Postal Rate Commission
Submitted 10/19/2006 4:05 pm
Filing ID: 54289
Accepted 10/19/2006

BEFORE THE
POSTAL RATE COMMISSION
WASHINGTON, D.C. 20268B0001

POSTAL RATE AND FEE CHANGES, 2006	Docket No. R2006-1

RESPONSES OF GREETING CARD ASSOCIATION WITNESS HARRY KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE (USPS/GCA-1-12)

(October 19, 2006)

The Greeting Card Association ("GCA") hereby provides the responses of Harry Kelejian to the following interrogatories of the United States Postal Service filed on October 2, 2006: USPS/GCA-1-12, which were filed as institutional requests and have been redirected to Professor Kelejian. Each interrogatory is set out verbatim followed by the response.

Respectfully submitted,

/s/ James Horwood
James Horwood
Spiegel & McDiarmid
1333 New Hampshire Ave. NW
2nd Floor
Washington, D.C. 20036

Date: October 19, 2006

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-1. Please provide a list of all organizations for whom Prof. Kelejian has had occasion to estimate empirical econometric equations. (Within that list, please specifically identify which occasions involved demand analysis and the empirical estimation of price elasticities.) Please provide a description of the econometric techniques which he employed on these occasions.

RESPONSE: My consulting work goes back to the 1970s. I do not remember the details of much of it. As I indicate below, I generally give econometric advice concerning model formulation, estimation, and testing. Some of my consulting work led to academic papers which were published. References are given in my vita.

From my vita it should be clear that I have published a lot of papers, and have consulted for a variety of institutions. Concerning my consulting activities, the institutions that the work I did "most likely" involved, in one way or the other, the estimation of econometric models were

1. W.R. Grace and Company, 1972. This was a long time ago. I really do not remember it except that I did econometrically estimate some equations.

2. Econ. Inc., Princeton NJ, 1973-1975. I worked with D. Bradford of Princeton University for them on value of information problems and its estimation as it relates to remote censoring of crops, such as wheat. Among other things, we used Bayesian analysis and quadratic dynamic programing. We also estimated demand equations for wheat, and other related things.

3. C&P telephone, Washington DC 1976. Generally I gave econometric advice relating to modeling. I also served as an expert witness in a rate case for them. That case involved the estimation of elasticities and an estimation problem. However, at this point I really do not remember details about it.

4. Arthur Young & Company, 1977-1980. I gave econometric advice concerning modeling and estimation. I do not remember details.

5. The Federal Trade Commission, Washington DC, 1979. I really do not remember the nature of my role with them, nor do I think my work with them involved estimating demand elasticities.

6. AT&T Communications, 1980-1990. My work with them was mostly of a theoretical sort, although we did estimate some models. We worked a good deal on model formulation involving a random parameter approach. I even published a few papers with two of their researchers, namely J. Gatto and S. Stephen. These papers were:

(A) "A Random Coefficient Qualitative Choice Model of the Demand for Telecommunications", (with J. Gatto and S. Stephan), in Economic Letters, Vol. 35, 1991, pp. 45-50.

This is a theoretical paper that does not involve an empirical model.

(B) “A Note Concerning Specifications of Interactive Random Coefficient Regression Models”, (with J. Gatto and S. Stephan), in Journal of Econometrics, Vol 47, 1991.

This is a theoretical paper that does not involve an empirical model.

(C) “Interstate/InterLata Telecommunication Demand Modelling”, (with J. Gatto and S. Stephan), Review of Business, Fall 1989, Vol. 11, No. 2, pp. 25-31. Among other things, in this paper we estimated telecommunication demand models using a random parameter approach in a panel data setting. Essentially, our objective in this study was to asses the feasibility of developing telecommunication demand models on a disaggregated level.

(D) “Stochastic Generalizations of Demand Systems with an Application to Telecommunications”, (with J. Gatto and S. Stephan), Information Economics and Policy Vol. 3, 1988, pp. 283-309.

In this paper we considered a stochastic generalization of a system of telecommunications demand equations via a random parameter approach. In this approach there are stochastic generalizations of various restrictions that one might consider in a demand system, such as Slutsky symmetry and a weak separability restriction. The generalization of the Slutsky symmetry condition was in terms moments, rather than as exact restrictions. This paper also contains empirical results.

7. Booz, Allen, and Hamilton, Inc. 1980. I gave econometric advice, but do not remember details.

8. The World Bank, 1982. I mostly gave econometric lectures there and interacted with some of the researchers on their modeling.

9. Wilkes, Artes, Hendrick, and Lane, 1983. Again, I gave econometric advice and do not even remember if we estimated models together.

10. Glassman-Oliver 1988-1990, 1994-1995. I gave econometric advice concerning modeling and estimation.

11. Association of American Railroads 1991. I really do not remember any details concerning this.

12. D.C. Public Service Commission, 1991-1992. I gave econometric advice, but do not remember details. I also testified on their behalf, but I really do not remember details.

13. InterAmerican Development Bank, 1998-1999. I gave econometric advice and estimated genie coefficient equations. I do not recall details of that work which never came to anything.

14. Interindustry Economic Research Fund, Inc. 2006. I gave econometric advice relating to demand modeling for various medical services. We gathered data and tabulated various descriptive statistics but never got to the formal model estimation.

15. Washington Economic Consulting Group, 2006. The work I am now doing.

16. I am now working with two people at the The Bank of Greece, in Athens Greece, 2006, namely George Tavlas and George Hondroyiannis. My work with them relates to modeling and estimating equations which describe contagion in foreign exchange markets. The econometric approach being taken is spatial econometrics. I have worked with them before on the same topic using the same modeling.

