

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

Postal Rate and Fee Changes, 2006

Docket No. R2006-1

**MOTION OF THE GREETING CARD ASSOCIATION TO
STRIKE PORTIONS OF THE UNITED STATES POSTAL
SERVICE'S REPLY BRIEF DISCUSSING EXTRA-RECORD
ECONOMETRIC EVIDENCE**

(January 10, 2007)

INTRODUCTION

Throughout the course of the hearing, GCA witnesses Clifton and Kelejian have criticized witness Thress's First Class mail econometric model as it is the product of unsound methodology. In particular, Drs. Clifton and Kelejian faulted Thress for not using an accepted methodology to select his relied-upon model from some 23 experimental candidates. (GCA-T-1 at 34; GCA-T-5 at 15).

Witness Thress was unable to identify any formal basis for his least mean squared error test methodology.¹

¹ When cross-examined concerning his initial testimony, witness Thress could not identify any support for his test method, other than that was what he always did. When cross-examined concerning his rebuttal testimony, Thress claimed that the 3rd edition of William Greene's textbook, Econometric Analysis, supported his test methodology. (Tr. 38/13112). However, when Thress was handed a copy of the textbook he could not locate any passage supporting his least mean squared error test. He acknowledged that his experimental models were nonnested. (Tr. 38/13114). After reviewing the section in the Greene textbook titled "Choosing Between Nonnested Models" he admitted that the least mean squared error test was not one of the identified formal methods for choosing between nonnested models. (Tr. 38/13115-116).

Now, following the close of the record, the Postal Service seeks to rehabilitate its witness, and a substantial part of its case, by using its Reply Brief to introduce extra-record evidence. This is improper and a violation of GCA's due process rights. As shown below, the extra-record econometric text book passages set out in the Postal Service's Reply Brief are fundamentally inconsistent with witness Thress's own testimony, and seem otherwise inapplicable to Thress's test models.

The Postal Service cannot properly avoid the acid test of the hearing process by seeking to rely upon un-vetted extra-record evidence. All of the references in the Postal Service's Reply Brief to the improper extra-record matter should be struck. These passages are identified with particularity in Attachment A.

ARGUMENT

I. THE POSTAL SERVICE CANNOT INTRODUCE NEW EVIDENCE FOLLOWING THE CLOSE OF THE RECORD

On December 19, 2006, pursuant to Presiding Officer Ruling No. R2006-1/129, "the evidentiary record in Docket No. R2006-1 ... closed."² The Postal Service is not free to disregard this ruling. However, on January 4, 2007, following the close of the evidentiary record, the Postal Service identified and quoted passages from two documents not previously submitted in the record, *i.e.*, page 159 of the Fifth Edition of William Greene's textbook Econometric Analysis and pages 594 and 595 of the second edition of Jan Kmenta's textbook Elements of Econometrics. (Reply Brief at 45-46).

² With the exception of a joint objection not relevant to the instant matter.

The Postal Service's purpose in attempting to introduce (selected) portions of these documents into the proceeding is to rehabilitate its witness, Mr. Thress. In response to GCA's challenge that witness Thress's least mean squared error had no formal methodological basis in the econometric literature, the Postal Service coyly states that "some further discussion on why Mean Square Error comparisons are a formal procedure as witness Thress indicated ... may be helpful." (Reply Br. at 44). But what follows is not mere "discussion." 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.

