ORIGINAL

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

RECEIVED
FEB 6 4 33 PM '98
POSTAL RATE COMMISSION
OFFICE OF THE SECRETARY

POSTAL RATE AND FEE CHANGES, 1997

Docket No. R97-1

RESPONSE OF UNITED STATES POSTAL SERVICE TO
NOTICE OF INQUIRY NO. 4 ON MAIL PROCESSING VARIABILITY
(February 6, 1998)

The United States Postal Service hereby provides its response to Notice of Inquiry No. 4, issued on January 16, 1998. The Postal Service's substantive comments on the NOI are contained in the attached Statement of Professor Michael D. Bradley. This statement is also being provided to the Commission on diskette. In addition, the programs and results of Dr. Bradley's analysis are being filed today as Library Reference H-339, Econometric Programs and Results Provided in Response to NOI#4. In accordance with Presiding Officer's Ruling No. R97-1/95, issued February 4, 1998, Dr. Bradley is prepared to adopt his statement and the accompanying library reference as testimony, to respond to appropriate written discovery and to appear for cross-examination on March 2, 1998.

The Postal Service understands the due process concerns and the goal of having a complete record which led to issuance of Presiding Officer's Ruling No. R97-1/95 and intends to comply. Nonetheless, the Postal Service has reservations about both the timing of the procedures established for the NOI.[1]

POSTAL RATE COMMISSION
DOCKETED
FEB 6 1998
OFFICE OF THE SECRETARY

[1] The extra work required to respond to the NOI and to prepare for hearings and rebuttal scheduled for March 2 and March 9, respectively, cannot help but interfere with the Postal Service's and other parties' abilities to prepare rebuttal to intervenor cases, also scheduled for March 9, 1998.

Moreover, its seems that the effect of the NOI is to allow parties yet another opportunity to rebut Dr. Bradley's analysis. Intervenors had their opportunity with the filing of their direct cases on December 30, 1998. To the extent that intervenors chose not to file testimony or to file testimony covering certain issues, and to the extent that the parties will be filing rebuttal to each other's cases on March 9, 1998, the NOI seemingly affords a "second bite at the apple" to those opposed to Dr. Bradley's analysis. The Postal Service believes that had this matter been raised earlier, before the filing of intervenor testimony, the Postal Service and other parties would have been able to proceed in a fashion which afforded all participants better opportunity to argue and defend their positions.

Respectfully submitted,

UNITED STATES POSTAL SERVICE

By its attorneys:

Daniel J. Foucheaux, Jr.
Chief Counsel, Ratemaking

Susan M. Duchek

Eric P. Koetting

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260–1137
(202) 268–2990; Fax –5402
February 6, 1998

Statement of Professor Michael D. Bradley

on

The Postal Rate Commission's

"Notice of Inquiry No. 4 On Mail Processing Variability"

February 6, 1998

I. INTRODUCTION

When an econometrician is fortunate enough to have access to panel data, he or she has information along two dimensions — through time and across sites. This is a considerable advantage over either cross sectional data, that can only vary across sites, or time series data, that can only vary through time. However, the dual dimension of the data does raise an issue about how the data should be organized for estimating econometric equations. The "Notice of Inquiry No. 4 On Mail Processing Variability" (hereinafter NOI #4) touches on a single dimension of this issue and thus seeks statistical evidence on only a single hypothesis.

To understand the statistical test that the Commission requested and to understand the implications of the results, it is important to understand the context for the single, specific, statistical test requested in NOI #4.

If one starts from the position that the individual sites in the panel have very little in common, then one would think of the panel data as a set of individual time series.[1] In that case, an econometric model would be estimated individually for each site. Formally, this would imply a set of equations of the form:

---

[1] Of course, this approach requires that there are sufficient continuous data on each site to permit estimation. If the panel data set includes just two or three data points for a site, then such an approach would not be possible.

$$y_{it} = \alpha_i + \beta_i x_{it} + \eta_{it}$$

If one believes that there are some commonalities in production, or if one is trying to estimate the response of the entire system of sites to a change in volume, then one imposes a restriction on the above equation. By restricting the slope parameters (the β_j) to be equal across sites while letting the intercept parameters vary (α_j), one captures the common or, in some sense, "average" response to volume changes. Yet this approach does not impose the strict requirement that the production process is identically equal across all sites. In fact, the use of a flexible functional form (like the translog) within this approach allows for varying elasticities across sites with out the need for estimating additional parameters. This intermediate and widely used approach to exploiting the richness of panel data is often referred to as the "fixed effects" model and can be written formally as:

$$y_{it} = \alpha_i + \beta x_{it} + \eta_{it}$$

A formal statistical test for this restriction is provided by testing the hypothesis that $\beta_i = \beta_j = \beta_k \ldots B_N$, where N is the number of sites in the data set. This is the statistical test requested in the NOI #4.