17. As my vita indicates, I worked with the US Army Corps of Engineers from 1989-1999. During that time I mostly worked with a researcher there, namely Dennis Robinson. We published a number of papers together. Looking at my vita,

(A) The paper with D. Robinson and R. Vigil (2005) related to characteristics of cities. It did not contain estimated demand equations, but we did estimate a poverty rate equation using spatial econometric techniques.

(B) The paper with D. Robinson (2004) was a theoretical paper.

(C) The paper with D. Robinson (2000) relates to navigation infrastructure and involved, among other things, the estimation of final demand equations for selected industries. In doing this we used a cubic in time to proxy for a missing variable.

(D) The paper with D. Robinson in (1998) was essentially a theoretical paper suggesting a test. We then illustrated that test in terms of an empirical poverty equation model.

(E) In the paper with D. Robinson (1997) we estimated a production function using spatial modeling techniques. We also considered other techniques, such as systems estimation, autocorrelation, and heteroskedasticity

(F) The paper with D. Robinson (1995) was a theoretical paper.

(G) The paper with D. Robinson (1993) was a theoretical paper the results of which were illustrated in terms of a county police expenditure spatial econometric model.

(H) The paper with D. Robinson (1992) was a theoretical paper the results of which were illustrated in terms of a country expenditure model involving spatial correlation.

(I) I also wrote with D. Robinson, Local Public Finance Impact Model: User's Guide and Technical Documentation, IWR Report 94-FIS-10. In this report we used spatial econometric techniques, among other things.

(J) I also did a number of studies for the Army Corps of Engineers relating to vessels, lockage stalls and stall duration, vessel fuel tax system and lock chamber performance and suggestions for maintenance.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-2. If any of the following cannot be confirmed, please explain fully. On page 2 of the document originally filed as Appendix C to GCA-T-1, witness Thress's model is described as follows:

$$Ln(V) = a + b(X^{\lambda}) + ...$$

The "Box-Cox Model" is described as follows:

$$Ln(V) = a' + b'(\frac{X^{\lambda} - 1}{\lambda}) + ...$$

a. Please confirm that the "Box-Cox Model" above could be re-written as follows:

$$Ln(V) = a' + \frac{b'}{\lambda}X^{\lambda} - \frac{b'}{\lambda} + ...$$

b. Please confirm that the equation in a. could be re-written as follows:

$$\begin{aligned} Ln(V) &= a + bX^{\lambda} + ... \\ a &= a' - \frac{b'}{\lambda}; \text{ and } b = \frac{b'}{\lambda} \end{aligned}$$

c. Please confirm that rewriting the equations in this manner shows that the statement on page 2 that "the transformation that Thress used in formulating the internet variable was not the Box-Cox transform" is not correct.

RESPONSE: (a) The Box-Cox model

$$Ln(V) = a' + b'(\frac{X^{\lambda} - 1}{\lambda}) + ... \quad (1)$$

can be re-written as you suggest as

$$Ln(V) = a' + \frac{b'}{\lambda}X^{\lambda} - \frac{b'}{\lambda} + ... \quad (2)$$

(b) The model in (2) can be re-written as

$$\begin{aligned} Ln(V) &= a + bX^{\lambda} + ... \quad (3) \\ a &= a' - \frac{b'}{\lambda}; \text{ and } b = \frac{b'}{\lambda} \end{aligned}$$

(c) Given Thress's description of his Box-Cox model on page 37, I assumed that the restrictions in the second line of (3) were not maintained and the parameters a and b were taken without the restrictions

$$a = a' - \frac{b'}{\lambda}; \text{ and } b = \frac{b'}{\lambda}$$

If these restrictions were, in fact, maintained throughout the analysis, which includes restricting the estimated parameters, then the models in (1) and (3) can be viewed as equivalent. Otherwise they are not the same.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-3. Suppose that one believed that the true model for the demand for First-Class single-piece letters was equation (2) hypothesized by Dr. Clifton at line 3 of page 18 of his testimony (GCA-T-1) in this case:

$$\log(Q) = a - b\log(P) + b_2\log(P_2)$$

Suppose further that the true value of log(P2) was not known.

a. Would it be appropriate in this case to attempt to find some variable, call it z, to serve as a proxy for $\log(P_2)$ within equation (2)? If not, why not?

b. Suppose that there was some variable, X, and some constant, y, such that X^y appeared to be very highly correlated with $\log(P_2)$. Would it be appropriate in this case to substitute X^y into equation (2) as a proxy for $\log(P_2)$? If not, why not?

c. If X^y as described in part b. were used instead of $\log(P_2)$ in equation (2), would the estimated value of b be biased? If so, please provide the precise mathematical formulation for the expected value of b expressed as a function of the true value of b?

d. If X (not raised to the power y) as described in part b. were used instead of $\log(P_2)$ in equation (2), would the estimated value of b be biased? If so, please provide the precise mathematical formulation for the expected value of b expressed as a function of the true value of b?

