Rate Commission has traditionally assumed that this cost elasticity is 1, or 100 percent. So far as I have been able to discern from the work I have read—there is a summary in Dr. Bozzo's testimony[1]—this value is based essentially on judgment, impression, 'common sense,' and intuition. No rigorous statistical procedures were ever used to arrive at this parameter. Drs. Bradley and Bozzo have used quite complex multiple regression methods and a large data base generated within the U.S. Postal Service system to measure this effect, and have found a large amount of evidence that disagrees with the traditional assumption. They found consistent evidence that the volume variability factors for a large number of specific activities in the mail processing chain is considerably less than 100 percent

Witnesses Neels and Smith have raised a large number of criticisms of the data, methods and models used by Drs. Bradley and Bozzo and, by implication, their results. Primary among these issues are:

[1] USPS–T–15 at 4–13.

- Data quality problems and the issue of measurement error;
- Problems of nonrandom sampling that they suggest arose as a consequence of the methods by which the data were purged of possibly incorrect, misreported, or missing values;
- The issue of functional form relating to the use of certain 'panel data' style models—the 'fixed effects' model in particular;
- Certain other issues concerning the ways in which the regression model was formulated—among them the problem of missing variables.

I intend in my testimony to limit my attention to issues of econometric technique and model building. There is an enormous amount of fine detail in all the testimonies I read, about the specifics of and technical aspects of mail processing procedures and costs, work flow, and technology. Many of these details are advanced by Drs. Neels and Smith as severe complications that cast doubt on the econometric results. Although I believe that some of their comments in this regard are superfluous to the questions at hand, I will nonetheless not be addressing any of this material, and offer no testimony as to their relevance to the econometric modeling. Some of my testimony will be somewhat technical. Unfortunately, this is unavoidable. Some of the issues that the intervenors have raised, such as the problem of 'sample selection,' are, themselves, fairly esoteric.

My testimony will be related to the following general topic areas:

- The criticisms of the methods by which the data set was 'scrubbed' miss some important points about sampling, random sampling in particular, the nature of model building, and, very importantly, the issue of 'sample selection.'
- The discussions in the Neels and Smith testimonies relating to issues of measurement error rely upon some widely held

misconceptions about the topic. Most of their discussion on this issue is incomplete, and some of it is incorrect.

- Much of the discussion of the 'fixed effects' model for panel data, particularly the claim that it is inferior to a pooled regression or a regression based on group means, is incorrect.

My testimony will briefly review the models developed by Dr. Bradley and Dr. Bozzo, and the criticisms of them raised by Drs. Neels and Smith. A more detailed summary appears in Dr. Bozzo's testimony. I will then turn to the specific econometric issues listed above. To summarize my conclusions, I find that while some of the criticisms raised by Drs. Neels and Smith might provide useful guidance for refinement of the data used for estimating models for volume variability, many of the more methodological among their comments are exaggerated and/or misleading. I disagree with the suggestion that virtually all of the flaws suggested by the intervenors would have acted systematically to bias Bradley's and Bozzo's estimates of volume variability downward. On the contrary, from what I have read, I believe that the Bradley and Bozzo studies provide strong evidence that the 100% volume variability assumption should be reconsidered. While I am not prepared to commit to any specific value for any activity, I do believe that the two studies combined provide a strong suggestion that the right results will be substantially less than one.

III. CONCLUSIONS DRAWN FROM MY EXAMINATION OF THE STUDIES

I would not say at this juncture that every econometric or modeling issue that could possibly be addressed by Dr. Bradley or Dr. Bozzo has been addressed. I would definitely conclude that they have provided a substantial amount of evidence that the Commission should take very seriously.

The Commission should have taken a much more favorable view in 1997, and should at this time consider the panel data, fixed effects form of econometric analysis an appropriate platform for continuing work on developing a model for mail processing costs. The aggregate means models and time series regressions advocated by Drs. Smith and Neels discard far more useful information than the data scrubbing operation of which they have been so critical. Dr. Smith is simply wrong that the simple regression of group means on each other is the "least bad" model. Given the data set at hand, the simple regression of group means on each other is not the 'least bad' model; it is the second most bad model. The worst is the grossly aggregated time series regression proposed by Dr. Neels, followed by Smith's site means model, and the best of the lot is the fixed effects model. The arguments advanced by Smith and Neels in favor of their alternatives are based on flawed statistical reasoning, and should be rejected on this basis alone. The same conclusion applies to the visual devices advocated by Dr. Smith. I do not believe that the Commission should accept this kind of visual approximation as a substitute for careful econometric analysis.

The MODS and PIRS data are obviously far from perfect. But, from my vantage point, they appear to be quite good, and in the absence of a well designed and sharply focused data set designed specifically for studying volume variability, are as good as an analyst of mail processing costs could hope for. What is important is for the Commission and other researchers to evaluate these data using appropriate criteria. The criticisms raised in the Neels and Smith testimonies are, in many cases, inappropriate. Likewise, it sometimes happens that intuitively appealing evidence is misleading. For example, the standard deviations of the measurement error cited by the Commission in its Opinion (discussed below), which suggest an alarming amount of measurement error, appear to be much more discouraging than they really are. The intervenors in this case have thrown up an array of criticisms of the data set that raise a standard that could never be met. Apparently, the MODS data were not created for the purpose for which they were used in this proceeding. But that is usually the case with large micro level data sets. Nonetheless, it does seem reasonable to assert that there is useful information in the MODS data for the determination of volume variability. I would suggest that the Commission take the view that researchers should extract from these data what useful information they contain, not go to great lengths to discredit the data, and then discard them and the analysis based on them.

On the other hand, if the Commission desires to pursue the line of research begun in these studies of volume variability, then continued development of micro level data should be undertaken. In that connection, it is a maxim in econometrics that micro level data are always better than aggregates. The reason is almost self-evident. Aggregation almost always discards information contained in micro level data, and never creates new information. On the other hand, if it is genuinely believed that the micro level data contain no useful independent information, then they can be aggregated. This process cannot be reversed. By this construction, I am unable to agree with Drs. Neels and Smith that analysis of the MODS data should be done using site means of the same data set that could be used in disaggregated form.

Finally, what kind of model should be developed? It is clear that it is appropriate to use multiple regression to model the response of labor costs to output—the appropriate definitions of these two variables and how to measure them is an issue to be settled elsewhere. A simple regression of hours (or its logarithm) on output of any sort (or its logarithm) will surely ignore many other factors that that should be in the equation, including the site specific differences that Dr. Bozzo has analyzed. I also assume that the various models proposed will be based on the same data that have been used in this set of studies. In this instance, given the availability of micro level data, the fixed effects models proposed by Drs. Bradley and Bozzo are appropriate. At a minimum, they can do no worse, and will generally do better, than the site means ('between groups') models suggested by Drs. Neels and Smith. Aggregation discards information. The more crude the aggregate, the more information is discarded. At the very worst, if the disaggregated data really do not contain information beyond that which is contained in the group means, then a model based on the disaggregated data would simply mimic the model based on aggregated data.