A more restrictive set of assumptions is embodied in what is known as the "pooled model." In this model, the production process is assumed to be identically equal across all sites and a single set of coefficients is estimated for all sites. The pooled model, which is nested inside and thus more restrictive than the fixed effects model, is given by:

$$y_{it} = \alpha + \beta x_{it} + \eta_{it}$$

The formal statistical test for the pooled model is provided by testing the hypothesis that $\beta_i = \beta_j = \beta_k \ldots B_N$ **<u>and</u>** that $\alpha_i = \alpha_j = \alpha_K \ldots \alpha_N$, where N is the number of sites in the data set. The NOI #4 did not request testing this hypothesis, presumably because the Commission was aware that rejecting the fixed effects model relative to the site-by-site <u>necessarily</u> implies rejecting the pooled model in the same comparison.

In the balance of these comments, I explain how one goes about testing these restrictions for the data set and model at hand, and I provide the results of such testing. I then explain the implications of the test results.

II. PRELIMINARY CONSIDERATIONS

There are procedures in econometrics for which specifying a "textbook" set of statistical tests is easy in theory, but for which actually carrying out those tests with real data and a real model is much more complex and difficult. The F-tests requested in NOI #4 are an example of such a procedure. Because the calculation of the requested F-statistics is not completely straightforward, the Commission should be aware of important issues that must be addressed in making those calculations. It should also check to be sure that all responses to NOI #4 address these issues. To inform the Commission on these issues, I provide a brief discussion of each.

A. The Issue of Non-invertability

The object of the F-test specified in the NOI #4 is to test the restriction that all of the estimated non-intercept parameter values are the same at all of the individual sites. This requires, of course, that the same set of parameters be estimated for each site. Although this is clear in theory, in practice this can be difficult because the data for a particular site may not permit estimation of the equation. In particular, one may not be able to invert the matrix of right-hand-side variables for a specific site and thus may not be able to estimate the set of regression coefficients for that site. Non-invertability can arise because of perfect collinearity among the limited number of data points for a site.[2] In these instances, OLS estimates for the set of model parameters cannot be obtained.[3]

To ensure direct comparability in calculating the F-statistics, I re-estimated the fixed effects and pooled models with data from just those sites for which the data permitted estimation of the set of site-specific parameters. All subsequent F-tests are, and should be, performed on a consistent set of data and estimated parameters.[4]

B. The Issue of Serial Correlation

F-tests of the type requested in the NOI #4 presume well-behaved residuals from the estimated regressions. In the particular case at hand, it is known that the residuals are serial correlated. Therefore, the F-tests should be done on the models corrected for serial correlation. In addition, one must recognize that an equation for each site is being estimated individually. Just as the α and β coefficients are allowed to vary across sites, so should be the ρ (serial

---

[2] It is important to recognize that this is not just the multicollinearity problem described in my response to POIR #7. Under that type of multicollinearity, the estimated coefficients are unreliable. Here, in contrast, the matrix cannot be inverted so the specified model cannot be estimated for the site.

[3] The test requested by the NOI #4 highlights the advantage of working with a data set that includes at least 39 observations per site. To the extent the data set was populated by small, fragmented chains of data, these tests could not go forward.

[4] Parameter estimates and the calculation of the F-statistics are included in Library Reference H-339, "Econometric Programs and Results Provided in Response to NOI #4."

correlation) coefficients.[5]

C. The Issue of Mean Centering.

When I estimated the fixed effects and pooled models for my testimony, I mean centered the data for computational convenience. Under mean centering, each observation is expressed as its deviation from the overall sample mean. Mean centering allows calculation of the desired elasticity directly from the TPH coefficients. However, when the models are estimated on a site-by-site basis, there is no advantage in constructing the data as deviations from the overall sample mean.[6] The site-by-site regressions are not on mean centered data. Therefore, to ensure direct comparability of the results of the various estimations, I re-estimated the fixed effects and pooled model on the data without mean centering[7]

---

[5] I do, however, present the F-tests based upon the uncorrected residuals. Although inferences should not be based upon these calculated statistics, I present them to reassure the Commission that the F-test results are not dependent upon the serial correlation correction.

[6] If one wanted to preserve the mean centering approach, one could mean center each site on its own data. The important thing is to be consistent among the site-specific regressions and the fixed effects and pooled models.

[7] When the data are not mean centered, one cannot simply calculate the variability from the coefficients on the TPH variable. Instead, one must calculate the derivative of log hours with respect to log TPH. This means that various cross product terms are included in the elasticity calculation. Also, because the models are designed to calculate the response in hours to a sustained increase in volume, the derivative should also include the response in hours to the lagged TPH terms.