RESPONSE:

(a): If the model is

$$\log(Q) = a - b\log(P) + b_2\log(P_2) \tag{4}$$

and the value of $\log(P_2)$ is not known and the model is estimated in terms of a proxy variable, z, then the results will depend upon the relationship of that proxy and $\log(P_2)$. There are many possibilities of such a relationship. As one example, if

$$\log(P_2) = \gamma_0 + \gamma_1 z + u \tag{5}$$

where u is an error term which is independent of z, then the model in (4) expressed in terms of z will be

$$\log(Q) = a - b\log(P) + b_2\gamma_0 + b_2\gamma_1 z + b_2 u \tag{6}$$

This model would probably be re-written and estimated as

$$\begin{aligned} Log(Q) &= A_0 - b\log(P) + A_1 z + e \\ A_0 &= a + b_2\gamma_0; A_1 = b_2\gamma_1 \end{aligned} \tag{7}$$

where A_0 and A_1 are parameters to be estimated, and $e = b_2 u$ is an error term. Under typical assumptions, the parameters of (7) which can be estimated are the ones that appear in the equation, namely $A_0, b,$ and A_1. Note that

$$\begin{aligned} A_1 &= b_2\gamma_1 \\ &\neq b_2, \text{ unless } \gamma_1 = 1 \end{aligned}$$

Therefore, unless $\gamma_1 = 1$ there is a problem concerning the estimation of b_2. The estimation of b_2 requires information about γ_1.

On the other hand, suppose if instead of (5), that the relationship between $\log(P_2)$ and z were

$$z = \delta_0 + \delta_1 \log(P_2) + u \tag{8}$$

where u is an error term which is now assumed to be independent of $\log(P_2)$. For future reference note, however, that the error term u would not be independent of z since via (8), z directly depends on u. In this case (8) implies

$$\log(P_2) = \frac{1}{\delta_1}[z - \delta_0 - u]$$

and so the model in (4) would become in terms of z

$$\begin{aligned} Log(Q) &= A_0 - b\log(P) + A_1 z + e \qquad (9) \\ &= A_0 + b[-\log(P)] + A_1 z + e \\ A_0 &= a - b_2\delta_0/\delta_1 \text{ and } A_1 = b_2/\delta_1 \end{aligned}$$

where $e = -\frac{b_2}{\delta_1}u$. Now there are two problems in the estimation of (9). One is that $A_1 \neq b_2$ unless $\delta_1 = 1.0$. The other is that the variable z is correlated with the error term, u and therefore with e. If the focus of attention is on the parameter b, on an intuitive level, the unbiased (more formally consistent) estimation of b has to account for the correlation between z and e. One way to do this is by the instrumental variable technique, an important special case of which is two stage least squares.

(b): Please see the answer to Part (a) above. In other words, X^y could be viewed as z.

The only additional concern I would have here relates to how the constant y is determined. If it is an estimate based on data, then additional problems arise because y would be random as described by its estimated variance.

(c): Please see Parts (a) and (b) above. In the first case described above which lead to equation 7 there would be no bias concerning the estimation of b. Now consider the case as described by (8), which suggests a bias if the correlation between z and u is not accounted for.

Let X be the regressor matrix corresponding to equation (9). Let

$$q' = [0, 0, -\frac{b_2}{\delta_1} cov(z, u)]$$

where cov is the covariance between z and u. Let the sample size be T and let Σ be the limit of $T^{-1}X'X$. Let $M_{2.}$ be the second row of Σ^{-1}. Then, on an intuitive level, the bias, say Bias, in the least squares estimator of b (more formally the inconsistency) would be, under standard assumptions

$$\begin{aligned} Bias &= M_{2.}q \\ &= M_{2.}[0, 0, -\frac{b_2}{\delta_1} cov(z, u)]' \end{aligned}$$

(d): Please see Parts (a), (b), and (c).

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-4. On page 4 of the document originally filed as Appendix C to GCA-T-1, Prof. Kelejian challenges witness Thress's assertion that "Holding all other factors constant, the total volume leaving First-Class single-piece mail due solely to changes in worksharing discounts should be exactly equal to the volume entering First-Class workshared mail." Specifically, he argues that "the statement relating to these equal but opposite volume flows between First-Class single-piece mail and First-Class workshare mail rests on an assumption that there are no spill-over effects with respect to any other forms of mail!"

a. If "spill-over effects with respect to any other forms of mail" were considered to be one of the "other factors" which witness Thress assumes to be held constant, would witness Thress's assertion be true? If not, why not?

b. Is it believed that witness Thress has failed to properly account for "spill-over effects" between First-Class and Standard Mail?

c. If the answer to b. is affirmative, please explain the answer, considering that witness Thress explicitly includes the average price difference between First-Class workshared letters and Standard Regular mail in his demand equation specification for First-Class workshared letters.

d. If the answer to b. is negative, is there any practical value to the assertion that "the statement relating to these equal but opposite volume flows between First-Class single-piece mail and First-Class workshare mail rests on an assumption that there are no spill-over effects with respect to any other forms of mail!"? If so, please explain.

RESPONSE: (a): If one assumes that there are no spill-over effects of any kind, which should include stoppages of mail that are not replaced by other forms of mail, and no new forms of mail that might come into existence, then the equal and opposite assumption might be true. In other words, one is essentially assuming that nothing can change except these two forms of mail and so what goes out "one door" must go in the "other door". I think this is a strong assumption - i.e., it is very restrictive.

(b): My statement was more hypothetical: if there are spill-over effects of any kind, some of which are described in Part (a), then the equal but opposite flow assumption would be violated.

(c): Please see my answers to Parts (a) and (b), especially Part (a). Also, I think that the average price difference between first class workshared letters and standard regular mail does not account for all possibilities concerning spill-overs that might arise. Finally, it should be noted that the nature of my comments on this issue were based on Thress's presentation on page 54 and his equations II.4, II.3, and II.5. Finally, in my opinion, this is a lesser point given the problems I identified with the econometrics as to how the symmetry condition was imposed.

(d): Please see parts (a) -(c).I still think this is a strong assumption!

