

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

POSTAL RATE AND FEE CHANGES, 2006

Docket No. R2006-1

RESPONSE OF THE UNITED STATES POSTAL SERVICE IN OPPOSITION TO GCA
MOTION TO STRIKE PORTIONS OF THE POSTAL SERVICE'S REPLY BRIEF
(January 17, 2007)

On January 10, 2007, GCA filed a motion to strike portions of the Postal Service's reply brief. The Postal Service hereby responds in opposition to that motion. The motion contains factual statements that are obviously incorrect, distorts both the record and the Postal Service's reply brief, and includes arguments on the merits that extend well beyond those necessary to address the relief sought by the motion. The materials from the econometrics literature contested by the motion appropriately respond to statements made in GCA's initial brief which attempt to characterize the econometrics literature as a whole, and which therefore invite responsive presentations from the econometrics literature. The motion to strike should be denied.

The motion to strike is ostensibly triggered by the citation and quotation in the Postal Service's reply brief of materials from two econometrics textbooks. Two sections of the GCA motion (Section I from pages 2-5, and Section II from pages 6-9) appear to make the argument why GCA (erroneously) believes that reference to those materials should not have been included within the reply brief. A third section, however, (Section III from pages 9-11) is simply an untimely attempt by GCA to address broader issues, beyond the scope of any questions raised by the inclusion or exclusion of the portions of the two econometrics texts. Apparently regretting its decision not to bother with its own

reply brief, GCA is now inappropriately attempting to use its motion to strike as a vehicle to respond to other arguments made in the Postal Service's reply brief. While many parties would presumably appreciate the opportunity to present sur-reply arguments, the procedural schedule does not allow it, and a rate proceeding could easily become unmanageable if every party claimed the right to get in the "last word."¹

GCA having injected such arguments into Section III of its motion, however, the Postal Service is compelled to address them in its response. This discussion, though, merely underscores the counterproductive nature of GCA's approach to the process by which the Postal Service and the Commission ultimately strive to achieve the best possible volume forecast as a necessary building block of the postal ratemaking process. GCA instead views the process as an abstract exercise to argue about the standards that academic econometricians would prefer to see in an ideal world.

The shell game that GCA is playing in Section III on pages 9-11 of its Motion is to take statements out of context in an attempt to surreptitiously equate a model that theoretically "might" be wrong with a model that actually is wrong. On page 9, GCA patches together two parts of a sentence from the Postal Service's reply brief:

Removing any doubt as to their position, the Postal Service states that "whether witness Thress 'might' have chosen the wrong model" is "an abstract question." (*Id.* at 40) (emphasis supplied).

In full, the paragraph in which that sentence appears in the Postal Service's brief reads:

¹ If, however one were seeking to identify egregious examples of parties truly stretching the bounds of reply brief practice to the breaking point, one need look no further than the OCA's unabashed attempt to file additional computer files in support of new empirical analysis in conjunction with its reply brief on mail processing costs. See NOTICE OF THE OFFICE OF THE CONSUMER ADVOCATE OF SUBMISSION OF FILES USED TO PRODUCE TABLE IN REPLY BRIEF (Jan. 4, 2007).

As explained in the rebuttal testimony of witness Thress (USPS-RT-2 at 60-61), and reiterated in the Postal Service's Initial Brief at page 61, model selection criteria (among the models estimated by witness Thress) is essentially a non-issue in this case because no other witness advocated in favor of any model other than the one selected by witness Thress. Neither Dr. Clifton nor Dr. Kelejian actually employs any alternate model selection criterion. Both were asked which of the models presented in LR-L-65 they would have chosen using the model selection criterion of their choosing, and both declined to identify any of Mr. Thress's alternate models as being superior to his preferred model. Tr. 24/8761, 29/9845. Moreover, Dr. Clifton in fact used Mr. Thress's chosen model as the starting point for his analysis of First-Class single-piece letters. Tr. 29/9845. Therefore, the abstract question of whether witness Thress "might" have chosen the wrong model has no practical application unless and until some other witness is prepared to come forward to select some other model he or she believes to be better, and to explain why on the basis of some other selection criterion. Since that did not happen in this case, the entire issue is fundamentally a red herring.

