Postal Rate Commission
Submitted 9/6/2006 12:13 pm
Filing ID: 53140
Accepted 9/6/2006

OCA-T-3
Docket No. R2006-1

DIRECT TESTIMONY

OF

J. EDWARD SMITH

ON BEHALF OF

THE OFFICE OF THE CONSUMER ADVOCATE

SEPTEMBER 6, 2006

TABLE OF CONTENTS Page

I. STATEMENT OF QUALIFICATIONS 1

II. PURPOSE OF TESTIMONY AND SUMMARY OF MAJOR CONCLUSIONS 2

III. COMMENTS ON WITNESS BRADLEY'S MODEL 3

A. The underlying database is deficient. 3

B. The density variable should be eliminated from the analysis. 3

C. The underlying City Carrier model appears to be of an *ad-hoc* equation estimation nature rather than being based on an economic analysis. 8

IV. COMMENTS ON THE EQUATIONS 9

A. Twenty four alternatives were considered 11

B. The recommended equation is CC5 14

V. COMMENTS ON DOIS DATA 16

A. Equations based on the approach used in witness Bradley's analyses are reported in Table 3. 17

B. Accordingly, based on the limited analysis of the database up to this point, I would advocate the use of the ND6 unrestricted volume variability. 20

VI. LESSONS LEARNED AND COMMENTS 21

A. The theoretical underpinnings of the carrier cost analysis continue to be weak. 21

B. One variable missing from the DOIS database is collection volume. 21

C. There has been extensive discussion of some of the problems associated with the database used by witness Bradley. 22

VII. CONCLUSIONS ON MODELING EFFORTS 23

DIRECT TESTIMONY
OF
J. EDWARD SMITH

I. STATEMENT OF QUALIFICATIONS

My name is J. Edward Smith, and I am an econometrician with the Office of the Consumer Advocate of the Postal Rate Commission. I have previously worked in a variety of economic assignments in industrial, academic, consulting, and governmental positions. My economics degrees are an A.B. from Hamilton College, and an M.S. and Ph.D. from Purdue University. I have testified before the Postal Rate Commission in Docket No. R97-1, Docket No. R2000-1, and Docket MC2002-2.

II. PURPOSE OF TESTIMONY AND SUMMARY OF MAJOR CONCLUSIONS

In his testimony on City Carrier Costs,[1] witness Bradley states that:

> There was simply insufficient time to refine the City Carrier costing study along the lines indicated by the Commission. Because of this time limitation, Postal Service decided that the best available city carrier street time costing methodology was the one chosen by the Commission in Docket No. R2005-1.

Accordingly, the conclusions from the Carrier Cost study presented by witness Bradley in Docket No. R2005-1 provide the theoretical and empirical basis for the Postal Service's City Carrier cost analysis in the current case. My testimony on the methodology of the original City Carrier cost model provides alternative volume variability estimates based on the database available in Docket No. R2005-1, provides a brief analysis of recently received DOIS data, and concludes that additional improvements in the estimation of City Carrier volume variability and data availability are needed.

[1] Testimony of Michael D. Bradley, Docket No. R2006-1, USPS-T-14, at 10.

III. COMMENTS ON WITNESS BRADLEY'S MODEL

A. The Underlying Database is Deficient.

The Commission has identified a number of flaws associated with the City Carrier Street Time Study (CCSTS) database underlying the City Carrier study.[2] In working with the data I have found regressors with signs that appear to be wrong on an *a priori* basis and variables with Variance Inflation Factors (VIF) values greater than 10. In addition, the relationship between the marginal costs of various products has frequently been different from that which would be expected on an *a priori* basis. The database appears to have significant multicollinearity problems, and it is not clear that meaningful conclusions can be obtained.

B. The Density Variable Should Be Eliminated from the Analysis.

Equations with the density variable should not be used in the estimation of volume variability. The City Carrier equations model delivery time as a function of the quantities of the various types of mail being delivered. Clearly delivery points should also be included, for carriers need to pass each delivery point in order to complete the route: one of the outputs of the delivery process is the passage by a carrier past a delivery point whether or not any mail is delivered.

Witness Bradley included density as a control variable. Measured as delivery points per square mile, one could propound density as accounting for differences in the

[2] Given that Appendix 1 of the Opinion and Recommended Decision, Docket No. R2005-1, delineated a number of deficiencies of the CCSTS database, I have focused my comments on methodology rather than data.

physical layout of ZIP codes—e.g., congestion, urban/suburban/rural, etc. However, inclusion of the density variable is theoretically incorrect.