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-5. a. On page 7 of the document originally filed as Appendix C to GCA-T-1, Prof. Kelejian suggests that witness Thress should have replaced the term $[Ln(d_w s)/(V_w s/V_s p)]$ (called "z" by you) with a fitted value $\hat{z}$. He goes on to assert that "[i]t should be clear that $\hat{z} \neq [Ln(d_w s)/(V_w' s/V_s' p)]$. Please explain why $[Ln(d_w s)/(V_w' s/V_s' p)]$, as estimated by witness Thress would not represent a satisfactory instrumental variable.

b. Again, on page 7 of the document originally filed as Appendix C to GCA-T-1, Prof. Kelejian states that "[i]f the variable $z = [Ln(d_w s)/(V_w s/V_s p)]$ ... is replaced by a variable such as $[Ln(d_w s)/\hat{w}]$, the resulting parameter estimates will not be consistent, i.e., on an intuitive level, there would be biases."

Please explain precisely why this substitution of $\hat{w}$ within witness Thress's demand equation will produce biased elasticity estimates. Please provide the precise mathematical formulation for the expected values of witness Thress's elasticity estimates relative to their actual values, given witness Thress's use of $[Ln(d_w s)/\hat{w}]$ as an instrumental variable for $[Ln(d_w s)/(V_w s/V_s p)]$ in this case.

c. What is the precise specification which is recommend be used by witness Thress to estimate the elasticity with respect to the average worksharing discount in his First-Class workshared letters equation?

RESPONSE: (a): First, please see the discription of the two stage least squares procedure as given in the literature.[1] On an intuitive level, let $w_t = (V'_{ws}/V'_{sp})$ and let $\hat{w}_t$ be the predicted value of w_t from the regression on the instruments. Also, let

$$\Delta_{t1} = Ln(dws)/(V'_{ws}/V'_{sp}) - Ln(dws)/\hat{w}_t \tag{10}$$

Finally, let $\hat{z}_t$ be the predicted value of $Ln(dws)/(V'_{ws}/V'_{sp})$ from a regression on the instruments, and let

$$\Delta_{t2} = Ln(dws)/(V'_{ws}/V'_{sp}) - \hat{z}_t \tag{11}$$

Then, if the instruments that were used include all the predetermined variables in the equation being considered, as well as the ones indicated by Thress, Δ_{t2} will be orthogonal to all of the predetermined variables appearing in the model because they were all used as part of the instruments. Please see the above indicated references; this orthogonality is a condition for consistency. This would correspond to the use of $\hat{z}_t$. On the other hand, because of the nonlinearity involved in (10), Δ_{t1} will not be orthogonal to all of

[1] A good reference for models that contain non-linearities is T. Amemeiya, Avanced Econometrics, Harvard University Press, Cambridge, Mass. 1985. See especially chapter 8. A simpler presentation of some of this material is given in H. Kelejian and W. Oates, Introduction to Econometrics, Principles and Applications, (3rd edition), Harper and Row Publishers, 1989. Concerning this, see chapters 7, which relates to linear models, and 8 which relates to nonlinear models. Concerning the selection of instruments please see pp. 277-280.

the predetermined variables appearing in the model even if they were all used which I do not think they were because Thress describes his procedure on page 55 in terms of three instruments, namely a dummy variable, a time trend, and its square. Actually, another issue is that Thress indicates that he replaced the ratio which appears in the model, namely V_{ws}/V_{sp} by the calculated value of $Ln(V'_{ws}/V'_{sp})$. I really do not know why he did this because the ratio that appears in the model is not in log terms.

(b): Please see Part (a). On an intuitive level, there would be biases. More formally, let X be the regressor matrix corresponding to the use of $Ln(d_{ws}/\hat{w})$. Let β be the vector of parameters appearing in the model, and let the estimator of β be $\hat{\beta}$. Again, let the sample size be T, and let Σ be the (probability) limit of $T^{-1}X'X$. Finally, let Δ_1 be the $T \times 1$ vector of values of Δ_{t1} corresponding to the use of $Ln(d_{ws}/\hat{w})$. Then, under typical assumptions, the bias (more formally the inconsistency) of each of the elements of the vector $\hat{\beta}$ is

$$\Sigma^{-1}[p\lim T^{-1}X'\Delta_1] \tag{12}$$

The expression in (12) would be zero if $\hat{z}_t$ were used and were based on the three instruments Thress used, as well as all predetermined variables appearing in the model. The reason for this is that $X'\Delta_2 = 0$ because of their orthogonality.

(c): I can not answer this question because I have not studied this market, nor do I know all the details which I would need to know in order to give a proper answer.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-6. At the bottom of page 7 of the document originally filed as Appendix C to GCA-T-1, Prof. Kelejian states that witness Thress ”implicitly assumed that the error terms in the First Class single piece and worksharing equations are uncorrelated.” At the top of page 8 of that document, he suggests that this assumption ”may not be reasonable.”

a. Is there any specific basis for questioning the reasonableness of this assumption? Please provide all mathematical or statistical evidence that was used to form this basis.

b. Please confirm that the correlation between the regression residuals for First-Class single-piece letters and First-Class workshared letters calculated by witness Thress in this case, which are presented in Library Reference LR-L-64, sponsored by witness Thress, is equal to -0.056 over the sample period for which both sets of residuals exist.

c. In general, would two variables which exhibit a correlation of -0.056 be considered to be ”uncorrelated”? If not, why not.

