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

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

RESPONSES OF MAGAZINE PUBLISHERS OF AMERICA, INC.
WITNESS CROWDER TO INTERROGATORIES OF THE
UNITED STATES POSTAL SERVICE (USPS/MPA-T5-23-28)

(July 3, 2000)

The Magazine Publishers of America hereby submits the responses of witness Crowder to interrogatories USPS/MPA-T5-23-28, filed on June 19, 2000. Each interrogatory is stated verbatim and is followed by the response.

Respectfully submitted,

James R. Creegan twm

James R. Creegan
Anne R. Noble
Counsel
Magazine Publishers of America, Inc.
Suite 610
1211 Connecticut Ave., N.W.
Washington, D.C. 20036
(202) 296-7277

USPS/MPA-T5-23. Consider the following regression model:

$$Y_i = \beta X_i + \varepsilon_i$$

Please recognize that this is the population regression equation in which β is the parameter estimate which captures the deterministic portion of the relationship between Y and X, not an estimated coefficient. Please also recognize that ε is the stochastic disturbance in the regression model and is not the residual from a least squares regression. This means that if β were known, then for any given value of X_i and ε_i, the corresponding value of Y can be calculated without error.[1]

In this model, Y is the dependent variable, X is the independent variable and ε is a zero mean, constant variance stochastic error term. Let the value of the actual (not estimated) β be equal to 0.45.

a. Confirm that there is no intercept in this model. If you cannot confirm, please provide the analytical expression for the intercept.

b. Confirm that if the value for X were zero, the model would predict a value of zero for Y. If you cannot confirm, please provide what you think is the correct prediction of Y given a value of X of zero and the mathematical formula underlying that prediction.

c. Confirm that β is the actual, not estimated, slope of the regression line and that the slope is equal to 0.45. If you do not confirm, provide what you think is the slope of the regression line and the mathematical formula underlying that slope.

d. You are given the following values for X and ε. Confirm that the actual values of Y generated by this model (with β=0.45) for these values of X and ε are as shown in the following table. If you do not confirm, please provide what you think are the true values for Y given these values for X and ε and the mathematical formula underlying those values.

ε	X	Y
3.0224	122	57.9224
-14.1576	455	190.5924
0.244257	177	79.89426
1.276474	289	131.3265
-12.257	550	235.243
1.733133	113	52.58313

---

1 For a further discussion of the population regression equation, see, William H. Greene, Econometric Analysis, Macmillan, 1993 at 143.

-2.18359	399	177.3664
-0.23418	255	114.5158
1.095023	446	201.795
-11.548	337	140.102
6.9887	108	55.5887
2.278	111	52.228
-1.84691	140	61.15309
-0.97763	155	68.77237

e. Confirm that an ordinary least squares regression would (including an intercept) of Y on X (as shown in d., above) would yield the following coefficient and t-statistics (within two decimal places):

	Intercept	β
Estimate	5.913	0.4201
t-statistic	2.409	51.2396

If you do not confirm, please provide what you think the estimated coefficients and t-statistics are and attach regression output supporting your results.

RESPONSE:

a. Confirmed.

b. Confirmed.

c. Confirmed.

d. Not confirmed. The constructed regression from the fourteen observations yields parameter estimates for a different model. The estimates and upper and lower limits at the 95 percent confidence level for the parameter estimates are:

	Intercept	β
Estimate	5.913	0.420
Upper Limit	11.262	0.438
Lower Limit	0.564	0.402

The assumed true values of zero for the intercept and 0.45 for β are outside the confidence bands indicated for both variables. Therefore, we are dealing with a different model. The problem is with the assumed values for ε in the regressions. The sum of the given ε values is -26.567, far distant from the true expected value of

zero (the average is -1.898), with which you premised your question. Since the confidence bands around the estimated parameter values exclude your assumed true values, the residuals from the regression must be estimates for true ε values different than the values you have assumed.