Lastly, there is the question of econometric practice. The worst extreme I see here is Dr. Smith's willingness to rely on gross and misleading, crude two-dimensional scatter plots to defend a specific estimate of a parameter. Between this and the appropriate model lie the pooled regressions suggested by the intervenors, in which they impose restrictions on a regression model, then argue, in direct contradiction to long established results, that these results have improved the estimates. In particular, the suggestion that a pooled regression that imposes the restriction that there are no site specific effects somehow removes a bias inherent in the fixed effects model is simply not true—exactly the opposite is the case. Imposing restrictions can induce biases, relaxing them cannot. At the other end of the scale are Drs. Bradley's and Bozzo's carefully developed econometric models that embody current practice using an elaborate panel data set. The models have been subjected to numerous specification and sensitivity tests, and include features such as dynamic structure, time and site effects, models for autocorrelation, and flexible functional form for the estimated equation. As I noted earlier, I believe that this is the appropriate framework within which the Postal Service should be analyzing mail processing costs.

IV. THE VOLUME VARIABILITY MODELS

A. Dr. Bradley's Cost Equation Model

Dr. Bradley's model was estimated using a data set that provided for a large number of specific installations at numerous points in time. The three crucial variables were:[2]

HRS_{it}	= hours of labor at site i in period t
TPH_{it}	= total pieces handled at site i in period t
$MANR_{it}$	= manual ratio, a site specific measure of one aspect of the technology at site i in period t.

The specific equation estimated for 'Direct Activities' (sorting, etc.) includes linear, quadratic, and all cross products of these three variables, time effects contained in a time trend which allows for a discrete change in the time effect at a midpoint in the period of the analysis, one period lagged terms for the logTPH variable and its square, and a site specific dummy variable which allows for the site specific constant, or 'fixed effect.' All told, the equation includes 15 variables plus seasonal dummy variables, plus the site specific constants, so it is quite large. Additional lagged effects are introduced into the model through the use of a correction for autocorrelation in the disturbances. A similar, but slightly more involved, model was specified for the 'Allied Activities.'

The data used for the study contained numerous obvious flaws, and as a consequence, they were 'scrubbed' by a procedure that removed from the sample all observations:[3]

(1) that were not part of a continuous sequence of 39 consecutive observations that were otherwise 'clean;'

[2] See Docket No. R97–1, USPS–T–14 at 12–22.
[3] Id. at 30–37; see also USPS–LR–H–148.

(2) for which variables were obviously missing or erroneously coded as *zeros*;

(3) that were 'outliers,' in that they were in the top or bottom one percent of the distribution of the variable HRS/TPH.

Dr. Bradley subjected this model to numerous specification tests, including tests for whether a fixed or random effects model was appropriate—the latter rarely survives this test—tests for the presence of any site specific effects at all, and a test for autocorrelation.[4]

Estimates of the crucial elasticity of hours with respect to TPH for the direct activities ranged from 0.395 to 0.945; none of the estimates exceeded one.[5] The counterparts for the Allied Activities ranged from 0.720 to 0.829.[6] A number of other regression results were presented for other activities, all with similar results. The consistent outcome was that the volume variability varied across operations, rarely approached one, and almost never exceeded it. The equations were subjected to various specification tests, as noted, and estimated using several different methods, for example without the autocorrelation correction. The elasticity estimates were quite robust to the changes in the *estimation methods*.

Dr. Bradley conducted an analysis of the effect of measurement error in the TPH variable as well, using a method suggested in Hsiao's monograph on panel data.[7] As he points out, with panel data one can compute two different, albeit inconsistent, estimators of the slope coefficient and, at the same time, two inconsistent estimators of the measurement error variance. Solving two

---

[4] Id. at 41–51.
[5] Id. at 54.
[6] Id. at 63.
[7] Cheng Hsiao, *Analysis of Panel Data,* Cambridge University Press 1986, at 63–65.

equations in two unknowns, it is possible to obtain one consistent estimator of each of these two parameters. Dr. Bradley carried out the analysis in a restricted model, and found that the consistent estimator was quite close to the fixed effects estimator. As Dr. Neels pointed out, Hsiao's method can (and in this case, does) produce a negative variance estimator.[8] This is a small sample issue—Hsiao's results are based on infinite sample results. I confess some skepticism of this procedure, not over Hsiao's analytical results, which are correct, but whether this is the best way to approach the analysis. Hsiao's result applies in a very narrow specification, and produces the disturbing variance result in a very ordinary circumstance. It is prone to this finite sample problem. I emphasize, the test is not biased and is not misleading. It is simply one possible test and, I suspect, not the most robust one that could be constructed.

B. Dr. Bozzo's Updated Version of the Bradley Model

Dr. Bozzo's model is similar to Dr. Bradley's. In constructing it, Dr. Bozzo attempted to remedy some of the defects in Dr. Bradley's model that were argued by the intervenors and by the Commission, including the use of the data scrubbing procedure, and the absence of several other variables, including one relating to the capital stock and another relating to wage rates. As before, the model fit was a translog (log quadratic) model with site specific intercepts (fixed effects). The translog model was specified with four lags of the logTPH variable and its square, as opposed to one in the earlier study.[9] The data preparation for Dr. Bozzo's model is considerably more elaborate than Dr. Bradley's. The equation is also considerably more elaborate, involving the following variables:

[8] Docket No. R97–1, TR. 28/15637.
[9] USPS–T–15 at 117–118. Note that since Bozzo also changed from AP level to quarterly data, his model embodies a lag structure that is effectively 13 times longer than Bradley's.

HRS_{it} = the log of hours
TPH_{it} = the output (volume) variable (enters with four lags)
CAP_{it} = the capital stock variable
DEL_{it} = deliveries, to capture network and density effects
$WAGE_{it}$ = the wage variable
$TREND_{it}$ = trend variable to capture smooth time effects
$MANR_{it}$ = the manual ratio
QTR_t = specific quarterly dummy variables.

Dr. Bozzo estimated the model without transforming the data to overall mean deviations, unlike Dr. Bradley. The point is important as, in the current case, all relevant elasticities become lengthy functions of the parameters and the variables. The estimated elasticities obtained are similar to Dr. Bradley's, ranging from 0.522 to 0.954. (USPS–T–15 at 119–120; 126). Since considerable attention has been paid to the effects of different methods of estimation and forms of the equations estimated on the quantitative results, it is worth noting that Dr. Bozzo examined his results for sensitivity to different methods of estimation and computing of the elasticities, and found that the various computations produced very similar results. Id. at 130–131, 140-141, 151–160. See also USPS–LR–I–107.