III. TESTING THE RESTRICTIONS

As explained in the introduction, the test requested in the NOI #4 is actually part of a sequence of statistical tests which can be used on panel data.[8] In this section, I present and explain the complete sequence of tests and present the results of applying the sequence of tests to the current data. Recall that there are three sets of possible parameters and three hypotheses to be tested.[9]

A. Hypothesis 1: Pooled Model vs. Site-by Site Regressions

The first hypothesis that is tested is that of the restrictions embodied in the pooled model. The hypothesis of common intercept and slope coefficients, which I call H_1, amounts to a set of (K+1) * (N-1) restrictions,where K is the number of (non-intercept) right-hand-side variables and N is the number of sites. This hypothesis can be tested by calculating the following F statistic:

---

[8] For a discussion of this sequence of F-tests please see, Cheng Hsiao, Analysis of Panel Data, Cambridge University Press, Cambridge, U.K., 1986 at pages 12-18.

[9] Formally speaking, there is a fourth parameterization that could be investigated. It is technically possible to envision a set of parameters in which the intercept is the same at all sites but the slope coefficients vary. However, I follow Hsiao, (see p.13) in not testing this case:

> Because it is seldom a meaningful question to ask if the intercepts are the same when the slopes are unequal, we shall ignore the type of restrictions postulated by (2.2.3).

$$F_1 = \frac{(SSE_P - SSE_S) / [(N-1)(K+1)]}{SSE_S / [NT - N(K+1)]}$$

where SSE_p is the sum of squared errors from the pooled regression and SSE_s is the sum of the sum of squared errors from the site-by-site regressions. If this F statistic is not significantly different from zero, then it would suggest pooling the data and estimating a single set of parameters.

Hypothesis 2: Fixed Effects Model vs Site-by-Site Regressions

The second hypothesis compares the fixed effects model with the site-by-site regression. The fixed effects model assumes common slope coefficients but different intercepts. I term this hypothesis H_2, and it amounts to testing a set of N-1 restrictions on the coefficients. This hypothesis is tested by calculating the following F statistic:

$$F_2 = \frac{(SSE_F - SSE_S) / [(N-1)(K)]}{SSE_S / [NT - N(K+1)]}$$

where SSE_F is the sum of squared errors from the fixed effects model.

B. Hypothesis 3: Pooled Model vs Fixed Effects Model

Under the maintained hypothesis of common slope coefficients, one can test the fixed effects model against the pooled model. I call this last hypothesis H_3. Testing this hypothesis requires testing a set of N-1 restrictions on the α

coefficients given the maintained hypothesis that the β_j are equal. This hypothesis is tested by the following F statistic:

$$F_3 = \frac{(SSE_P - SSE_F) / [(N-1)]}{SSE_F / [NT - N - K]}$$

The calculated F-statistics for each of the three hypothesis are presented in Table 1 below. Results are presented for residuals corrected for serial correlation as well as for uncorrected residuals.

TABLE 1
F-Statistics for Hypothesis Tests

	Manual Letter	Manual Flat	LSM	OCR
	Serial Correction Included			
F_1	5.79	5.60	6.04	3.48
F_2	4.03	3.48	3.45	2.21
F_3	27.75	36.71	43.42	27.15
	No Serial Correction Included			
F_1	40.26	40.79	41.13	23.40
F_2	9.90	10.05	9.63	9.61
F_3	232.00	241.77	244.53	106.33

IV. INTERPRETING THE RESULTS

Because of the large number of observations in the data set, the critical value for all of the F-tests can be taken as 1.0. Thus, in any instance in which the calculated F-statistic exceeds 1.0, rejection of the null hypothesis is indicated. The results for F_1 presented in Table 1 indicate that the null hypothesis of site-specific homogeneity is rejected. This implies that site-specific heterogeneity is important and, consequently, a pooled equation should not be estimated for the mail processing activities. The question, then, is how to model the site-specific heterogeneity. Site-specific heterogeneity can be modeled with either a fixed effects approach or a site-by-site regression approach.

The second F-statistic, termed F_2, provides information on this question as it investigates the hypothesis that the 27 (non-intercept) slope coefficients are identical across sites. The results for F_2 suggest that this hypothesis is also rejected, albeit with much smaller F-statistics. This result is not surprising, as I suggested in an earlier interrogatory response, because the test requires equality of 27 different regression coefficients across hundreds of sites. In fact, it would be surprising if the null hypothesis did hold.