RESPONSE: (a): The assumption I am referring to is the lack of correlation between the two equations. If the two equations have correlated disturbance terms, then the prior information used to estimate the first class single piece mail will not be independent of the disturbance term in that equation. In other words, I think one should consider the possibility that there is a correlation. If there is one, then the development in the referred-to book by Judge et al. on pages 57-59 will indicate that there should be a covariance term in the matrix equation (3.2.20b) in that book by Judge et al. on page 58. Actually, on page 58 Judge et al. explicitly assume that the prior information is independent of the model’s disturbance term.

(b): I was not aware of this estimated correlation. However, if that estimated correlation is in terms of estimated residuals which are obtained from procedures which are subject to question, then the estimated correlation is also subject to question. For example, if the estimated residuals from the single piece equation are obtained after the symmetry conditions are imposed as was indicated by Thress, then I would interpret that estimated correlation with suspicion and likely subject to bias.

(c): In practice, given typical sample sizes, etc. I would assume that two variables are uncorrelated if their “properly estimated” correlation is -0.056. However, please see Part (b).

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-7. On page 9 in the document originally filed as Appendix C to GCA-T-1, Prof. Kelejian states that "model selection via a minimization of a mean squared error could very well lead to an incorrect model."

a. He cites, as one example of how a mean-squared error criterion could "lead to incorrect results", "the case in which the various models considered have different numbers of parameters." Please explain why a mean-squared error criterion could lead to incorrect results when comparing two equations with different numbers of parameters.

b. He cites, as another example of how a mean-squared error criterion could "lead to incorrect results", "the case in which a variety of complicated estimation procedures are considered." (i) What specific "complicated estimation procedures" are you referring to with respect to witness Thress's First-Class single-piece letters equation? (ii) Please explain why a mean-squared error criterion could lead to incorrect results when comparing two equations which are estimated using "complicated estimation procedures."

c. In discussing alternatives to a mean-squared error selection criterion, William Greene, in the third edition of his Econometric Analysis (1997), says the following on 401: "Although intuitively appealing, these measures are a bit unorthodox in that they have no firm basis in theory."

In the document originally filed as Appendix C to GCA-T-1, at page 1, Prof. Kelejian criticizes witness Thress because "some rather intuitive procedures were used that have no formal basis. This lack of a formal basis is important and not just a concern raised by an 'ivory-tower' academic."

Please confirm, therefore, that the "lack of a formal basis" for these alternative model selection criteria is a serious problem, and that Prof. Kelejian's statement argues strongly in support of a mean-squared error selection criterion, since the mean-squared error selection criterion has a "firm basis in theory". If not confirmed, please explain.

RESPONSE:

(a): Assume two models are to be compared and they have a different number of parameters. In my letter I had assumed that the mean square error relates to the model's dependent variable. That is, it is a residual sum of squares divided by something. My discussion below is based on this assumption.

As an illustration of "problems", in general if the model is estimated by least squares, the more parameters there are the smaller will be the residual sum of squares. In and of itself, this implies that one can not compare two models based only upon the residual sum of squares because the model with the most parameters will always "win". It would follow that if models are to be compared in terms of the residual sum of squares, there must be some penalty factor that will discount that sum of squares for the number of parameters which are used.[2] The mean square error can be calculated in a variety of ways. If it is

[2] As an illustration, see pages 159-160 in W. Greene, Econometric Analysis, 5th ed., Prentice Hall, Upper Saddle River, New Jersey, 2003.

just the sum of squared residuals divided by the sample size then it would not be useful because, again, the model with the most parameters would "win". A better calculation of the mean square error would be the residual sum of squares divided by the sample size minus the number of parameters contained in the model. In this case a given value of the residual sum of squares will lead to a larger value of the mean square error because the denominator would be smaller if a larger number of parameters are used.

I do not know which form of the mean square error Thress used. However, in a model such as his subject to the problems associated with his implementation of the Box-Cox formulation, his imposition of the symmetry condition, and his procedure for accounting for autocorrelation, I would have no idea as to what an appropriate penalty factor would be that would lead to a correct model selection.

(b): Please see Part (a). I am referring to the imposition of the Box-Cox formulation, the symmetry conditions, and the procedure for accounting for autocorrelation.

As a simple example that the mean square error might lead to an incorrect model consider a fairly minor complication of heteroskedasticity. This is not one of the complications I raised, nor is it one of the problems that Thress faced; I am using it for illustrative purposes only because of its simplicity.

Consider the two models

$$\begin{aligned} Model\ 1 &: y = X\beta + u; u \sim (0, \sigma^2 I) \\ Model\ 2 &: y = X\beta + u; u \sim (0, \Omega) \end{aligned} \tag{13}$$

Suppose Model 2 is the true model in the sense that it describes how the dependent variable y is generated. For ease of exposition, suppose Ω is known. In this case the researcher would estimate Model 1 by least squares and obtain the residual sum of squares, say RSS_{OLS}. The researcher would estimate Model 2 by generalized least squares and obtain the residual sum of squares, say RSS_{GLS}. By construction, the least squares procedure minimizes the residual sum of squares and so $RSS_{OLS} < RSS_{GSL}$. Since both models have the same number of parameters, and since the least squares procedure, by construction, minimizes the residual sum of squares, Model 1 would be selected even though Model 2 is the true model.

(c): Please see Part (b). Furthermore, in looking through the specification analysis and model selection Chapter 8 in the newer 5th edition by W. Greene, Econometric Analysis, 5th ed., Prentice Hall, Upper Saddle River, New Jersey, 2003 what I see is a reasoned evaluation of many procedures. I do not see a blanket statement that the various measures discussed in that chapter are unorthodox and have no firm basis in theory. I also do not see a statement which indicates that the mean squared error has a firm foundation in theory and can be applied to models which have various complications, let alone models like Thress's which contain improper procedures.