Postal Service Reply Brief at 39-40 (emphasis added, footnote omitted). The sentence on which GCA wishes to focus is the penultimate one in the paragraph, underlined in the above quotation. The word "might" in that sentence, appearing in quotation marks, was intended as a paraphrase of the testimony of Prof. Kelejian (GCA-T-5 at 14), that model selection via a minimization of mean square error "could" lead to an incorrect model.

The mere *possibility* that a model "might" or "could" be wrong is indeed an abstract concept. Review of Prof. Kelejian's candid response to a Postal Service interrogatory reveals why such an abstract suggestion is not particularly helpful to the ratemaking process:

USPS/GCA-8. On page 9 of the document originally filed as Appendix C to GCAT-1, Prof. Kelejian states that, in order for a model selection procedure to be valid "the complete set of models that are being considered must include the correct model." Please confirm that, as a practical matter, it is never possible to know the "correct model" in any empirical econometric work. If not confirmed, please explain fully.

RESPONSE:

In a sense, my statement was an obvious statement. It simply states that the true model can not be selected if it is not even considered. Also, formal statistic tests are based on modeling assumptions. These assumptions typically relate to a true model, which is the model which generates the dependent variable. In economics we often deal with complicated issues, as well as uncertainty. In practice, I think one can not be sure of anything.

Tr. 24/8760. By stating that, in practice, “one can not be sure of anything,” Prof. Kelejian’s testimony amounts to an acknowledgement that **any** model might or could be a “wrong” model. Moreover, Prof. Kelejian is also implicitly (if not explicitly) acknowledging that no statistical test ever devised can unfailingly detect “wrong” models, or unfailingly identify the “right” model. Therefore, GCA’s suggestion (Motion at 9) that the Commission “think long and hard about the implications of the Postal Service’s statements and the sentiments they portray,” is merely a suggestion to think long and hard about the testimony of its own expert witness.

GCA, however, misleadingly jumps from discussion on page 9 of a model that in some abstract sense “might” be wrong to a claim on page 10 that the “contention that a ‘wrong [expert] model’ constitutes substantial evidence is Orwellian; it robs words of their meaning.” (Brackets in original.) GCA, of course, leaves distinctly unclear exactly who it suggests is allegedly making this contention. The Postal Service, however, has two points to make in response. First, the notion that the Postal Service, through witness Thress, has failed to provide substantial evidence to support its forecasting methodology in general, or for single-piece First-Class Mail specifically, is absurd on its face.² As explained in the Postal Service’s Reply Brief at pages 41-42, the Commission

² Meeting the substantial evidence standard requires “such relevant evidence as a

has relied on the same basic forecasting methodology and the results of the same model selection criterion over numerous cases. Witness Thress has explained in great detail the careful process by which he chose his preferred single-piece model from among those he investigated. It is gross hyperbole for GCA even to suggest that the testimony of witness Thress is necessarily excluded from consideration as substantial evidence merely on the bare testimonies of its witnesses that the model selection procedure employed by witness Thress “could” fail to identify the model with the best fit, without any further effort to counterpropose a model with better fit. GCA is illogically seeking to impose on the Postal Service the obligation to prove a negative: that its preferred model is not wrong.