To see how changes affect the parameter estimates, I assumed zero ε values for the second and fifth observations instead to produce a revised sum of -.152 (the average is -0.011). I also changed the corresponding Y values only and then ran a regression on the revised data. The coefficient estimates, 95 percent confidence bands and t statistics are as follows:

	Intercept	β
Estimate	2.349	0.441
Upper Limit	7.045	0.457
Lower Limit	-2.347	0.425
t Statistic	1.090	61.290

Note that the t statistic for the intercept requires acceptance of the null hypothesis (zero value). Also the 0.45 true coefficient estimate you assumed is within the confidence band.

e. Confirmed. Also see my response to (d) above.

USPS/MPA-T5-24. Consider a simple route with five SDR stops. Further suppose that the "true" load time (as you define it in footnote 1 on page 2 of your testimony) at a stop is given by:

$$LT = 3 + .5 * V - .0002 * V^2$$

a. Given that each of the 5 stops have the following volume, confirm that the "true" load time at each stop is as shown below. If you do not confirm, please provide what you think the "true" load time for each stop would be and the mathematical formula supporting that calculation.

Stop	Volume	Load Time
1	5	5.450 seconds
2	7	6.402 seconds
3	3	4.482 seconds
4	8	6.872 seconds
5	0	0.000 seconds

b. Please confirm that the total load for the route is 23.206 seconds. If you do not confirm, please provide what you think the route's load time is along with a mathematical formula supporting that conjecture.

c. Now suppose that all of the existing volume stays on the route, but two additional pieces are added. Further suppose that they both go to stop 5, the previously uncovered stop. Confirm that the load time at stop 5 rises to 3.992 seconds. If you do not confirm, please provide what you think that additional load time is along with a mathematical formula supporting that conjecture.

RESPONSE:

a. Confirmed subject to the following stop load time equation:

$$LT = 3 * B + .5 * V - .002 * V^2,$$

where B = 0 when V = 0, and B = 1 when V > 0. The modification is required to show that the incurrence of stop load time is conditional on positive stop volume (i.e., conditional on the need to load/collect at the stop). Also please see my response to USPS/MPA-T5-1(g).

b. Confirmed, subject to the sum of the load times given by the modified stop/ load time equation in a. above.

c. Confirmed. Stop load time for stop 5 rises from zero to 3.992 seconds.

USPS/MPA-T5-25. Please refer to footnote 1 on page 1 of Appendix B to your testimony. There you indicate that you will use, in your regression equation, possible deliveries as a proxy variable for the two missing workload variables, "route volume and number of actual deliveries."

a. Please provide the econometric conditions under which a variable can serve as a effective "proxy" for an omitted variable. Include citations to the econometric literature to justify your statement of conditions.

b. Please provide a list of criteria you use in choosing a particular proxy variable.

c. If the proxy variable is not correlated with the omitted variable, can it serve as an effective proxy variable? Why or why not?

d. Is "possible deliveries" a better proxy for "route volumes" or for the "number of actual deliveries." Why?

e. Would it be feasible to use possible deliveries as a proxy for actual deliveries?

f. Are there any criteria under which a variable could not serve as a proxy variable for an omitted variable?

g. Are the econometric qualities of a regression with an omitted variable affected bythe quality and choice to the proxy variable? If so, please provide the qualities that are affected. If not, please explain why one is not free to use any variable as a proxy variable with no impact on the estimated regression.