V. ECONOMETRIC ISSUES RAISED BY THE INTERVENORS

As noted earlier, an extremely long list of problems with the preceding analyses was raised by intervenors Neels and Smith. Many of these related to whether the variables used in the analyses were appropriate or accurate measures of the activity being analyzed, e.g., whether total pieces handled

(TPH)[10] was an appropriate volume measure and whether the Postal Service's plant and equipment data really contain useful information about the capital stock. I will not be commenting on these concerns as I do not have the necessary background or knowledge about the specifics of the Postal Service. However, they did, as well raise several issues related to the econometrics.

- Dr. Neels: Most of Dr. Neels's rebuttal focused on the data issues mentioned above. He did make a number of points about the effects of measurement error that he feels persistently bias the estimated elasticities toward zero—that is, toward a result less than one. He was also critical of the screening of the data which produced the estimating sample. I will address this below.
- Dr. Smith: Dr. Smith has raised a daunting litany of criticisms of both the Bradley and Bozzo studies. I will focus my testimony on only a few of these:

  (1) He, as does Dr. Neels, criticizes the data scrubbing procedure.

  (2) He feels that the analysis is 'short run' in nature, and is therefore inappropriate for the phenomenon being studied.

  (3) He feels that observable (with his eyes) evidence contradicts the results of Dr. Bozzo's analysis.

  (4) He is critical of the panel data, fixed effects model that Dr. Bozzo used.

A fair amount of Dr. Smith's testimony is centered on issues of underlying microeconomic theory. Some of this is used to criticize the theoretical underpinnings of Dr. Bozzo's study. It is not my intention in this testimony to address issues of the underlying theory of any of this analysis; I come to this proceeding as an econometrician. However, I am familiar with the economics and econometrics of the estimation of cost and production functions. My doctoral

[10] Total pieces fed (TPF) was used in place of TPH in the automated and mechanized operations. See USPS–T–15 at 50–52.

dissertation and a subsequent paper published in the *Journal of Political Economy* with Laurits Christensen have long been regarded as pioneering studies (they were among the first) of the marriage of theory and empirical estimation of cost and factor demand equations—they are, for example, a standard application presented in microeconomics textbooks. While there are valid points in Dr. Smith's discussion of the theory behind an appropriate cost function, there are also some noteworthy errors. For example, Dr. Smith states that "Dr. Bozzo's treatment of homotheticity appears to lead to incorrect conclusions." Tr. 27/13196. He then states:

> In his testimony, Dr. Bozzo asserts that "... capital and labor variabilities will be identical in equilibrium under the assumption that cost-pool-level production (or cost) functions are *'homothetic'* ... Homotheticity implies that changing the level of output of the operation will not alter relative factor demands such as the capital/labor ratio, in equlibrium (and other things equal). However, the Postal Service testimony is replete with examples of the implementation of major investments designed to reduce costs. ... The focus is on the elimination of major labor costs via capital investment to achieve an overall reduction of total costs. Accordingly, the application of a homotheticity assumption appears to be an inappropriate assumption. (Id.)

Nowhere does the theory state that the firm in equilibrium will never adjust its capital labor ratio in response to changes in relative prices. Even if the technology is homothetic, the firm will respond to a change in relative prices by substituting away from the factor that is becoming more expensive unless it is unable to. This has nothing to do with whether the production function is homothetic or not. It is a question of factor substitution, and I do not believe that either Dr. Bozzo or Dr. Smith argued that the Postal Service operates under conditions of fixed coefficients, in which it would never substitute capital for labor in the face of increasing wage rates. The wages that appear in the labor demand functions estimated by Dr. Bozzo allow for adjustment in response to changes in wages over time, and are not consistent with a fixed coefficients assumption. Dr. Smith seems, as well, to suggest that the technology of production in the Postal Service is nonhomothetic, which it may well be. But no empirical evidence for this has been presented, and the mere fact that the Postal Service has invested in labor saving capital does not say anything on the subject one way or the other.

A. Sample Selection and the Data 'Scrubbing' Procedure

As noted, Dr. Bradley subjected the MODS data to a screening process denoted 'scrubbing' that was intended to remove observations that were obviously erroneous and unrepresentative of the underlying relationship he was attempting to uncover. This data scrubbing—systematic removal of observations from the sample—is the subject of a considerable amount of discussion in this proceeding. There are two issues that are brought forth by such a procedure. The first is biases. Under certain circumstances, selection of observations based on specific criteria (as opposed to randomly) can induce biases in the estimates obtained with the resulting sample. The second concerns the issue of efficient use of sample data—the problem of 'throwing away information.' In point of fact, efficiency has not been an issue in this proceeding. However, at some points, comments by the intervenors that are related to this issue have nonetheless been made, evidently to cast doubt on the Bradley and Bozzo studies. This section will discuss these issues.

To review, Dr. Bradley's screening procedure involved the following steps:

(1) He removed observations with obviously missing values, zeros coded for certain activities known to be taking place at the facilities observed, and observations for which reported output was below a specified threshold.

(2) He imposed a continuity requirement that the remaining data set for a site contain at least 39 useable consecutive observations.

(3) He removed the top and bottom 1% of observations based on productivity—the ratio of pieces handled to labor hours.

The productivity screen could disrupt the 'continuity' of the data series for some sites, so his data scrub was iterative in that after step (3) it was necessary to revisit step (2).

Among the issues raised by the intervenors was that this screening process removed an extraordinary amount of data from the sample.[11] The response to this point has been made in passing by Dr. Bozzo, but it bears repeating. The samples involved in these analyses are large, even after the data scrub.[12] However, irrespective of the size of the samples, if we are agreed at the outset that the underlying model that we seek to discover applies to all the data points, then absent the problem of nonrandom selection discussed in the next paragraph, the amount of data discarded has no bearing on whether the results obtained with the remainder are biased or not. Under the assumption, Dr. Bradley could have simply randomly thrown away three quarters of the observations, and safely based his results on the remaining quarter. Certainly, intuition would correctly suggest that this waste of information would be costly. But the cost is that of having a smaller sample, which leads to less precise estimates than one might otherwise obtain—i.e., larger standard errors. It has no relation at all to whether or not those estimates are systematically biased in one direction or another. The issue of how many data were discarded is a red herring.

[11] E.g., Docket No. R97–1, TR. 28/15609–619, 15632-15633, 15853. In the present docket, see Tr. 27/13163, 13172 and TR. 27/12796–12800.
[12] USPS–T–15 at 20–22, 95–102.