Finally, as a check on the nature of the site-specific heterogeneity, one can test the restriction that the α_i the same across facilities, given the maintained hypothesis that the β_i are. This set of restrictions is tested by comparing the fixed effects model with the pooled model, and the results are presented in the F_3 row in Table 1. The F-statistics for testing this hypothesis are very large and thus indicate a “strong” rejection of the pooled model in favor of the fixed effects model.[10]

If the approach to econometric modeling was based solely upon this set of F-tests, the results would indicate a preference for site-by-site estimation of the regression equations. However, in a more informed determination, the econometrician should consider a number of factors in making the modeling choice. These factors include:

1. The severity of the restrictions being testing.
2. The goal of the econometric research.
3. The ability of the data to support reliable site-specific equations.
4. The possible reasons for rejection of the restrictions.

As I discussed above, the complexity of the econometric specification implies that a relatively large number of parameters need to be estimated for each site. This increases the "tightness" of tests of the fixed effects regression and makes it more likely that the hypothesis will be rejected. Yet the richness of the specification is a strength of the analysis, not a weakness, and the fixed effects approach allows one to estimate a more sophisticated model. This advantage must be traded off against the "cost" of imposing the restrictions.

This tradeoff can be evaluated relative to the goal of the research. The

---

[10] This result corroborates the Gauss Newton specification tests presented in my testimony. Those results strongly suggested that site-specific differences are important.

goal here is not to estimate individual site equations or aggregate equations to be applied to individual sites for the purpose of, say, evaluating the individual sites. Rather, the goal is to produce an aggregate variability or elasticity for the relevant cost pool.[11] This means that even if site-specific elasticities are calculated, they must be averaged or aggregated in some way. This aggregation, in itself, imposes some implicit restrictions, so in choosing between the models, the validity of these implicit restrictions must be balanced against the validity of the explicit restrictions in the fixed effects model.

In addition, the site-specific variabilities are dependent upon the ability of the data to estimate reliable equations for each site. If, for example, reliable equations cannot be estimated for a number of sites, then the aggregate elasticity derived from this approach will be less representative than the elasticity derived from the fixed effects model. Moreover, given the size and complexity of the mail processing analysis, evaluating the set of site-specific equations for just one activity would be a time-consuming job. Evaluating and defending the site-specific equations for each site, for each activity, is an overwhelming job. This also reduces the attractiveness of the site-by-site approach — it is far more difficult to get a thorough review of the econometric results.

The site-specific approach also has the weakness that does not provide

[11] The fact that the site-specific equation approach is not consistent with estimating an aggregate variability does not imply that it is necessarily the best econometric model to estimating site-specific variabilties. As demonstrated in my response to POIR #7, the fixed effects model provides a complete set of sensible variabilities for individual sites. In fact, the fixed effects model can provide variabilities for many sites for which the site-specific approach cannot.

variabilities for costs generated at sites not included in the econometric analysis. The fixed effects approach provides a representative variability that is applicable to the costs for all sites included in the data as well as for those that are not. Because the site-specific approach is just that — site-specific — its estimated variabilities are not directly applicable to other sites. Thus, it does not provide a method for generating variabilities for sites not included in the estimation process.

Finally, the inability of the site-specific data to estimate reliable site-specific equations may be a contributor to the rejection of the null hypothesis. To the extent that data inadequacies at the site-specific level cause variation in the estimated coefficients, the likelihood of rejection of a "true" restriction is increased. It is clear, for example, that the site-specific equations suffer from multicollinearity, which generates instability in the parameter estimates. It is thus possible that the calculated F-statistics are reflecting this problem. Moreover, traditional remedies for multicollinearity include adding more data to the analysis and imposing restrictions on the parameters to be estimated. These are just the remedies provided by the fixed effects approach. Use of the fixed effects approach thus avoids the problems generated by the site-specific data inadequacies.

When all of these factors are considered, I believe that the fixed effects approach continues to be superior to the site-by-site approach even given the results of the F-tests. For the reasons discussed above, I would encourage the Commission to adopt the fixed effects approach.

Nevertheless, to the extent the Commission felt compelled to accept the site-by-site approach, it is important to understand the implications for the variability estimates. To gain that understanding, I used the site-specific equations to calculate an elasticity for each site. This calculation is performed by using each site-specific regression equation with the means of the data for just that site. The derivative of log hours with respect to log TPH (and lagged log TPH) is calculated from the regression equation for each site in the data. It is then evaluated at the sample mean values for that site alone.

The clear implication that emerges from this exercise is that the site-by-site results strongly suggest that the true underlying variability is not 100 percent. If the true underlying variability were 100 percent, the distribution of site-specific variabilities would be massed around that value. However, as the attached histograms show, the vast majority of sites have a variability *between* zero and 100 percent and the variabilities are massed at a value far below 100 percent.