RESPONSE:

a. When the available variable is known or expected to be correlated with the missing variables then the available variable can be used as a proxy for the missing data. The relationship between the dependent variable and the proxy variable need not be directly causal. For example, time is often use as a proxy to capture the effects of technology changes on an otherwise static analysis of economic variables.[2] However, no one would claim that changes in time cause changes in the dependent variables, just as no one should claim that changes in possible deliveries cause changes in cost. If a statistically significant relationship between the dependent

---

2 See also Pindyck & Rubinfeld, *Econometric Models & Economic Forecasts*, McGraw Hill, 1991, p. 413 - 416.

variable and the proxy variable exists, then the inference should be that there is statistically significant relationship between the dependent variable and the causal variables for which the proxy is used.

b. When one or more of the causal variables are not available, then a proxy variable which is correlated with the missing data can be used.

c. No. The correlation is required in order infer causality between the missing variable and the dependent variable.

d. I did not determine whether possible deliveries was correlated better with volume or actual deliveries in a statistical sense, nor was that necessary to assess the results from the MPA regression model. The important point is that possible deliveries is expected to be correlated with the two causal variables affecting load time a priori. This expectation formed the basis for the regression model form. The expectation was effectively evidenced with the statistically significant load time coefficients for the possible delivery variables.

e. Only if the volume variable is omitted. If the volume variable is part of the regression equation, then possible deliveries acts as a control variable to allow the volume effects on load time to be accurately determined. Also see my response to USPS/MPA-T5-2 (b) (3).

f. Yes, if the variable is not correlated with the omitted variable.

g. Yes. If there is no or poor correlation between the proxy and the omitted variable, then I would expect a lower R^2 because of larger unexplaned variations in the dependent variable caused by the missing causal (omitted) variable. Also the coefficients in the causal variables included in the regression may or may not be biased depending on correlation between them and the missing variable.

USPS/MPA-T5-26. Please refer to the regression equation presented on page 4 in Appendix B to your testimony.

a. How long did it take you (in person hours) to estimate the regression?

b. Confirm that you (or others directed by you) did not run any other regressions on the ES data other than the one appearing on page 4 of Appendix B. If you do not confirm, please provide a listing of all other regression equations that you ran on the ES data long with the program listing and logs.

c. Did anyone assist you in in estimating the regression on page 4 or any other regresssions on the ES data? If so, please provide the names of the people who assisted you, the qualifications of those individuals, the type of assistance they provided you and the amount of time they assisted you.

d. Please provide the date when you started the estimation process that led to the regression on page 4 in Appendix B to your testimony and the date when that work was completed.

RESPONSE:

a. In total my staff and I spent approximately 80 hours in reviewing various aspects of the relationship between the load time proportions, possible deliveries, and mode of delivery, discussing and developing the concept, preparing the data, running the regression, and evaluating the results fully.

b. Confirmed.

c. Yes. Ms. Lindsay Turpin prepared the data used for the analysis. Dr. William Miller, Mr. Fred Kelsey, and I reviewed various aspects of the data that led us to the regression analysis. Dr. Miller developed, ran, and evaluated the regression. I estimate that Ms. Turpin spent about 10 hours, Mr. Kelsey spent about 30 hours, I spent about 15 hours, and Dr. Miller spent about 28 hours.

Dr. Miller has an M.B.A. in Finance, an M.A. in Economics and a Ph.D. in Economics and has over twenty years' experience assessing the economic and financial performance of government and commercial programs, projects,

and activities. He has been associated with TRANSCOMM since 1987 and has been a key investigator in TRANSCOMM's postal and other activities.

Mr. Kelsey has an M.S. in Applied Science and an M.B.A. in Finance and over thirty years' experience in the analyzing utility demand, costs, and rates. He has been associated with TRANSCOMM since 1989 and has been a key investigator in TRANSCOMM's postal and other activities. This experience includes considerable statistical analysis of large data sets.

Ms. Turpin is a research assistant with Project Performance Corporation (PPC) and has worked on a variety of projects while at PPC and has experience in managing large data sets like the one in USPS LR I-163.

d. The process began in late March and ended April 25th.

USPS/MPA-T5-27. Please provide a list of all published econometric studies that you have authored or coauthored.

RESPONSE:

None. However, over the years, Dr. Miller, Mr. Kelsey, and I have had considerable practical experience in developing and reviewing proprietary econometric models used to estimate either demand or cost characteristics for commercial and governmental clients. Dr. Miller and I have also previously studied econometric modeling issues associated with Postal Service city carrier out-of-office costs.