There is a substantive issue concerning how the data were removed from the sample. The overriding issue is whether the criteria used to discard data points were themselves related to the quantitative measure being studied, in this case, the log of the hours variable. This immediately raises a consideration that does not appear to have been noted by the intervenors or by Drs. Bradley or Bozzo. In particular, the missing or questionable values in the data set upon which the scrubs were based were the output variable, an independent variable, and the hours variable, the dependent variable. In the former case, once again, removal of data from the sample need not impart any particular bias to the results. Removal of observations from the sample because the output variable is missing or miscoded simply makes the sample smaller. Once again, the underlying relationship still applies to, and is discernible from, the observations which remain. Discarding observations based on values of the output variable is similar in its impact to throwing away observations randomly. On one hand, it may amount simply to wasting possibly useful information. On the other, if the output variable is erroneous while the hours variable is correctly coded, then my rule that the model must apply to all the data points would not hold, and the observation should be discarded. For an obvious example, positive output coded with zero hours makes no sense.

That leaves the missing or badly coded data on the dependent variable, hours. Bradley and Bozzo note a few cases.[13] Zero values recorded within a sequence of positive values are obviously erroneous. These once again violate the assumption that the model applies to all data in the sample, and they should be discarded. Bradley identifies another case, that of a 'ramping up' period, in

[13] E.g., USPS–T–15 at 109–110. See also Docket No. R97–1, USPS–T–14 at 30.

which the hours data would be unrepresentative.[14] As I noted in my introduction, I am not able to comment on specific technical aspects of the production process in mail handling. As such, I leave it to the others involved in this discussion to settle whether this is an appropriate omission. My own reading of the testimony suggests to me that it is.

The final case, and the one that does merit some attention is the trimming operation. Dr. Bradley eliminated the extreme values of hours per piece handled, from his sample, reasoning that these were unrepresentative and should be treated as erroneous data.[15] This is a specific sample selection rule that could, in principle, color the results obtained with the remaining sample. Dr. Bradley removed the top and bottom 1% of the distribution with this rule. Dr. Bozzo used a more detailed screen of this sort.[16] This productivity screen has called forth a criticism about "sample selection." Dr. Bozzo has commented specifically on the issue, but I believe there is a consideration that should be added. First, *sample selection* has become something of a bugaboo in econometric analysis, so we should be precise in our use of the term. What the productivity screen could potentially induce is a *truncation bias.* The distinction is important in this context because not very much is known about sample selection bias—except that it is bad—but a fair amount is known about truncation, and some of what is known has direct bearing on this case.

Dr. Bradley's productivity scrub of the data amounts to a trimming operation. Although the term 'selection bias' has been used in this context, the proper term is 'truncation.' Extracting data based on values of the dependent variable does have the potential to do some mischief. The pure theory of the

[14] Docket No. R97–1, USPS–T–15 at 30.
[15] Id. at 32.
[16] USPS–T–15 at 101–102, 110-112.

issue (discussed at various points in Chapter 20 of my text) does suggest that trimming the tails of the distribution would bias the least squares regression estimator toward zero. There are two crucial variables here, the asymmetry of the trim and the amount of the distribution that remains after the culling. The lesser the former and the greater the latter, the less 'damage' is done. In this regard, Dr. Bradley's productivity scrub scores well on both counts, in that he removed a fixed and very small percentage—one percent—from each tail. Dr. Bozzo's scrub was more complicated, in that he did not symetrically cull observations from the tails of the productivity distribution as per Bradley, but rather used cutoffs based on *a priori* information on maximum and minimum TPH per hour. It is impossible to tell what if any truncation bias would result from this. But, in any event, looking at his Table 3 (USPS-T-15 at 107) we see that, with two exceptions, the numbers of observations removed from the sample by the productivity scrub are generally so small that it would be hard to argue that the truncation effect would be substantial. His Appendix A (id. at 140) is suggestive. By foregoing the productivity screen and using "All Usable Observations," he obtains largely similar results. What I find surprising, and encouraging, about these results is that the theory suggests the estimates should rise, not fall, when the suspect observations are put back in the sample. In fact, most of the estimates fall, some substantially. Dr. Bozzo's type of screen does not conform to the assumptions in my text, so I don't see this as a contradiction. I do conclude that concerns about attenuation due to truncation of the data set are probably misplaced.

This leaves an important consideration, which entered both the Bradley and Bozzo discussions, the data continuity issue. Dr. Bradley imposed a 39 contiguous observation threshold on his sample.[17] Since he was fitting models with autocorrelation, this was primarily a practical consideration. Dr. Bozzo used tools (the econometric software package, *TSP*) in which time series with gaps are permissible, so the continuity requirement becomes a nonissue. But, in either case, it would be a question of sample size, not systematic coloring of the sample.

I am reluctant to generalize from narrow results to sweeping conclusions (though, in fact, both Dr. Neels and Dr. Smith have done so, using results taken from my book). But I do believe, based on the considerations discussed above, that the attention paid to the criticisms raised by Neels and Smith concerning the data scrubbing procedures has been exaggerated. Data that contain recording errors and other obvious flaws must be cleaned before being used. The samples used were large to begin with, and remained so after the trimming operations. By and large, the trimming operations were innocent. To the extent they were not innocent, the received theory suggests that the problems created are likely to be very small.

The foregoing is not meant to be glib. Data cleaning of this sort must be done carefully, particularly when the end result of the statistical process will be an input into a process as important as setting postal rates. Moreover, I do not dispute the possibility that the data scrubbing procedures used by Dr. Bradley were less than ideal, perhaps even less perfect than it potentially could have been had it been done differently at the time. Dr. Neels has raised some valid criticisms of the procedures; his observation that "unusual observations ... may also provide the clearest possible picture of how processing costs vary with volume" is well taken.[18] In his update of Dr. Bradley's model, Dr. Bozzo backed

[17] Docket No. R97–1, USPS–T–14 at 31.
[18] See Docket No. R97–1, TR. 28/15613.

away from some of Dr. Bradley's procedures. But in its Opinion and Recommended Decision from Docket No. 97–1 (PRC Op., R97–1, Volume 2, Appendix F), in the discussion of the cost models, the Commission devoted 11 of 45 pages (pp. 24–34) to this issue, and the conclusions it reached were quite dire. I believe that while many of the issues raised were appropriate, the conclusions were unduly pessimistic. After reviewing the procedures, the Commission stated "Witness Bradley's productivity scrub is exactly the kind of data elimination that econometricians try to avoid. Since the scrub eliminates values that are accurate as well as those that are erroneous, it leaves a sample that cripples the econometrics." Id. at 26–27. This is not true. Notwithstanding the truncation issue I raised above, discarding the extreme, though still valid, observations will indeed reduce the quality of the sample; it will do so by producing a model that is less precise (in terms of statistical measures such as standard errors) than it might be otherwise. But "cripples" overstates the case. The screen left more than adequate variation in the sample to allow econometric estimation of the model. Discarding anywhere from a quarter to a half of a sample might seem extreme, but it must be recalled that the sample that remained contained thousands of observations, not dozens, and the analysis attempted to estimate only a relative handful of coefficients. Faced with a need either to use obviously erroneous data or to discard with those data some observations that might have improved his estimates, I feel that Bradley took the right course. In order to argue that this data scrubbing "crippled the econometrics," one would have to argue that all or nearly all the data were bad, not just some of them.