Second, there is nothing on the record to suggest that the model selected by witness Thress is, in fact, “wrong.” Witness Thress explained why that model is his choice among the models he examined. Contrary to what GCA is suggesting, it is perfectly reasonable to insist that someone wishing to postulate that a witness has *actually* (rather than merely *possibly*) chosen the “wrong” model from among those he investigated must be willing and able to identify some other model as better in order to sustain that accusation. Not only did Prof. Kelejian decline to take this step, but, as noted in the quote from the Reply Brief above, Dr. Clifton actually used the model

reasonable mind might accept as adequate to support a conclusion.” *Vieques Air Link v US Dept of Labor*, 437 F3d 102, 104 (1st Cir 2006). The cases cited by GCA on page 10 of its Motion do not address the substantial evidence standard; they instead address admissibility of testimony under the *Daubert* standard, and the Postal Service’s initial brief (footnote 5 on page 41) already showed how GCA has overstated the rigidity of the *Daubert* standard. The substantial evidence standard is also a flexible one, as “the possibility of drawing two inconsistent conclusions from the evidence does not prevent an administrative agency’s findings from being supported by substantial evidence.” *Domestic Securities Inc. v SEC*, 333 F3d 239, 249 (DC Cir 2003).

chosen by witness Thress as the starting point for his own research, as opposed to any other model tested by Mr. Thress.³ Tr. 29/9845. Attempting to portray the model of witness Thress as “wrong” is simply wishful thinking on the part of GCA. Rather than Orwell, the more appropriate literary allusion would appear to be something from *Alice in Wonderland*.⁴

In contrast with Section III of the motion to strike, the arguments raised in Section I (and, to a much lesser extent, Section II) are at least germane to the relief sought.

³ Nor can GCA excuse the failure of its witnesses to choose an alternative model by hiding behind the far-fetched possibility that testing might have shown that “none of Thress’s test models were any good.” See GCA Brief at 46. If GCA’s witnesses had actually conducted the series of tests that would be necessary to support that contention, one might have some sympathy for their reluctance to select an alternative. Instead, they conducted no tests at all, thus robbing statements regarding that absurdly remote possibility of any substance whatsoever.

⁴ In some sense, GCA’s focus on the term “wrong” to characterize forecasting models underscores the degree to which the entire discussion is removed from reality. Anyone who expects a forecasting model to be “right” or “correct,” if that means generating forecasts of every category of mail in every forecast quarter that are correct to the nearest thousand pieces of mail, simply does not understand forecasting. The objective in forecasting is to generate the best possible forecast, not one that anyone necessarily believes will achieve perfection. (The role of forecasting in a rate case is fundamentally different from other types of expert analyses used to assist in reaching the types of binary determinations much more common in court proceedings, such as did the defendant kill the victim, or did the defendant’s actions cause harm to the plaintiff. Conversely, forecasting models are entirely comparable to the myriad of other models used in the ratemaking process which may not achieve theoretical perfection, yet can and do produce reliable results for the purposes of fulfilling their functions within the ratemaking process.) Since a volume forecast is absolutely necessary for omnibus ratemaking under the current structure of the Commission’s rules, as a practical matter, it is incumbent upon parties wishing to argue that the forecast (or forecasting parameters) can be improved to propose and defend whatever adjustments they believe will yield a better and more reliable forecast. GCA started down that road in this case because, whatever its other faults, the testimony of Dr. Clifton was an attempt to present a concrete alternative model. For reasons that should be obvious, however, GCA now wishes to abandon a debate over the relative merits of the alternatives on the table, and instead suggest that the relative merits of the competing proposals are essentially irrelevant. Such an impractical proposal is not helpful to the ratemaking process, or to the Commission.

Nonetheless, those sections are rife with inaccuracies and distortions, and they do not justify the extraordinary relief requested.

For example, GCA claims:

The Postal Service introduces and quotes at length textual material from Greene and Kmenta (*id.* at 44-45) never submitted or even specifically referenced in the evidentiary record notwithstanding GCA's repeated efforts on cross-examination of witness Thress to have him identify any formal support for his least mean squared error test.