USPS/MPA-T5-28. Please refer to the regression presented on page 4 of Appendix B of your testimony. For that regression please provide:

a. A general description of the program that estimated that regression including the objectives of the program and the processing tasks performed.

b. Any methods and procedures employed, along with a listing of input and output data, in machine readable form.

c. A listing of the source code in hardcopy and machine readable form.

d. For all input data:

   1. Designation of all sources for such data.
   2. Explanations of any modifications to such data made for use in the program.

e. Definitions of all input and output variables or sets of variables.

f. A description of input and output data file organization.

g. For all source codes, documentation sufficiently comprehensive and detailed to satisfy generally accepted software documentation standards appropriate to the type of program and to its intended use in the proceedings:

h. All pertinent operating system and programming language manuals.

i. If the requested program is user interactive, a representative sample program run, together with any explanation necessary to illustrate the response sequence.

j. A presentation of the economic theory underlying the economic study.

k. A complete description of:

   a. The econometric model(s)
   b. The reasons for each major assumption and specification.

l. The definition of the variables selected and the justification for their selection.

m. For any alternative model whose computed econometric results influenced the choice of the preferred model:

   1. A statement of the reasons for rejecting the alternative.
   2. An identification of any differences between that alternative and the preferred model with respect to variable definitions, equation forms, data or estimation methods, and

3. The computed econometric results for that alternative.

n. For every econometric technique used in the estimation process and the reasons for selecting each:

1. A referrence to a detailed description in a text, manual, or technical journal, or
2. A description and analysis of the technique that is sufficient for a technical evaluation.

o. A complete report of the econometric results, including, where applicable,

1. coefficient estimates,
2. standard errors and t-values,
3. goodness of fit statistics,
4. other appropriate test statistics,
5. the variance/covariance matrix of the estimates,
6. computed residuals for results computed from samples composed of fewer than 250 observations.

p. Descriptions of all statistical tests and hypotheses and the results of such tests.

RESPONSES:

a. The Excel regression program included as part of that software was used for the analysis.

b. See MPA-LR-6.

c. The source code is part of the Microsoft Excel 97 program and is non-extractable.

d. The load time data used for the analysis were from the ES data set in USPS LR I-163. For each of the 336 routes used by witness Baron to develop out-of-office time proportions, daily load tallies were averaged. The possible delivery and mode of delivery data were from USPS LR I-219. Please see Appendix B of my testimony.

e. Please see pages 2 and 3 of Appendix B.

f. The data used for the regression are arranged in tabular format in MPA-LR-6.

g. Please see my response to (c) above.

h. Please see my response to (c) above.

i. We did not run any sample programs. The only analysis run was the regression analysis documented in Appendix B.

j. Please see page 1 of Appendix B.

k. Please see Appendix B.

l. Please see Appendix B.

m. Given the limitations of the data we had to work with, the model presented was always our preferred model. No other model influenced the choice of our preferred model.

n. The regression equation was developed using ordinary least squares. This method provides the best linear unbiased estimators under the classical assumptions. Also please see chapters 4 and 5 from Pindyck & Rubinfeld, *Econometric Models & Economic Forecasts*, McGraw Hill, 1991.

o. 1.-4. Please see Appendix B and MPA-LR-6.

5. The variance/covariance matrix was not computed.

6. There were no residuals computed as part of the regression analysis.

p. The regression resulted in t-statistic values requiring rejection of the null hypothesis (zero value) for the intercept, all residential possible delivery types, business centralized deliveries, and two delivery mode dummy variables. Please see Appendix B (page 4) to my testimony for these t values.

DECLARATION

I, Antoinette Crowder, declare under penalty of perjury that the foregoing answers are true and correct, to the best of my knowledge, information, and belief.

ANTOINETTE CROWDER

Dated: July 3, 2000

CERTIFICATE OF SERVICE

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

Thomas W. McLaughlin

July 3, 2000