The Commission makes one final argument about the data scrubbing process, that the process did not truly purge the sample of erroneous data. Id. at 33–34. This may well be true, but it is a side issue—the screen was not intended for this purpose. They cite certain values derived by Dr. Bradley to illustrate the extent of measurement error remaining in the data. Two aspects of this observation should be noted. The first is already made above. The screen was intended to provide complete and appropriate data, not data free of measurement error. Whether or not TPH is an appropriate measure of the output or whether errors are introduced by the conversion of some other measure to the TPH are valid concerns, but they are separate issues from the screening of the data discussed here. The second point concerns two numerical estimates of the extent of measurement error that are given. These measures are interesting, but are prone to misinterpretation, as I discuss in the next section.

B. The Issue of Measurement Error

A large amount of the criticism leveled at the Bradley and Bozzo studies concerns the issue of measurement error. Specifically, Dr. Neels argues that the output measure used, pieces handled in Dr. Bradley's case and pieces "fed" in Dr. Bozzo's case, do not correspond to the true measure of output that should enter the calculation of volume variability.[19] He concludes that the output variable which appears on the right hand sides of both regression models is measured with error. From this, he concludes:

(1) "It is a well established econometric principle that measurement error in an independent variable causes downward bias in coefficient estimates." (Docket No. R97–1, Tr. 28/15604. He goes on to state a quote from page 437 of the third edition of my text.)

(2) "Measurement error in an explanatory variable of a linear regression model renders the estimator inconsistent and frequently biases

[19] Tr. 27/12792–12793, 12802 et seq. See Also Docket No. R97–1, Tr. 28/15598–600.

coefficient estimates towards zero." (Tr. 27/12800. In this instance, he does not invoke my text.)

The statements above are based on a widely cited, fairly simple result from econometric theory. Suppose that the analyst wishes to estimate the slope parameter in a regression model:

$$y = \alpha + \beta x^* + \varepsilon$$

where x* is the desired independent variable, volume in this instance. Let x denote the observed independent variable, pieces handled, however measured. We further assume that x deviates from x* by a random measurement error, denoted u, so that $x = x^* + u$. In order to obtain the results that form the backbone of Dr. Neels's criticism, it must now be assumed that (a) the measurement error and the true variable are uncorrelated, (b) all variables are strictly uncorrelated across time and with other observations—i.e., we are using random samples—(c) the variances of u and x* are denoted θ^2 and λ^2, respectively. With these in place, we obtain the fundamental result that the slope estimator in a least squares regression of y on x (the observable data) will estimate consistently, not β, but

$$\gamma = \beta \times 1 / (1 + \theta^2/\lambda^2).$$

Two important points to note are, first, that if there is no measurement error, then θ is zero and least squares does what it should (it estimates β), and, second, when θ is not zero, least squares estimates β with a persistent downward bias. This is the source of Neels's result stated above.

There are quite a few misconceptions about measurement error in the discussions on the subject that I have seen in this case.

(1) The suggestion that measurement error biases all coefficients downward is generally not correct. The preceding statement is true under the circumstances assumed. However, none of the models discussed in the present docket or the preceding one involve a simple regression of a dependent variable on a single independent variable measured with additive error. In a multiple regression in which one variable is measured with error in this fashion, the coefficient on the badly measured variable is indeed biased downward, though not by the same amount as in the simple regression case. Also, other coefficients in the regression are affected as well, in unknown directions. There is one easy case to analyze. In the preceding example, with measurement error, the constant term is biased *upward*, not downward. The effect of the measurement error is to tilt the regression line, not to push it down. This observation is important in this case because all models are multiple regression models, not simple ones. (This result appears in my text four pages after the familiar one cited by Neels.)

(2) Whether or not the bias in the coefficients carries through to biases in functions of those coefficients, such as the volume-variability factors, is unknown. Any function of the coefficients in a multiple regression in which a variable is badly measured is a mixture of coefficients, some of which may be biased downward and others of which might be biased upward. The end result is not knowable.

(3) In time series data with autocorrelation in the variables, the effect of the measurement error will be mitigated if the underlying variables are correlated across time and the measurement errors are not. This has particular relevance here because lagged values of the output variable

appeared in the model, through the estimation of the autocorrelation model.

(4) In a model in which more than one variable is measured with error, essentially all bets are off. The familiar attenuation result can no longer be shown. The directions of biases, if any, are unknown. Since the models fit by Drs. Bradley and Bozzo are translog, quadratic in the logs of the output variable, this result applies here. In addition, note that the square of the erroneously measured variable appears in the models estimated by Drs. Bradley and Bozzo, so the assumption of additive error which enabled the derivation of the multiple regression case in my text is also lost.

(5) If the original data were measured with additive error, surely the logs of them are not. This means that the blanket statements made by Neels cited above are incorrect. The results would obtain if the logs were measured with additive error, which would be the case if the original data were measured with multiplicative error. Thus, the analytic results above have to be qualified, in ways that are not obvious.

Lost in this discussion is an assessment of the likely magnitude of the quantitative impact of measurement error. Without a large amount of very high quality data, we cannot say much with any precision on this subject. We can form some impressions, though. First, the familiar result on the previous page can be written in the form

$$\gamma = \beta \times \rho^2$$

where ρ is the correlation between the true variable and the one measured with error. As noted, I am not able to make a judgment on this sort of calculation. I do note that the R^2s in the regressions reported by the various authors are exceedingly high, sometimes above 0.99. Another effect of measurement error is to bias the fit of the model downward. Given values this high, I suspect that measurement error is not a major factor here. There is another way to approach this. Suppose the measure were multiplicative. It is possible to show that in this instance, the result becomes a bit cleaner,

$$\gamma = \beta/(1 + \theta^2).$$

Now, what value do we use for the measurement error variance? Suppose that the pieces handled variable varied in either direction by as much as 20 percent from its true value. This would imply a relative standard deviation of the measurement error of about 0.1, or a relative measurement error variance of about $\theta^2 = 0.01$. This is trivial. While a 20 percent error rate in the reporting of piece handlings seems fairly large, it implies only a 1% bias in the estimated coefficient, since with these values, $\gamma = 0.99\beta$.