When the evidentiary record closed, the state of this matter was clear: witness Thress had failed to identify any formal support for his test methodology. On August 9, 2006, GCA's counsel asked Mr. Thress to identify what econometric authorities supported the use of a mean squared error diagnostic to choose among competing econometric models. His answer was: "I don't have anybody off the top of my head." (Tr. 6/1333). When questioned again on this matter, on December 6, 2006, Mr. Thress claimed that the 3rd edition of the econometric textbook Econometric Analysis by William Greene identified the least mean squared error metric as a test for choosing between and among econometric models. (Tr. 38/13112). Upon further questioning Mr. Thress acknowledged that all or most of his 20 odd experimental models were "nonnested" models. (Tr. 38/13114). But when counsel for GCA handed witness Thress a copy of the 3rd edition of William Greene's Econometric Analysis and had him

review it, he acknowledged that the 3rd edition of the Greene textbook did not list a least mean squared error approach as an appropriate form of model selection. (Tr. 38/13116). Witness Thress acknowledged that the only tests Greene identified for choosing between and among nonnested models (i.e., the type of First Class econometric estimations models that Mr. Thress generated) were: the (1) J-Test; (2) the Cox test; and (3) a Bayesian test approach. (Tr. 38/13114-115).

Now, on brief, and after the record has closed, the Postal Service's lawyers attempt to provide the evidence their witness did not. The Postal Service's lawyers introduce statements from Greene and Kmenta as evidence concerning the use of an adjusted-R squared test. (Reply Br. at 45-46). 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 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.

Pursuant to the governing schedule, the Postal Service and witness Thress had no fewer than seven different opportunities to submit the Greene and Kmenta material into the evidentiary record: 1) as part of witness Thress's initial testimony; 2) in response to discovery; 3) on cross-examination concerning Thress initial testimony; 4) on re-direct of same; 5) as part of witness Thress's rebuttal testimony; 6) on cross-examination concerning his rebuttal testimony; and 7) on re-direct of same. Had the Postal Service and witness Thress done so, GCA would have had the opportunity to develop record evidence to the contrary or, at least challenge its relevance and probativeness through oral examination. There is nothing that excuses the Postal Service from its failure to avail itself of these opportunities. Having failed to properly and timely submit the Greene and Kmenta material in the evidentiary record, the Postal Service cannot properly do so now, and thereby defeat GCA's rights. The Postal Service is "not entitle[d] ... to introduce [evidence] at a stage in the proceeding at which the opposing party will not have an opportunity to respond." *Pittsburgh and Lake Erie R.R. v. ICC*, 796 F.2d 1534, 1544 (D.C. Cir. 1986).

II. THE POSTAL SERVICE'S INTRODUCTION OF POST-RECORD EVIDENCE IS PREJUDICIAL TO GCA

The newly proffered Greene material in the Postal Service's Reply Brief addresses the use of an adjusted R squared approach to assess the impact "when a variable is added to a model." (Reply Br. at 46) (quoting Greene at 159). The quoted passage goes on to state that "[t]he applicable result appears in Theorem 3.7; [adjusted R squared] does not rise when a variable is added to a model unless the t ratio associated with that variable exceeds one in absolute value." (*Id.*). Had the Postal Service timely and properly submitted the Greene material into the evidentiary record, GCA would have shown that Greene's adjusted R squared method does not apply to Thress's experimental situation.

In the first instance, Theorem 3.7, and thus the passage as a whole, relates to linear models. (See Greene at 26-35). Thress's 23 experimental models are not linear; they are loglinear models because they contain an ersatz Box-Cox transformation. (See, e.g., USPS-RT-2 at 62 [discussing "non-linear transformation of the internet variable"]). Further, Theorem 3.7 essentially relates to the importance of a single variable, or to a comparison of 2 models when one model has all the variables of the

other except that it also has one additional variable.³ Theorem 3.7 has no application where the two models differ because they have different sets of variables, as is the case with nonnested models. Witness Thress admitted on cross-examination that his 23 experimental models are nonnested. (Tr.38/13114). Indeed, the opening sentence of the quoted Greene passage indicates that the discussion which follows does not apply to nonnested model testing, i.e., “[t]he preceding discussion suggested some approaches to model selection based on nonnested hypothesis tests.” (Reply Br. 45) (quoting Greene at 159 [emphasis supplied]). There is record evidence as to Greene’s position with respect to choosing between nonnested models. Witness Thress admitted on the stand that Greene did not identify a least mean squared error test as a formally accepted test for choosing between nonnested models. (Tr.38/13115).⁴

³ A more general result is given in Kmenta (1986) on page 594.