I can think of 3 procedures which I would view as formal.

1. The Bayesian posterior odds approach: in this framework one starts off with a prior probability that each considered model is the true one. One then considers the data and after going through certain manipulations, arrives at the posterior odds that each model is the true one. One would then select a particular model as the true one as the one which has the largest posterior odds.

2. A test involving the nesting of 2 models: In this framework one would consider two models. The one which the researcher believes to be most likely the true model is viewed as the null hypothesis. The other model is viewed as the alternative. Say the variables which appear in the alternative model which do not appear in the null hypothesis model are Z. Then one adds Z to the null hypothesis model and estimates that expanded model. One then tests for the joint significance of Z. The null model is accepted as the true model if Z is not jointly significant. If Z is jointly significant, I would stop my testing procedure and go back to the drawing board to reconsider my theory. I would view this as having a formal basis in the sense that it is based on a formal testing procedure.

3. The J test: This procedure is somewhat related to the one in 2 above. Here one tests to see if the alternative model has any explanatory power, or more formally, any significance after the null model is considered. I would view this as having a formal basis in the sense that it is based on a formal testing procedure.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-8. On page 9 of the document originally filed as Appendix C to GCA-T-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.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-9. Based on the selection criteria of his choosing, which of the First-Class single-piece letters models presented by witness Thress in LR-L-65 would Prof. Kelejian choose? If the choice is different from the model used by witness Thress in this case, please explain the basis for the choice and describe the ways in which the chosen model is superior to the model used by witness Thress.

RESPONSE:

As I indicated, I focused on the econometric procedures rather than on the actual empirical estimations, and so I can not pick out one of the equations as the “best one”. What I can say, however, is that I would have approached the model formulation and estimation differently. For example, I would have tried hard to come to a model which, in some sense, was based on whatever economic theory I had access to. I would have viewed that model as my null hypothesis. If I felt that there are parameter restrictions across equations I would have tried to estimate those equations jointly as a system. If I believed in the symmetry restrictions assumed by Thress I would not have gone through the substitutions he did. I would also have tried to estimate all parameters of the model jointly.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-10. Dr. Harry Kelejian, in the document originally filed as Appendix C to GCA-T-1, says the following with respect to witness Thress's estimate of the Box-Cox coefficient, λ.

> "The implication of this statement is that λ was first estimated in a preliminary step which was which was prior to full model estimation which, I assume, would incorporate his stochastic symmetry conditions, etc. Now this may seem to be a very intuitive thing to do, but on a formal level problems are raised. For example, suppose the estimated value of λ is λ'. This statement then suggests that the internet variable that was used in the full estimation of the model was $X^{\lambda'}$. If this is true, problems arise! Actually, one's intuition may lead one to think that problems should not arise if λ is properly estimated in that preliminary step. Unfortunately this is not the case. That is, even if λ is properly estimated in a preliminary step, the explanatory variable $X^{\lambda'}$ is not an ordinary explanatory variable because it is based on an estimated coefficient and therefore has a random component. This random component should be obvious since Thress himself on page 37 gives t-ratios relating to it! If an explanatory variable has such a random component that randomness can not be ignored in the model's estimation, nor can it be ignored in the inferences that come from that model! Assuming there are no other problems with the model, all of this suggests that the estimation of λ must be done in the final model considered which should incorporate all the other parameter restrictions that are considered. On a somewhat intuitive level, problems arise because the randomness in such a model would not only come from the model's error terms, but also from the explanatory variable, $X^{\lambda'}$."
>
> What are the specific problems to which Dr. Kelejian refers here? Will the inclusion of $X^{\lambda'}$ directly in the model estimated by witness Thress bias the estimated values of the other coefficients in witness Thress's model (e.g., the estimated own-price elasticity)? Please provide citations to econometrics literature in support of the answer.

RESPONSE:

Typically, econometricians study the consequences of a particular problem in the absence of other problems. This sort of analysis indicates marginal effects of the particular problem being considered. When I wrote my letter, which has since become direct testimony, I thought that there were two important marginal effects relating to the use of a variable such as $X^{\hat{\lambda}}$ as an ordinary regressor. One of these, I thought, was an inconsistency (intuitively, a bias) relating to the estimates of the model parameters. Upon further thought this is not the case. As my demonstration below indicates, under typical conditions an inconsistency would not be one of the marginal effects. The other marginal effect, however, is that the estimated standard errors relating to the model parameter estimates would be incorrect unless the random nature of $X^{\hat{\lambda}}$ was explicitly considered. As my demonstration below will indicate this is a marginal effect. The importance of this, for example, is that one may think that a variable is significant, when it is not. My reasoning is described in detail below. On an intuitive level, and excluding systems issues, typically calculated standard errors (or t-ratios) are based on the assumption that the only random component in the model is the error term. However, if an explanatory variable is of the form $X^{\hat{\lambda}}$, then there are two random components to the model and so standard errors which do not account for both will be incorrect.

Please see pages 127-135 in T. Amemeiya, Advanced Econometrics, Harvard University Press, Cambridge, Massachusetts, 1985. These pages describe nonlinear least squares estimation of the model parameters and its properties. The presentation is such that there may be nonlinearities with respect to all the parameters, or just one. See also pages 368-370 in that book. These pages give results relating to a model in which the regressor is based on an estimated parameter. The results given clearly demonstrate that the variances of the estimators of the model parameters reflect the random nature of such a regressor. I am giving my direct demonstration below because I think it is much simpler. Another reference relating to nonlinear estimation is chapter 9 in W. Greene, Econometric Analysis, 5th ed., Prentice Hall, Upper Saddle River, New Jersey, 2003.