All of these results are narrow theoretical conclusions based on a hypothetical situation. But I do believe that they have relevance here. The overriding result, which will fall out of any analysis, is that the damage done by measurement error will be a function of the 'reliability ratio':

reliability ratio = variance of true variable / variance of measured variable.

This, in turn, is a function of the correlation between the true and the measured variables. In cross sections, in which researchers attempt to measure such things as education, the reliability of self reported data can be extremely low. In this setting, by contrast, we are considering very stable flow variables which evolve reasonably smoothly through time. I would strongly predict that the reliability of output data in this setting is exceedingly high. Consequently, I would argue that criticisms of the models based on measurement error, while certainly to be taken seriously, are greatly exaggerated. Moreover, isolated examples in which an observed flow rate of pieces handled differs noticeably from some known true value are not really informative. What matters is the *correlation* between the observed measure and what we are trying to measure. Even in the face of a few egregious reporting errors, this seems likely to be very high for these data sets.

Interestingly, there are a couple of estimates in the Commission's Docket No. R97–1 Opinion. They state, citing a table in Dr. Bradley's rebuttal testimony (which I have not seen),

> The standard deviations for total piece handlings (TPH) derived from the variances in Table 3 are 0.268 for manual letters and 0.297 for manual flats. The corresponding standard deviations for the measurement error are 0.123 for manual letters and 0.068 for flats. These results do not support the conclusion reached by witness Bradley that large and material measurement errors are absent from the piece handling data for these activities....In the Commission's opinion these results are inconclusive but tend to support exactly the opposite finding, that large measurement errors remain in the sample after witness Bradley's scrubs. (PRC Op., Docket No. R97–1, Volume 2, Appendix F at 34.)

This seems true. However, one must be careful. While standard deviations are in natural units, the crucial variables for our analysis are variances (squared standard deviations). For the values cited by the Commission, the reliability ratios are $0.268^2/(0.268^2 + 0.123^2) = 0.826$ for manual letters and $0.297^2/(.0297^2+0.068^2) = 0.950$ for manual flats. Whether these are large or small is debatable, but the impression one gets from the reliability ratio is certainly different from the raw data.

C. Panel Data Treatments

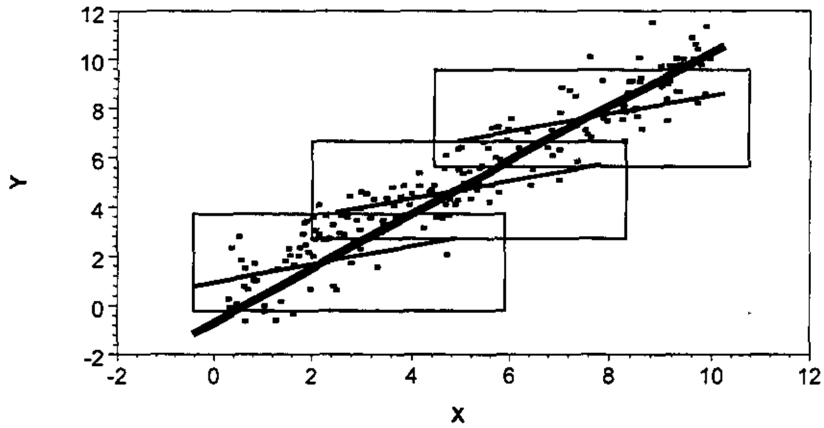
1. The Fixed Effects Model vs. a Group Means Model

For the purpose of this discussion, I will focus on Dr. Bozzo's model and the criticisms raised by Drs. Neels and Smith in this proceeding. Some of the discussion would apply to the earlier results, but it seems more useful to concentrate on the current efforts. Dr. Bozzo estimated a labor demand equation of the form:

$$\log HRS_{it} = \alpha_i + \Sigma_k \beta_k z_{itk} + \varepsilon_{it}$$

where α_i is a site specific constant and the remaining terms are the components of a conventional regression (though one that involves linear, quadratic, and cross terms in logs of the independent variables, time effects, lags of the output variable, and autocorrelation, which make it quite complicated). A central issue concerns the use of the fixed effects regression model.

Dr. Smith takes issue with the fixed effects approach altogether. TR. 27/13163–65, 13189–90, 13207–214. He argues, through the devices of a graphical and logical demonstration, that the fixed effects model is inappropriate, and that the appropriate model is a regression on means that does not have fixed site effects. To focus ideas, I use a caricature of his main figure, his Figure 4 (TR. 27/13210). I assume he would agree with this approximation to his depiction.



The three boxes in the figure represent the observed data for three sites. The line of best fit that passes through each box is that produced by a fixed effects regression. But, so Smith's argument goes, each of these is a short run regression, and the long run regression is the one that passes through the center of the data, which is the heavy line in the figure, and which has a slope much greater than the individual lines. The logic behind the demonstration is that the data inside the box embody a short run relationship that the site obeys, but in the long run the site obeys the steeper sloped long run relationship.

Missing from this demonstration is just when the short run becomes the long run. At some point, so the argument goes, the site in the lower box is transformed to the one in the middle box, as it is then that it climbs up this curve, and obeys the long run relationship. The problem with this discussion is that within each box (at least figuratively—in actuality within the MODS data) the sites' operations are observed for several years. What Smith characterizes as the "long run" regression relationship certainly should manifest itself at some point. Dr. Smith claims that the fixed effects model assumes that the capital stock is constant within a site for the full period of the observations, but this is not true.[20] Dr. Bozzo's model contains both a measure of the site's capital stock in each period (imperfect though it may be, it does not appear to be devoid of information) and a time trend. There is evolution of the technology built into the model.

Now, let's consider this argument statistically. Dr. Smith argues that the appropriate model is the group means regression. Let's suppose that it is. We'll make the assumption that the site means regression that Dr. Smith advocates is correct. That means that the linear regression model using the site means,

$$\bar{y}_i = \alpha + \beta\bar{x}_i + \bar{\varepsilon}_i$$

is appropriate. Suppose it is. Then it must be true that the disaggregated data obey the same regression:

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}.$$

Note the common constant term. It is there of necessity, because if this were not the case, then the group means regression suggested cannot be right. The only way there can be a single constant term in the group means regression is if there is a single constant term in the disaggregated data. Something is wrong here. Surely the data would have something to say about this. If the group means regression were appropriate, then when the fixed effects regression is fit, the site specific constants would all be the same, at least statistically so. But this is decidedly not the case. All the tests of this hypothesis decisively reject the hypothesis of no site effects.[21] The upshot is that it must be the group means

[20] TR. 27/13190–92.
[21] See USPS–T–15 at 123; see also Docket No. R97–1, USPS–T–14 at 41–43, Tr. 33/18021–22.

regression, which inappropriately restricts the regression model, that is biased—not the fixed-effects model.