Specifically, as in Greene, Kmenta considers the issue of selecting between two linear nested models. In Kmenta's case, one model has all the variables of the other as well as, perhaps, more than one additional variable. Kmenta gives an operational mean square error criterion which he demonstrates is equivalent to the R bar squared criterion for model selection. In this more general framework, Kmenta's results indicate that the operational mean square error criterion will select the model with the additional variables if an F statistic exceeds 1.0.

Equivalently, the hypothesis that all of the coefficients of the additional variables are zero will be accepted if the F statistic is less than 1.0. Kmenta then notes, on page 594, "This corresponds to a level of significance considerably greater than the conventional 5% or even 10%". This excessive level of significance noted by Kmenta is also implied by the results in Theorem 3.7 in Greene when models are selected by the R bar squared criterion. In that case one model differs from the other by only a single variable, and the smaller model is accepted if a t-ratio is less than 1.0 in absolute value. The implied level of significance is approximately .31, well beyond a typically acceptable degree of Type I error.

⁴ Thress testified concerning the 3rd edition of Greene but there is no relevant difference between the section in the 3rd edition titled “Choosing Between Nonnested Models” and Section 8.3 (“Choosing Between Nonnested Models”) in the 5th edition.

GCA would have shown that Kmenta describes the adjusted R squared approach as an “ad hoc” method under the heading of “Informal Model Selection.” (Kmenta at 598).

Further, as indicated above, GCA would have explored the inconsistencies involved in Thress’s preference for a least mean squared error model selection procedure, and Kmenta’s statement that “the [adjusted R squared] criterion” (which according to Thress is the same as the mean squared error approach) “is *exactly equivalent* to the operational mean square error criterion.” (Reply Br. at 46 n.8) (quoting Kmenta at 595).

GCA anticipates that in response the Postal Service will attempt to defend the application of the adjusted R squared test to Thress’s experimental nonnested models. But the time and place for that inquiry has passed: the evidentiary record is closed. The Postal Service cannot now fill the evidentiary hole in its case.

Nor can the Postal Service defend its extra-record reliance on the Greene and Kmenta matter on grounds of official notice. What is at issue are core adjudicative facts concerning technical econometric matters. Although witness Thress claimed to have relied upon a formally accepted test methodology, he was incapable of supporting that contention and all of the record evidence is to the contrary. The Commission cannot properly take official notice of an academic passage that on its face does not apply to Thress’s situation (*i.e.*, choosing between nonnested models) to find otherwise. Moreover, even under the Commission’s rule governing official notice, GCA must “be afforded an opportunity to show the contrary.” Commission Rule 3001.31 (j). In the

circumstances here, that would require nothing less than a re-opening of the evidentiary record with respect to this issue.

III. THE LACK OF SUPPORT FOR WITNESS THRESS'S METHODOLOGICAL APPROACH IS A SERIOUS ISSUE

In what can only be described as an astonishing and misguided proposition, the Postal Service claims that the entire issue of whether witness Thress relied upon a formally accepted method of test selection is irrelevant, and it makes no difference whether his model is wrong and unreliable. According to the Postal Service, “[d]ebates on model selection criteria are only relevant in rate proceedings in the context of selection from among real competing models, proposed by competing witnesses.” (Reply Br. at 49). 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). GCA respectfully suggests that the decision maker(s) think long and hard about the implications of the Postal Service’s statements and the sentiment they portray. The Postal Service claims, that unless the Commission relies upon some other witness’s elasticity model,⁵ it has no choice but to accept the results of witness Thress’s proffered measures of First Class elasticity no matter how bad or unsound the methodology employed in generating those results. This is not correct.