My Direct Demonstration

Suppose the model is

$$y_t = Z_t\gamma + bX_t^{\lambda} + \varepsilon_t \tag{14}$$

where γ, b, and λ are parameters, and Z_t is a vector of observations on explanatory variables, X_t is an explanatory variable, and ε_t is the disturbance term. Let T be the sample size. Suppose we have a consistent estimator of λ, say $\hat{\lambda}$. This means that as the sample size increases beyond limit (goes to infinity), the probability that $\hat{\lambda}$ and λ will differ by any amount, however small, goes to zero. To continue, suppose $\hat{\lambda}$ is based on data which are independent of the disturbance term in (14). [3]

Consider the variable $X_t^{\hat{\lambda}}$. Using the mean value theorem to expand about λ, we can express $X_t^{\hat{\lambda}}$ as

$$X_t^{\hat{\lambda}} = X_t^{\lambda} + X_t^{\tilde{\lambda}} \log(X_t)[\hat{\lambda} - \lambda] \tag{15}$$

where $\tilde{\lambda}$ is between $\hat{\lambda}$ and λ. Intuitively, since $\hat{\lambda}$ goes to λ, as the sample size goes to infinity, and $\tilde{\lambda}$ is between $\hat{\lambda}$ and λ, then $\tilde{\lambda}$ also goes to λ. Somewhat more formally, as the

[3] This independence assumption is a simplification. In the model considered by Thress this independence condition would not hold because $\hat{\lambda}$ would be based on the same data.

sample size increases beyond limit, the probability that $\tilde{\lambda}$ will differ from λ by any amount limits to zero.

Substituting (15) into (14) yields a model in which the variable being explained, y_t, is related to $X_t^{\hat{\lambda}}$, namely

$$y_t = Z_t\gamma + bX_t^{\hat{\lambda}} + bX_t^{\tilde{\lambda}} \log(X_t)[\hat{\lambda} - \lambda] + \varepsilon_t \tag{16}$$

The explanatory variables in (16) are Z_t and $X_t^{\hat{\lambda}}$. Let the T observations on Z_t and $X_t^{\hat{\lambda}}$ be given by the matrix H, and let the corresponding coefficient vector be

$$\beta' = (\gamma', b)$$

Finally, let the T values of $bX_t^{\tilde{\lambda}} \log(X_t)$ be given by the vector V. Then, the matrix form of (16) is

$$y = H\beta + V\ [\hat{\lambda} - \lambda] + \varepsilon \tag{17}$$

where ε is the vector of disturbance terms.

The least squares estimator of β based on (17), which takes $X_t^{\hat{\lambda}}$ as one of the explanatory variables is

$$\hat{\beta} = (H'H)^{-1}H'y \tag{18}$$

Replacing y by its expression in (17) yields

$$\begin{aligned} \hat{\beta} \ &= \ \beta + (H'H)^{-1}H'V\ [\hat{\lambda} - \lambda] + \\ &\quad\ (H'H)^{-1}H'\varepsilon \end{aligned} \tag{19}$$

which can easily be expressed as

$$T^{1/2}[\hat{\beta} - \beta] = [T(H'H)^{-1}]\ [T^{-1}H'V]\ T^{1/2}[\hat{\lambda} - \lambda] + [T(H'H)^{-1}]\ T^{-1/2}H'\varepsilon \tag{20}$$

As the sample size T increases beyond limit, typical modeling assumptions would imply limits of the sort:[4]

$$\begin{aligned}
[T(H'H)^{-1}] &\rightarrow \Sigma_{HH}^{-1} \\
[T^{-1}H'V] &\rightarrow \Sigma_{HV} \\
T^{1/2}[\hat{\lambda}-\lambda] &\rightarrow N(0,\alpha^2) \\
T^{-1/2}H'\varepsilon &\rightarrow N(0,\sigma_\varepsilon^2\Sigma_{HH})
\end{aligned} \tag{21}$$

where

$$\begin{aligned}
&\Sigma_{HH}^{-1} \text{ is a finite matrix} \\
&\Sigma_{HV} \text{ is a finite vector}
\end{aligned}$$

and where $N(0,\alpha^2)$ is the normal with mean zero, and variance α^2, and $N(0,\sigma_\varepsilon^2\Sigma_{HH})$ is the multivariate normal with mean zero and variance-covariance matrix $\sigma_\varepsilon^2\Sigma_{HH}$).

The implication of (20) and (21) is that the limiting distribution of $T^{1/2}[\hat{\beta}-\beta]$ is

$$T^{1/2}[\hat{\beta}-\beta] \rightarrow N(0,\Omega_1+\Omega_2) \tag{22}$$

where

$$\begin{aligned}
\Omega_1 &= \alpha^2\Sigma_{HH}^{-1}\Sigma_{HV}\Sigma_{HV}'\Sigma_{HH}^{-1} \\
\Omega_2 &= \sigma_\varepsilon^2\Sigma_{HH}^{-1}
\end{aligned}$$

In light of (22) one would base small (e.g., finite) sample inferences on the approximation

$$\hat{\beta} \sim N(\beta, T^{-1}[\Omega_1+\Omega_2]) \tag{23}$$

Notice that if $\hat{\lambda}=\lambda$, then $\alpha^2=0$ and the normal distribution in (23) would reduce to

$$\hat{\beta} \sim N(\beta, T^{-1}\Omega_2) \tag{24}$$

If the researcher ignores the randomness in an explanatory variable such as$X_t^{\hat{\lambda}}$, then inferences that would be made using standard computer results would correspond to an empirical version of (24).