GCA Motion at 3 (underlined in original). With respect to the Greene material, this claim is demonstrably false. Not only were the salient pages of the Greene text referenced on the record, but that reference was quoted by GCA on page 44 of its initial brief, which in turn was quoted in footnote 7 on page 45 of the Postal Service reply brief, one of the two pages of the reply brief actually cited by GCA in the above misstatement. It was Prof. Kelejian himself who made the reference, citing page 159 among the pages of Greene in which "many formal procedures which relate to model section" are "nicely described." GCA-T-5 at 15. Furthermore, while not actually reciting the page numbers, witness Thress likewise referred to this portion of the Greene text during his cross-examination. Tr. 38/13111. The Greene material, therefore, was specifically referenced in the evidentiary record.

Moreover, the claims about repeated attempts on cross-examination to have witness Thress identify formal support for a Mean Square Error comparison are misleading.⁵ Oral cross-examination is a particularly poor vehicle to attempt to

⁵ GCA's Nov. 30th notice regarding oral cross-examination of witness Thress made no mention of any intent to review the econometric literature regarding model selection, made no mention of any intent to question the witness about specific portions of any edition of the Greene textbook, and did not request that the witness bring his own copies of the Greene textbook. GCA is thus precluded from claiming that the resulting cross examination can in any way be suggested to constitute an exhaustive discussion

determine what type of support exists in the econometrics literature for a particular approach.⁶ Just because a witness cannot provide any cite “off the top of [his] head” does not mean that no support exists. Written cross-examination, on the other hand, would have been a much more appropriate vehicle to attempt to develop the record on this issue. GCA claims that witness Thress had the opportunity to offer citations to the literature in response to discovery. Motion at 5.

In fact, however, in responding to discovery, a witness can only answer the questions that have been posed. In discovery, GCA did not ask whether any literature citations would support use of Mean Square Error, or why Mr. Thress did not select his

of the contents of the various editions of the Green textbook, much less econometric literature in general.

⁶ GCA’s unfairly truncated statements regarding the oral cross-examination illustrate the types of confusion that mode of exploration can engender. The Motion to Strike (at pages 1 and 3) appears to suggest that witness Thress only cited (at Tr. 38/13112) the 3rd Edition of the Greene text as his recollection of where Mean Square Error is recognized as a model selection criterion. GCA, of course, would prefer that his comments had been limited to that edition, since that is the edition that GCA produced for him to examine on the stand, and is, equally obviously, not the Greene edition from which the materials now in dispute came (i.e., the 5th edition). In fact, however, witness Thress was not so specific in his recollection, citing both the 3rd and the 5th editions at Tr. 38/13111 and 13112. When Mr. Thress states at the bottom of Tr. 38/13111 that his recollection of the relevant portion of the Greene text “goes through model selection criteria, and he includes mean square error and R squared as sort of the starting point, and then he moves on from there to explain some of these other things that Dr. Kelejian has suggested,” it was clear that he had something in mind from the Greene text. In fact, there are portions of both editions of the Greene text that relate to Mr. Thress’s statements on the record -- the material cited in the Postal Service’s Reply Brief from the 5th edition (on pages cited in Prof. Kelejian’s direct testimony), or corresponding material on pages 399-401 of the 3rd edition (the passage from the 3rd edition cited by the Postal Service in an interrogatory to Prof. Kelejian at Tr. 24/8757, in which Prof. Greene himself also notes the equivalence of minimizing MSE and maximizing Adjusted R-Square). GCA never asked witness Thress to identify the portion of Greene to which he was referring in his remarks at Tr. 38/13111. All GCA did was ask him to confirm what is in the limited part of the 3rd edition of Greene that GCA presented to him on the stand. Tr. 38/13116. When the Postal Service specifically objected to a question which appeared to be intended to foreclose any future opportunity to identify other relevant parts of Greene, counsel for GCA recast the question. *Id.*