As the proponent of an order recommending an increase in rates the Postal Service bears the burden of proof. 39 U.S.C. § 3624(a); 5 U.S.C. §556(d); *Director, OWCP v. Greenwich Collieries*, 512 U.S. 267 (1994). The Postal Service has an

affirmative burden to present substantial evidence in support of its position. Where it fails to do so, as here, it “can: (a) credit GCA witness Clifton’s elasticity estimates; or (b) decide the issue of FCLM rates based upon the record evidence as a whole recognizing the deficiencies of the Postal Service’s case and discounting it accordingly.” (GCA Post Hearing Br. at 62).

The particular matter at issue, the Postal Service’s estimates of First Class letter mail elasticity, is significant. Pursuant to the Commission’s Rules of Practice the Postal Service is required to submit these econometric elasticity estimates as part of its case-in-chief. Numerous parties, including the Postal Service, rely upon witness Thress’s estimates of First Class elasticities to support, *inter alia*, the Postal Service’s proposed rates for First Class single piece mail and attendant allocation of responsibility for institutional cost coverage.

The contention that a “wrong [expert] model” constitutes substantial evidence is Orwellian; it robs words of their meaning. And as GCA explained in its Post Hearing Brief, several courts have held that as regards expert testimony in administrative proceedings the substantial evidence standard is informed by the Supreme Court’s *Daubert* – line of decisions. “‘Junk science’ has no more place in administrative proceedings than in judicial ones.” *Pasha v. Gonzalez*, 433 F.3d 530, 535 (7th Cir. 2005). Critical to determining whether a testifying expert has resorted to junk expertise is whether the expert has employed accepted and peer reviewed methodology. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 92-94 (1993). The record evidence

⁵ GCA does not reargue here the merits of witness Clifton’s elasticity study. It objects to the Postal Service’s attempt to introduce extra-record evidence in support of its witness, Mr. Thress.

establishes beyond all reasonable doubt that witness Thress did not employ formally accepted test methodology applicable for choosing between his 23 different nonnested experimental models.

The Postal Service objects to being held to a “narrow set of highly formal procedures,” and essentially claims (based on anecdotal, unvetted, extra-record and inapposite examples) that informal test selection methods are appropriate for the Commission’s purposes. (Reply Br. at 47-48). This is an argument nowhere advanced by its witness, Mr. Thress who sought (futilely) to defend his method of model selection as a formally accepted practice applicable to his method of elasticity estimation. (Tr.38/13112-113). It is understandable that witness Thress, unlike the Postal Service, did not attempt to defend his method of model selection as informal or ad hoc. The Kmenta treatise, in a passage not quoted by the Postal Service identifies the adjusted R squared test as an “ad hoc criteria” (at 598) and goes on to explain the unsoundness of informal model selection. “A researcher confronted by a list of regressors tries various combinations of variables until satisfactory results ... are obtained. This is known as ‘torturing the data until they confess.’” (Kmenta at 599). Kmenta goes on to explain that this “[d]ata mining is definitely unscientific and the results could be seriously misleading.” (*Id.*).

CONCLUSION

For all of the foregoing reasons, the Commission should strike the presentation and discussion of new evidentiary matter in the Postal Service's Reply Brief as identified specifically in Attachment A by black strike through, *i.e.*, all of the text beginning with "In the fifth edition of Greene" (Reply Br. at 45) through and including footnote 8 on page 46.

Respectfully submitted,

/s/ James Horwood

James Horwood

Peter Hopkins

Spiegel & McDiarmid

1333 New Hampshire Ave. NW, 2nd Floor

Washington, D.C. 20036

Date: January 10, 2007

Attachment A

witness Thress's "test for model selection has no basis in the literature or the econometric community." *Id.* In fact, such a selection criterion lies at the heart of Least Squares regression analysis, which provides the foundation and the framework for all of the econometric work in this case, not just that of witness Thress, and not just that relating to demand analysis.