There is another point of view here. If Dr. Smith is correct, then surely his group means regression could be fit using the disaggregated data. Disaggregating the data always relaxes restrictions, in this case, the restriction of equal constant terms for each site. It is a fundamental principle of statistics that when one relaxes a restriction, the very worst one can do is what one did before. Here, what this means is that if we don't assume the constants are the same, and they really are, then the regression on the disaggregated data should do no worse than the group means regression, and the site specific constants should resemble each other. Otherwise, the restrictions will appear to be incorrect. In fact, the MODS data are speaking loudly and clearly here. Statistically, Dr. Smith's argument in favor of the group means regression is not correct. Logically, it is weak as well, but whether that is the case or not, his graphical device cannot be used to support it, and his statistical interpretation is incorrect.

I would like to make one additional point at this juncture. The term "between" regression has been used at several points in the discussion, and it has been argued that the "between" regression (using group means) is more appropriate than the fixed effects model. The preceding addresses the site means issue. But it should be noted that there is an important distinction between the group means regression and the "between groups" regression. The fixed effects model is the 'within groups' regression. The "between groups" regression is a weighted regression of the deviation of site means of the dependent variable from the overall mean on the same transformation of the independent variables, *without a constant term*, and weighted by the number of observations made for each site. It is easy to show—it is done on page 619 in my aforementioned text, for example—that the same regression model applies to this data set as to the original data set. But this is not the group means regression that Dr. Smith has suggested. Dr. Smith's group means regression involves a simple regression of site means on site means, with a single constant term. This regression is only appropriate if the original model with no site specific effects is correct. Otherwise, the between groups estimator and the within groups estimator both estimate the same parameters, while Dr. Smith's regression produces results that are biased.

The preceding has addressed some econometric fine points. There is a substantive conclusion. Dr. Smith has advocated the group (site) means regression, with means constructed from the underlying data used to fit the fixed effects model, as somehow superior to the fixed effects model. Logically, this makes little sense. Statistically, it is simply incorrect. It is the group means regression which imposes the improper restriction, not the fixed effects regression. A fortiori, if Dr. Smith were correct about the means regression, then the worst the fixed effects model could do would be to mimic it. The fact that it does not is strong evidence that the assumption underlying the means regression is incorrect. His statement that the "between model" is the least bad model available is not correct either, even if he had fit the appropriate between groups regression.

2. Measurement Error

Dr. Neels has suggested that aggregating the data to group means helps to ameliorate the measurement error problem.[22] The logic is that averaging tends to average out the measurement error. It's an intriguing argument, and would be a very useful one if it were true. Unfortunately, it is not. Once again,

[22] Docket No. R97–1, TR. 28/15626-15630.

the statistical rebuttal is simple. It is true that in the averaged data, the variance of the measurement error is now divided by the number of observations. However, the variance of the mean of the systematic component is likewise. That leaves the crucial reliability ratio that I discussed earlier unchanged. If there is measurement error, it will exert the same influence on a model fit with site means as it would on the underlying disaggregated data.

3. A Pooled Time Series Regression

Dr. Neels has offered a pooled, aggregate yearly time series regression based on system wide costs and volumes as an appropriate tool for analyzing volume variability.[23] If it could be argued that there were no systematic variation in volume variability factors across sites or activities, no adjustment within calendar years in response to changes in output, and no long run adjustment in response to technological advances, this might be appropriate. None of these assumptions seems warranted. And whether they are or not, assuming them at the outset discards nearly all the useful information to be had from the disaggregated data set.

The (lack of) usefulness of the time series regression suggested is the same as that for the group means regression. Once again, the statistical result is clear. If it were appropriate to aggregate the data—in this case, that would mean no site specific and no period specific effects—then the aggregate and the disaggregated approaches would produce similar estimates of the same parameters. The disaggregated approach cannot make things worse. When they differ substantially, as they do here, the right conclusion to draw is that the

[23] TR. 27/12835–12843.

aggregated approach is causing the problem. To reiterate, the disaggregated data will give the right answer whether or not Dr. Neels' approach is correct.

D. Alternative Estimates Based on a Reverse Regression

In a demonstration intended to show that piece handlings rise faster than volume, Dr. Neels presents a set of results using a technique known as reverse regression (TR. 27/12805–12810). This technique originated roughly two decades ago in the sex discrimination literature. (See *Journal of Business and Economic Statistics*, (April 1984) in a symposium in which I have a contributed paper.) The logic of the technique is suggested by the following: In a regression of wages on qualifications, job characteristics, a dummy variable for sex, and other control variables, if there is discrimination, the coefficient on the dummy variable will be positive and the coefficient on qualifications will be too small. If so, then in a regression of qualifications on wages, the sex dummy, and the other variables, the coefficient on the dummy variable will be too high. In essence, if women are underpaid, then they are overqualified for the jobs they have.

Dr. Neels has extended this methodology, using an alternative approach to regression in the presence of measurement error. In particular, he states, "[t]o avoid the pitfalls of errors-in-variables bias, I estimate the elasticity of TPH/F with respect to FHP using the reverse regression of FHP on TPH/F and other variables..." TR. 27/12806. Then, to obtain the desired result, he takes the reciprocal of the elasticity of FHP with respect to TPH/F derived from the reverse regression. Id. The reasoning appears to be that in reversing the dependent and independent variables, he can pile the measurement error into the equation error on the right hand side and mitigate the measurement error bias that might affect the direct regression. Once again, he cites my text: "It is a well known result that measurement error in the dependent variable is absorbed in the error term and can be ignored." Id. (footnote omitted). The quotation is right (well, close enough), but the regression result is not. The reason is that, even with the rearranged equation, the measurement error is still in the independent variable, and the estimator remains inconsistent, as I now show.