models on the basis of a J-test, or a Cox test, or a Bayesian posterior odds approach. Instead, GCA focused its model selection questions on R-square, adjusted R-square, t-statistic, and coefficient estimates, asking Mr. Thress, for example, (Tr. 6/1229) to “confirm that the differences among the R-square in these models are so minimal that for forecasting purposes any one of these models could be used.” See Tr. 6/1229, 1234, 1238, 1241-42.⁷ It was in response to these questions that Mr. Thress explained why he prefers the metric which allows him to focus on the amount of unexplained variation (i.e., Mean Square Error, or MSE) instead of the flip side of the same coin, explained variation (i.e., adjusted R-square). Tr. 6/1231-32. But since GCA’s questions on written cross-examination did not mention Mean Square Error, there was absolutely nothing in them to cause Mr. Thress to believe that GCA was suggesting that Mean Square Error was inadequate, or that literature citations were necessary to explain the model selection criteria he favors (Mean Square Error) versus those criteria (R-Square, Adjusted R-Square, t-statistic, coefficient estimates) posited in the questions, much less those criteria not even mentioned (e.g., J-test, Cox test, etc.).

GCA, however, manages to muddle the state of the record on the issue of the relationship between Mean Square Error and Adjusted R-Square:

According to the newly proffered Kmenta material “the [adjusted R squared] criterion is *exactly equivalent* to the operational mean square error criterion.” (*Id.* at 46, n.10) (quoting Kmenta, *Elements of Econometrics* at 595 (2d ed.) (original emphasis). The Postal Service’s attorneys repeat and emphasize this point: “adjusted-R [squared] and

⁷ Thus, GCA posed a request to “explain fully why you did decide to choose model #23 over model #7, since it appears that the latter model has an essentially equivalent R-square and a much higher t-statistic” (Tr. 6/1234), and to “confirm that considering the t-statistic and R-square or adjusted R-square, model #4 is superior to your chosen model #1” (Tr. 6/1241).

mean-squared error, as defined by witness Thress, are identical selection criteria.” (*Id.* at 46).

This is not a reiteration on brief of material testified to on the record. Indeed, manifest inconsistencies between the Postal Service’s statements on brief, and witness Thress’s record testimony, make clear the unsoundness and unfairness of the Postal Service’s attempt to introduce the post-record documentary evidence at issue. The Postal Service claims that the two selection criteria are “identical,” and “exactly equivalent.” But according to witness Thress’s record testimony they are not: “[b]ecause of limitations of ... adjusted R [squared], [his] preferred diagnostic measure for evaluating demand equations is mean-squared error.” (*Id.* at 46) (quoting witness Thress’s response to GCA/USPS-T7-10). It cannot make any meaningful sense to prefer one “identical” test over another. Nor can one test have “limitations” not present in its “exactly equivalent” counterpart. The disconnect between the Postal Service’s selected scholarly passages and the testimony they ostensibly support is patent. And because the Postal Service has only now identified the Greene and Kmenta matter after the record has closed it is impossible for GCA to fairly explore what is either Thress’s illogic, or his disagreement with Kmenta and Greene.

Motion at 4-5. GCA is simply wrong to claim that restatement of the proposition that Mean Square Error and Adjusted R-Square are identical “is not a reiteration on brief of material testified to on the record.” Examination of the complete passage of testimony by witness Thress in response to GCA/USPS-T7-10, already quoted in full in the Postal Service’s reply brief at page 46, is apparently necessary again:

“[T]he goal of econometric estimation is not to maximize the explained variation but to minimize the unexplained variation within a model. While these two goals are, in some sense, literally identical there is an important distinction. Improving the adjusted-R² value in an equation from 0.986 (Model Number 7) to 0.990 (Model Number 23, which is used by me to make volume forecasts in this case) increases the explained variation in the model by 0.4 percent. Yet, reducing the percentage of variance which is unexplained from 0.014 (1 – 0.986) to 0.010 (1 – 0.990) reduces the unexplained variation in the model by nearly 30 percent.

Because of these limitations of R² and adjusted-R² measures, my preferred diagnostic measure for evaluating demand equations is mean-squared error. Mean-squared error is equal to the sum of the squared residuals divided by the number of degrees of freedom. This is equivalent to the square of the standard error of the model and can therefore be thought of as measuring the variance of a model.”