The histograms also show that there are a few outliers with variabilities that are negative or greatly larger than 100 percent. But given the multicollinearity problem that I described before, such results are expected. Nevertheless, the site-by-site results seem to validate the reliability of the data because for the overwhelming majority of sites, the data produce "sensible" variabilities.[12] Taken together, the points indicate that the true underlying

---

[12] I do not make a stronger claim for the site-by-site results because, given the number of individual sites, I could not review all of the individual site results and determine the reliability of each one. Therefore, I view the site-by-site results as being suggestive and corroborative.

variability is far from 100 percent, particularly for the manual letter and manual flat activities.

I believe that there are some serious concerns associated with applying the site-by-site approach, but if the Commission did pursue the disaggregated approach, the individual site variabilities would have to be combined into a single overall variability. To see the outcome of this experiment, I calculated the average variability from the site-by-site regressions and compared it with the variability for the pooled model and the fixed effects model for the same offices.[13] The calculated variabilities for the site-by-site approach, the fixed effects approach and the pooled approach are all presented in Table 2.

TABLE 2
Estimated Variabilities from the Three Approaches

	Manual Letter	Manual Flat	LSM	OCR
Site-by-Site	0.524	0.523	0.832	0.707
Fixed Effects	0.728	0.763	0.913	0.768
Pooled	1.030	1.071	1.024	0.978

Interestingly, in all cases, the average site-by-site variability is lower than the fixed effects variability and the pooled variability. In fact, as additional restrictions are imposed, the variability rises. This indicates that the pooling bias,

[13] Recall that the data for some offices do not allow estimation of the individual site regressions. To ensure an accurate comparison, I re-estimated the equation and variabilities for the pooled model and the fixed effects model for the exact set of offices for which an individual equation is available.

if there is one, is positive. A positive pooling bias means that the fixed effects model, if anything, may overstate the "true" variability.

V. CONCLUSION

The specification analysis implied by the NOI #4 reveals that a fixed effects approach appears to be a good approach for econometric modeling of mail processing activities. The fixed effects approach provides the best balance of raw statistical accuracy, accurate model specification, avoidance of bias, and overall econometric reliability. It makes use of all the data in an effective way and provides an accurate estimate of the overall variability without encountering severe multicollinearity problems. It permits specifying an equation that effectively controls for non-volume influences on hours, ensuring an accurate measure of the elasticity estimate

In addition, the specification tests establish that if one uses statistical tests to reject the fixed effects model, those same tests can only imply a simultaneous rejection of the hypothesis that the mail processing variability is 100 percent.