Two points should be noted about these results. First, since the distribution in (22) has a zero mean vector, the estimator $\hat{\beta}$ of β would not be inconsistent. Again, on an

[4]In obtaining (21) one should note Theorem 4.1.4 on page 112 in the book by T. Amemiya, Advanced Econometric, Harvard University Press, Cambridge Massachusetts, 1985.

intuitive level, this means there would not be a bias. On the other hand, since the variance covariance matrix in (23) is larger than the one in (24) (larger means that the difference is positive semidefinite), the t-ratios based on an empirical version of (24) will be overstated and so one might be lead to think that a particular variable is significant when in fact it is not.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-11. At page 9 in the document originally filed as Appendix C to GCA-T-1, Prof. Kelejian indicates an understanding that "Thress indicates that he tested for autocorrelation via the model which is his equation III.12."

a. In his testimony, on page 321, at lines 1 - 4, witness Thress describes the procedure whereby he tests for the presence of autocorrelation as follows:

The exact nature of the autoregressive process is identified by testing the significance of the partial autocorrelation of the residuals at one, two, and four lags. In general, a 95 percent confidence level is used to test for the presence of autocorrelation.

Is the methodology described by witness Thress in the above quote, testing the significance of the partial autocorrelation of the residuals, an appropriate method of testing for the presence of autocorrelation? If not, why not?

b. At page 9 in the document originally filed as Appendix C to GCA-T-1, Prof. Kelejian states that "since the parameter λ in Thress's version of the Box-Cox procedure was estimated prior to the full estimation of his model, and given the errors in the way he imposed the stochastic symmetry conditions, it is difficult to deduce just how to make proper inferences in terms of a model such as III.12."

If the Box-Cox coefficient, λ, and the stochastic symmetry condition were introduced as you have suggested they should have been by witness Thress, would there be any objection to the procedure which he used to test for and correct autocorrelation? Is so, please explain.

RESPONSE:

(a): First, I still do not see why the third quarter lag was not considered. It still seems confusing to me that one would consider one, two, and four quarter lags without also considering the third quarter lag. Ignoring this issue, I would find it difficult to determine the properties of the "seemingly reasonable" procedure used by Thress because of the other procedures he used, namely his version of the Box-Cox procedure, his imposition of the symmetry condition, and his model selection procedure. I have stated problems with respect to all of these procedures. The properties of the tests used by Thress are difficult to determine because, I assume, the residuals were obtained after these complex procedures were used. If they were obtained before these procedures were used, I would find that very difficult to understand because a researcher would typically look for autocorrelation as somewhat of a residual after his model contains all the available information.

(b): My objections would diminish, but not to zero. The reason for this is that I think the third quarter lag should also have been considered. Second, I might have accounted for an assumed pattern of autocorrelation by incorporating it directly into the model and estimating its parameters along with the other model parameters.

RESPONSES OF GREETING CARD ASSOCIATION WITNESS KELEJIAN (GCA-T5) TO INTERROGATORIES OF THE UNITED STATES POSTAL SERVICE

USPS/GCA-12. Please provide copies of a Statement of Work, or any other documents or correspondence containing instructions or directions defining the nature and scope of the task(s) that Prof. Kelejian was asked to perform in the preparation of the document which became the declaration originally attached to GCA-T-1 as Appendix C.

RESPONSE:

See attached documents.

From: James Clifton [clifton@wecg.com]

Sent: Wednesday, August 16, 2006 2:43 PM

To: 'Harry Kelejian'

Subject: Question

Harry:

Re: our discussion on Monday on just the critique of Thress, would you be willing to stand by those points by putting them in a typed letter under your signature for possible use as an Appendix?
Although I know from having used your intro McGraw Hill text that you write well or Oates does or McGraw Hill editors do, or all of the above, we would probably want to take a crack at editing a draft of it. By "we" I mean the attorneys mostly.

Ball-park cost is the second hurdle if you are amenable to this.

Regards,

Jim

P. S. Promise stronger, better coffee next time.

From: James Clifton [clifton@wecg.com]
Sent: Friday, August 18, 2006 11:05 AM
To: 'Harry Kelejian'
Subject: RE: Question

That is fine, Harry. Please just advise me if we are getting close to that $4,200 cap. If we are getting close you should know by that time more precisely whether: (a) you will be within that cap; and (b) if beyond the cap, how much more time you will need, etc.

Time is of the essence in getting a complete draft letter from you as the final must be submitted September 6th, but really needs to be complete for reproduction and related production reasons by Friday, September 1st. So, if we can get the draft by next week, that would be great.

Jim

-----Original Message-----
From: Harry Kelejian [mailto:kelejian@econ.bsos.umd.edu]
Sent: Friday, August 18, 2006 10:04 AM
To: clifton@wecg.com
Subject: RE: Question

James, concerning the ball park figure: I had a good talk with Reza yesterday. He said there were about 40 pages in the Thress testimony that I did not read but I should. I arrived at a ball park figure of $4200. This figure was constructed as follows. $175 per hour times:
Roughly

4 hours for carefully reading those 40 pages
4 hours for "getting into the writing mode
4 hours for each of the following 3 procedures : Autocorrelation, Box-Cox, and the symmetry conditions. The MSE procedure related to the Shiller procedure and so there is no reason to go over that. Also, a good discussion of that would take more time.
4 hours for iterations on comments I may receive from your lawyers.
The actual bill I send to you will relate to the hours I actually put in. That bill may be less than $4200 less.