Tr. 6/1232. Mr. Thress plainly acknowledges in the first two sentences in this quote that a selection criterion of minimizing Mean Square Error and a selection criterion of maximizing Adjusted R-Square are “literally identical.” Therefore, contrary to GCA’s claim, the Kmenta material is indeed reiteration on brief of a point testified to on the record.

GCA’s further claim of alleged “manifest inconsistencies” between the Kmenta material and the above testimony of witness Thress is easily refuted by even a cursory review of the interrogatory response. The distinction that Mr. Thress was making between MSE and adjusted R-Square is nothing more than the classic distinction between the perception of the glass as half empty versus the perception of the glass as half full. Mr. Thress was emphasizing, in effect, that when the comparison is actually between a glass that is 99 percent full and 1 percent empty, versus a glass that is 98 percent full and 2 percent empty, the change in perception based on the change in perspective becomes much more significant. (Recall that Mr. Thress was responding to a GCA request to “confirm that the differences among the R-square in these models are so minimal that for forecasting purposes any one of these models could be used.” Tr. 6/1229.) Yet if, for example, we expand the illustration to three glasses of water, the first 99 percent full (and 1 percent empty), the second 98 percent full (and 2 percent empty), and the third 97 percent full (and 3 percent empty), there is nothing in witness Thress’s response to suggest that looking at minimizing percent empty versus maximizing percent full would change the fact that, either way, the most full glass would always be selected first, and the least full glass would always be selected last. It is the difference in perception, not a difference in results, which Mr. Thress refers to as the

“limitation” with respect to maximizing R-Square, and which therefore causes him to prefer MSE. If GCA had any problems understanding the distinction that witness Thress was making, despite his affirmative statement that the two procedures are “literally identical,” it had ample opportunity to follow-up on his interrogatory response and conduct whatever exploration was deemed necessary. The Kmenta material is corroboration from the literature for statements made on the record by witness Thress.

In its initial brief in December, GCA claimed that “Witness Thress’s test for model selection has no basis in the literature or in the econometric community,” and that “[j]udged against the applicable literature, Mr. Thress’ technique of model selection was no better than throwing darts.” GCA Brief at 43, 47-48. Coming from a party that in July was urging Mr. Thress to confirm (Tr. 6/1229) that “the differences among the R-square in these models are so minimal that for forecasting purposes any one of these models could be used,” it is clear that GCA has had a major change of heart regarding model selection criteria. It is not surprising, therefore, that the record regarding this issue might not be as well developed as it could have been if GCA had maintained a consistent position throughout the proceeding. Nonetheless, the erroneous claims made by GCA in its initial brief clearly opened the door for the Postal Service to provide citations from the “literature” to show that GCA was grossly overstating its case. Contrary to GCA’s misstatement in its Motion, the material from the 5th edition of Greene is referenced on the record by GCA’s own witness as a literature source for relevant model selection criteria. Mr. Thress alluded to the Greene material during oral cross-examination as well. Contrary to GCA’s misstatement in its Motion, the Kmenta material is a source from the literature that reiterates a point made by Mr. Thress on the

record, and in no way conflicts with his testimony. The quotations from the two literature sources were properly included to assist the Commission in its evaluation of the claims made in GCA's initial brief. Moreover, even assuming *arguendo* that GCA should be afforded some opportunity to respond to those materials, GCA has already gratuitously used the Motion to Strike to respond to those materials, in addition to other portions of the Postal Service's Reply Brief as well. The motion to strike should be denied.

Respectfully submitted,

UNITED STATES POSTAL SERVICE

By its attorneys:

Daniel J. Foucheaux, Jr.
Chief Counsel, Ratemaking

Eric P. Koetting

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260-1137
(202) 268-2992, FAX -5402
January 17, 2007

CERTIFICATE OF SERVICE

I hereby certify that I have this date served the foregoing document in accordance with Section 12 of the Rules of Practice and Procedure.

Eric P. Koetting

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260-1137
(202) 268-2992, FAX: -5402
January 17, 2007