Appendix to

Statement of Professor Michael D. Bradley

on

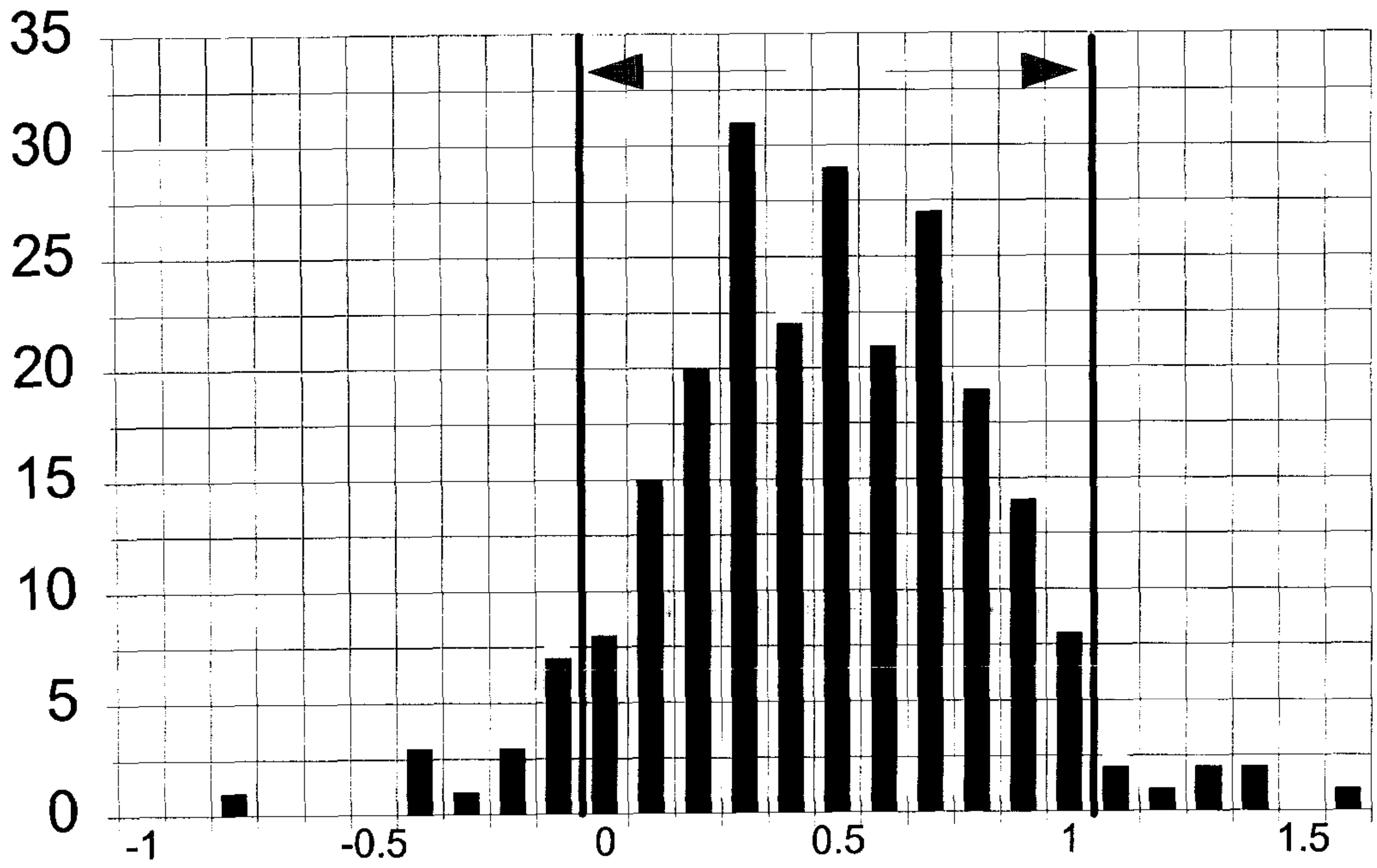
The Postal Rate Commission's

"Notice of Inquiry No. 4 On Mail Processing Variability"