The underlying premise of Least Squares Regression Analysis is the minimization of the sum of squared residuals. Obviously, a minimum mean-squared error model selection criterion flows directly from this. While certainly not endorsing it, even Prof. Kelejian states that "model selection via a minimization of a mean squared error is an intuitive thing to do." GCA-T-5 at 14. ~~In the fifth edition of Econometric Analysis by Dr. William Greene, which Mr. Thress cited (Tr. 38/13112) as a source for a mean-squared error selection criterion, chapter 8.4 starts on page 159 and is entitled "Model Selection Criteria".⁷ This section begins as follows (emphasis added):~~

~~"The preceding discussion suggested some approaches to model selection based on nonnested hypothesis tests. Fit measures and testing procedures based on the sum of squared residuals, such as R^2 and the Cox test, are useful when interest centers on the within-sample fit or within-sample prediction of the dependent variable. When the model building is directed toward forecasting, within-sample measures are not necessarily optimal. As we have seen, R^2 cannot fall when variables are added to a model, so there is a built-in tendency to overfit the model. This criterion may point us away from the best forecasting model, because adding variables to a model may increase the variance of the forecast error (see Section 6.6) despite the improved fit to the data. With this thought in mind, the adjusted R^2 ,~~

$$\hat{R}^2 = 1 - [(n-1)/(n-K)] \cdot (1 - R^2) = 1 - [(n-1)/(n-K)] \cdot [(e'e)/\sum_{i=1}^n (y_i - \hat{y}_i)^2]$$

⁷ In fact, the GCA, in their brief quotes their own econometric expert, Dr. Harry Kelejian, making specific reference to this very passage of Greene: "There are, of course, many formal procedures which relate to model selection. Some of these are nicely described in the econometric text by Greene on pages 152-160." GCA Brief at 44.

~~has been suggested as a fit measure that appropriately penalizes the loss of degrees of freedom that result from adding variables to the model. Note that R^2 may fall when a variable is added to a model if the sum of squares does not fall fast enough. (The applicable result appears in Theorem 3.7; R^2 does not rise when a variable is added to a model unless the t ratio associated with that variable exceeds one in absolute value.) The adjusted R^2 has been found to be a preferable fit measure for assessing the fit of forecasting models.~~

~~Perhaps GCA was confused by the fact that Dr. Greene did not explicitly mention the term "mean squared error". In fact, however, adjusted R^2 and mean squared error, as defined by witness Thress, are identical selection criteria.⁸ Witness Thress explained his preference for looking at mean-squared error as opposed to adjusted- R^2 in his response to GCA/USPS-T7-10:~~

~~"[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- R^2 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 R^2 and adjusted- R^2 measures, my preferred diagnostic measure for evaluating demand equations is mean-squared error.~~

~~⁸ Using Dr. Greene's nomenclature, mean-squared error (MSE), as defined by witness Thress, is equal to $(e'e)/(n-K)$, so that, $R^2 = 1 - \text{MSE} \cdot [(n-1)/\sum_{i=1}^n (y_i - \hat{y}_i)^2]$. For a comparison across equations for which the sample period and dependent variable are the same (as is the case in Mr. Thress's USPS-LR-L-65), the term $[(n-1)/\sum_{i=1}^n (y_i - \hat{y}_i)^2]$ will be constant across all equations and maximizing R^2 will be equivalent to minimizing MSE. The same point is made in another prominent econometrics text, Elements of Econometrics, Jan Kmenta, Second Edition, in which Prof. Kmenta's section 11-10 on "Model Choice" first explicitly lays out the Mean Square Error Norm as an "alternative criterion [for model choice that] involves the comparison of mean square errors of the two models." *Id.* at 594. Prof. Kmenta next lays out, as yet another alternative, the " R^2 criterion," and then demonstrates that "the R^2 criterion is *exactly equivalent* to the operational mean square error criterion. *Id.* at 595 (emphasis in original). Thus, both Prof. Kmenta (directly) and Prof. Green (less directly) confirm that Mean Square Error comparisons may properly play a role in model selection.~~