The prototype for the original regression is:

$$y = \beta x^* + \varepsilon$$

$$x = x^* + u$$

exactly as specified earlier. The thinking in Dr. Neels's reverse regression, apparently, is embodied in:

$$x^* = (1/\beta)y - (1/\beta)\varepsilon$$

so that:

$$\begin{aligned} x &= x^* + u \\ &= (1/\beta)y - (1/\beta)\varepsilon + u \end{aligned}$$

which is a conventional regression which appears to obey the claim from my text that was quoted earlier. We just absorb u in the disturbance, compute the regression, then take the reciprocal of the coefficient. It doesn't work. Relying on conventional regression results, and to avoid technical details, I go directly to the result. The least squares slope estimator in the reverse regression of x on y is an estimator of the quantity

$$\delta = \mathrm{Cov}[x,y] / \mathrm{Var}[y].$$

We can derive this just by going back to the original specification;

$$\delta = \beta\lambda^2 / [\beta^2\lambda^2 + \sigma^2]$$

where σ^2 is the disturbance variance. Neels' estimator would be

$$b = 1/\delta$$

which estimates not β but

$$1/\delta = \beta[1 + \sigma^2/(\beta\lambda^2)].$$

Whether or not there is measurement error—indeed, independently of the measurement error—the Neels estimator will overestimate the true coefficient that he seeks. His elasticity estimates are biased upwards.

What went wrong? What went wrong is that this manipulation simply trades one misspecification for another. Looking back at the rearranged equation,

$$x = (1/\beta)y - (1/\beta)\varepsilon + u$$

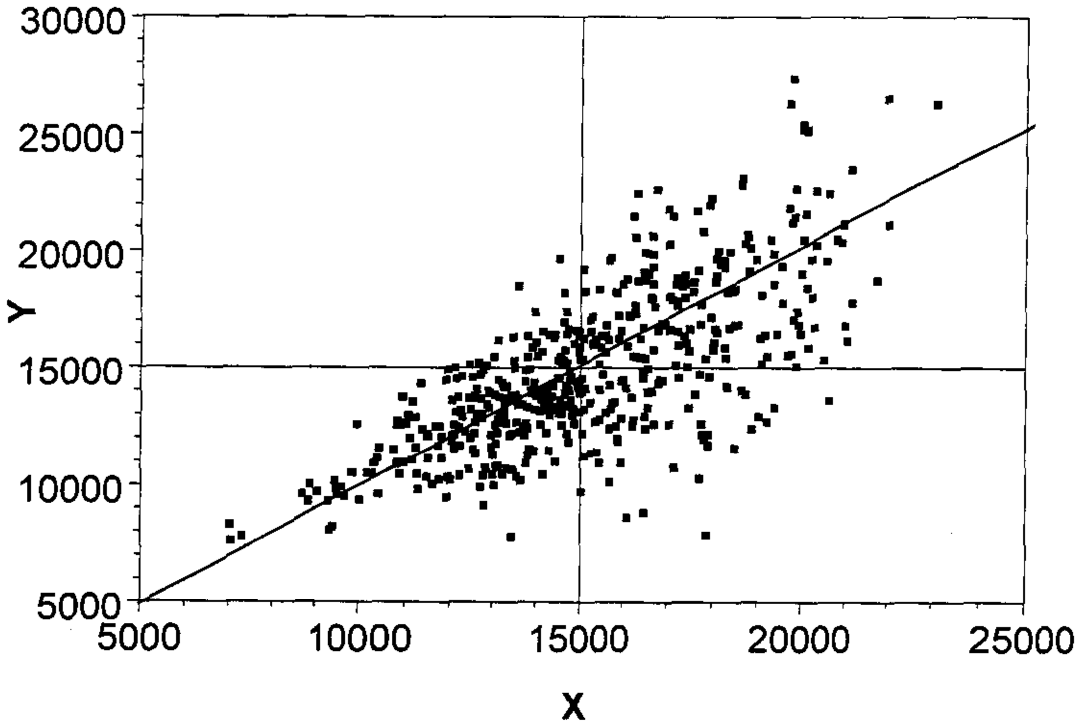
what we find by manipulating it a bit is that the 'independent variable,' y, is correlated with the 'disturbance,' $-(1/\beta)\varepsilon+u$; the covariance is $-\sigma^2/\beta$. This violates another assumption of the regression model, and renders least squares estimates from the reverse regression inconsistent. In fact, the accepted result on reverse regression in the presence of measurement error is that the reverse and direct regression estimators <u>bracket</u> the correct result, which is what is shown above.

I hesitate to generalize from narrow results. The reverse regression estimator is generally going to be vastly more complicated than I have made it, because it is usually embedded in a multiple, not simple regression model, and at least in this case, we are not completely agreed on the nature of the measurement error in any event. I do believe, however, that a firm conclusion is safe here. Reverse regression is not an appropriate way of "avoiding the pitfalls of errors-in-variables bias." This method and the estimates presented should not be accepted. Rather, if the problem is to be analyzed and solved, it should be done so directly.

E. Visually Compelling Plots

In both his 1997 and 2000 testimonies, witness Smith has relied on some low resolution figures generated by the SAS computer program to bolster his suggestion that the data and evidence in hand are consistent with 100 percent volume variability. Indeed, in the 1997 testimony he characterizes the figures as "visually compelling in demonstrating a variability approaching 100 percent between labor hours and mail volume."[24] Irrespective of any other evidence or results discussed in this proceeding, I would caution the Commission against relying on any visual devices of this sort. I do not believe that the eye is capable of resolving such evidence at a level approaching "compelling" or even convincing. I offer the figure below as motivation for this belief. The figure

[24] Docket No. R97–1, Tr. 28/15847.



contains a plot of 500 points that, as those in Dr. Smith's graphs do, certainly appear to be consistent with a 100% variability relationship. The solid 45 degree line plotted in the figure is intended to aid the eye in reaching its conclusion. However, they are not consistent with such a relationship—by what appear to be the proportions of interest in this proceeding, not even closely. The points in the figure were cleanly produced with a random number generator so that the values of *X* are drawn from a normal distribution with a mean of 15,000 and a standard deviation of 3000, while *Y* was produced so as to be equal to 2000 + 0.85 times *X* plus a random normal draw with mean zero and standard deviation *X*/100. In the terminology of this proceeding, the data were constructed with an 85 percent volume variability. (The standard deviation is intended to produce the kind of fanning effect in the figure that is typical in Dr. Smith's figures. This feature will not produce any effect on the slope of an underlying relationship; it will only produce systematic variation around that relationship.) In the context of the results I have reviewed in the various testimonies, the difference between 0.85 and 1.00 is substantial.

I realize that this little demonstration is simplistic. The data were carefully constructed so as to produce an impression, not to mimic any real outcome that an analyst might observe. The purpose is to suggest that visual devices such as this, which could be based on such real data, could be very misleading. I do not believe that one could rely on a visual inspection of such a figure as this, which is itself of considerably higher quality than Dr. Smith's, to draw a conclusion about a precise quantity such as the slope of a regression. Graphs such as those in Dr. Smith's testimony should not be substituted for careful statistical analysis, and should not be accepted as evidence that casts doubt on any such analysis.