February 6, 1998

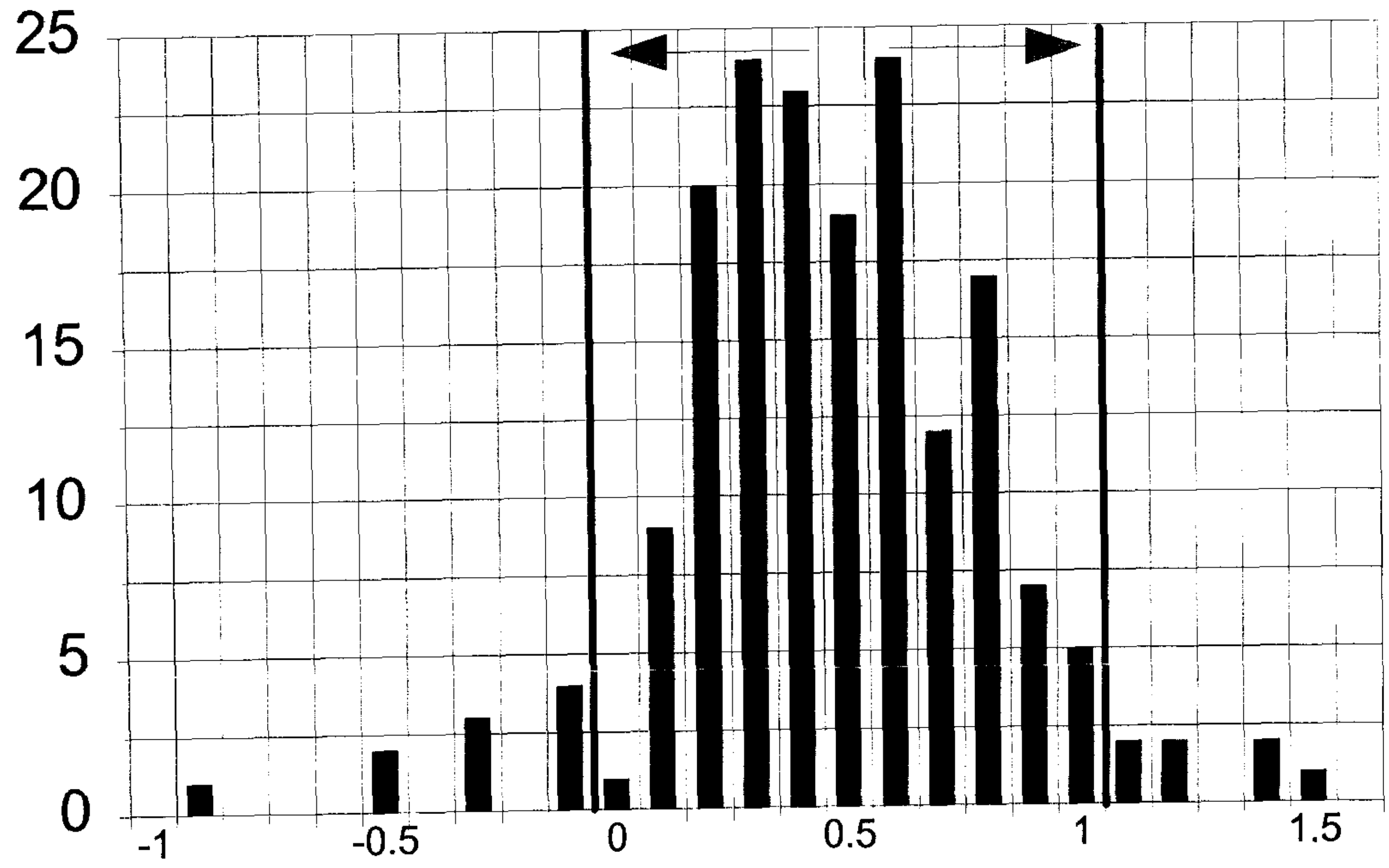
Manual Letter Variabilities

Distribution Across Sites



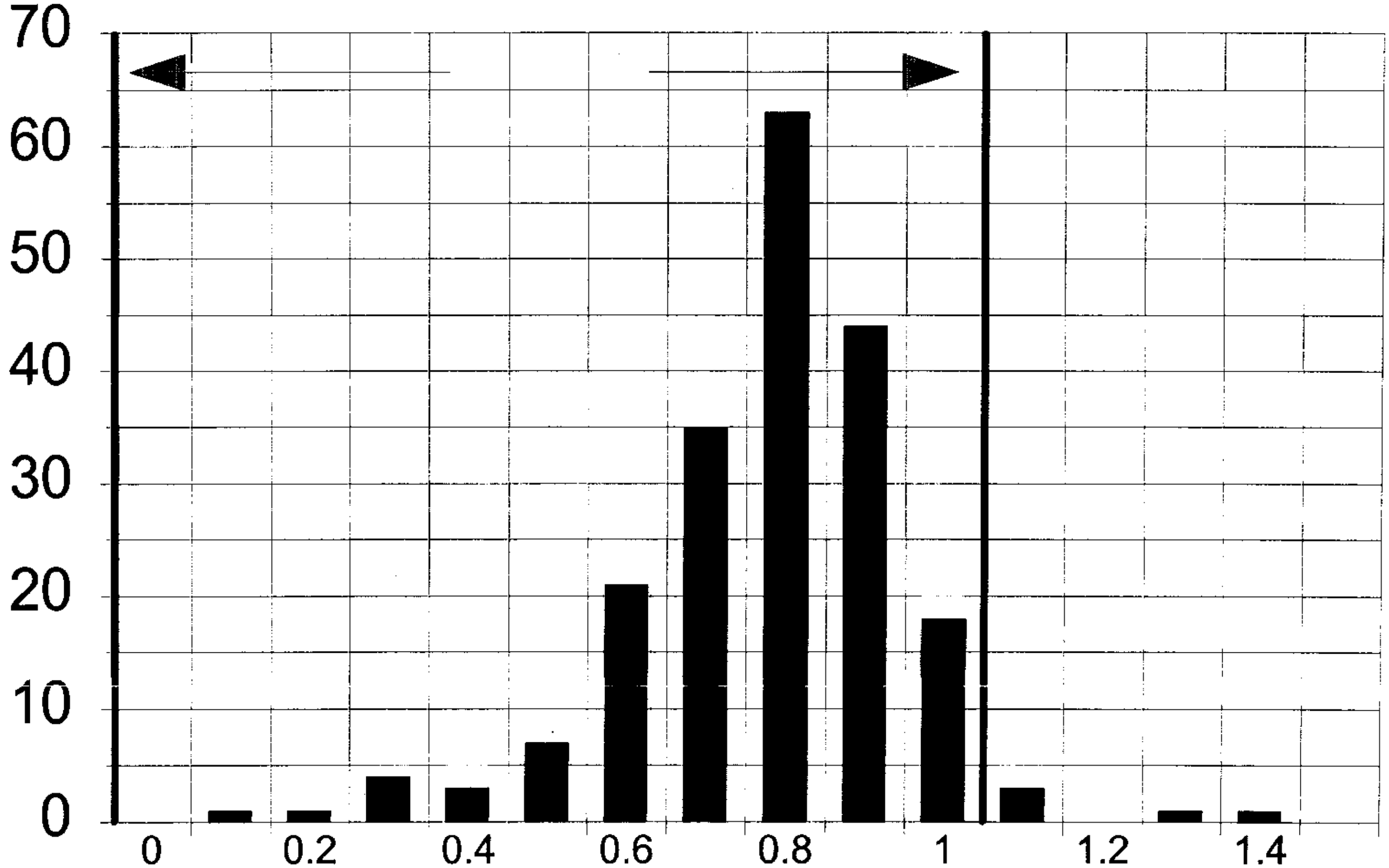
Manual Flat Variabilities

Distribution Across Sites



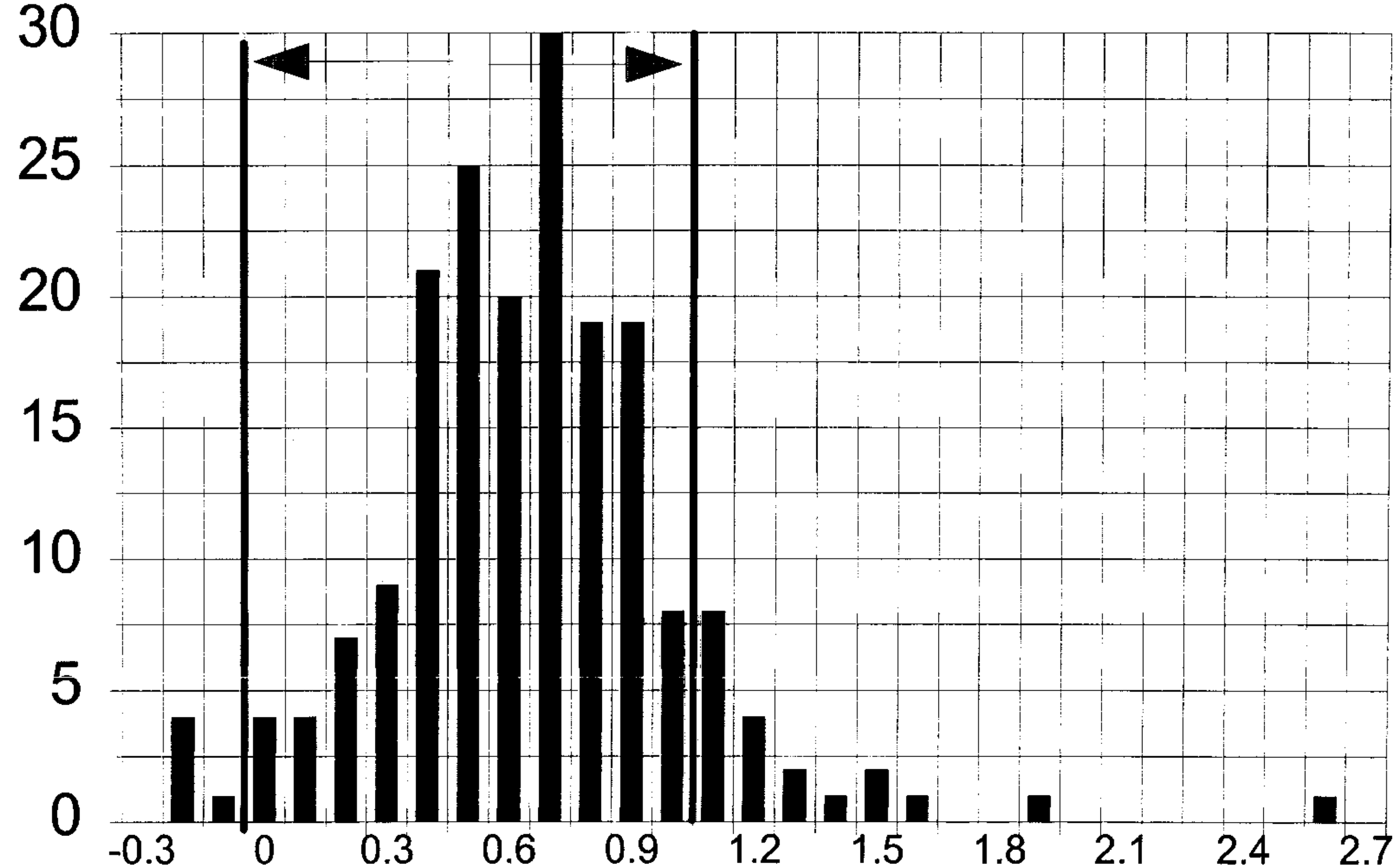
LSM Variabilities

Distribution Across Sites



OCR Variabilities

Distribution Across Sites



DECLARATION

I, Michael D. Bradley, declare under penalty of perjury that the foregoing statement is true and correct, to the best of my knowledge, information, and belief.

Michael D Bradley

Dated: Feb. 6, 1998

CERTIFICATE OF SERVICE

I hereby certify that I have this day served the foregoing document upon all participants of record in this proceeding in accordance with section 12 of the Rules of Practice.

Susan M. Duchek

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260–1137
(202) 268–2990; Fax –5402
February 6, 1998