In the modeling of an economic process one generally expects to see the maximization or minimization of a process subject to some type of constraint. This results in the development of various equilibrium conditions. Although no theoretical analysis of the underlying economic processes of mail delivery has been explicitly hypothesized in conjunction with the modeling effort, one could conclude that the equations model a cost function, with cost (measured in terms of time) as a function of output (pieces of mail delivered or collected plus coverage of the delivery points). Delivery points are also an output, given that the City Carrier must service delivery points by walking or driving by the point, regardless of whether any mail is delivered.

A cost curve for a firm models cost as a function of output. The costs are a result of production decisions and tradeoffs made by the firm's management in an attempt to minimize overall costs. The level of costs is a result of decisions made by the firm's managers in attaining a level of output for various levels of inputs. The function relating the level of output to the levels of inputs is denoted as a production function.

The managers of the firm attempt to minimize costs for a given level of output; the production function relates the choices of levels of various inputs to the level of output. For example, for a given level of output the firm can be thought of has having a budget, which can be spent on a variety of possible levels of inputs. A budget line delineates the various combinations of inputs that can be purchased for a given level of the budget; this budget line is denoted by economists as an isocost. Similarly, for a

given level of output the firm can make tradeoffs between the various inputs. A curve that delineates the various levels of inputs needed to produce a given output is termed an isoquant.

Isoquants and isocosts intersect. An intersection of an isoquant with the minimum possible isocost is an equilibrium: at this point the firm minimizes the cost of production. The tradeoffs between inputs take into account the tradeoffs between input costs and the efficiency with which the inputs can be used. In technical terms, the slope of the isoquant—denoting productivity tradeoffs between inputs, and the slope of the isocost—denoting cost tradeoffs, are equal. Accordingly, the set of intersections of isocosts with isoquants delineates the expansion path for the production process.

> The equilibrium of the firm in the long run, when both inputs can be freely varied, is at the tangency of an isocost to an isoquant. Only at such a point is output maximized for a given cost, or, equivalently, is cost minimized for a given output. The former follows by moving along any one isocost: if at any one point it crosses an isoquant it is possible to increase output with no additional cost—by moving toward the tangency point. Similarly, moving along any one isoquant, if at any one point it crosses an isocost, it is possible to decrease cost while holding output constant—by moving toward the tangency point. The locus of tangency points is the set of possible equilibrium points for the firm: it is called the expansion path and is characterized by the equality of slopes of isocost and isoquant. From the above results on these slopes, the geometric tangency is in fact equivalent to the algebraic condition (89.2.7), stating that, for profit maximization, the marginal rate of technical substitution must equal the ratio of wages.[3]

In the case of City Carrier delivery, the Postal Service management has a group of costs related to various inputs—for example, the costs of personnel, the costs of equipment, the costs of materials. Also, the Postal Service management has a group

[3] Michael D. Intriligator, Ronald G. Bodkin, Cheng Hsiao, *Econometric Models, Techniques, and Applications,* Second Edition, Prentice Hall, 1996, at 278

of tradeoffs between the ways in which inputs are used—i.e. decisions to use certain levels of personnel, overtime, equipment, materials in various possible route and delivery configurations. All of the costs and quantities are considered, and the management makes tradeoffs. An economist would refer to this effort as the attainment of equilibrium between isocosts and isoquants. The management probably thinks of the effort as making good decisions on how to get the job done as efficiently as possible—i.e., whether to work overtime, whether to pivot routes, whether to design routes in various configurations, etc. As a result of the consideration of tradeoffs, the cost to deliver a quantity of mail is determined. Density is an output of the process, not an input to the process. Density is determined partly by how the route is designed and partly by the characteristics of the service territory. What drives cost are the management's decisions on how to utilize resources to accommodate whatever level of mail and service territory characteristics are present.

Accordingly, it appears that witness Bradley's model is that of a cost curve with

$C = f(y_1, y_2, \ldots y_n)$.

Types of mail--e.g., letters, flats, sequenced mail, collection volume, and small packages—as well as delivery point are clearly outputs whose cost is being measured However, ZIP code density—i.e., dp/sqm—is a function of the arrangement of the City Carrier delivery routes, which would be achieved through the determination of a least cost solution to a production function through the attainment of equalities between various marginal rates of technical substitution and input/price ratios in a cost minimization process. The value of the density variable is an output of the cost minimization process; density is not an input to the cost function.

Therefore, it does not appear that the use of the density variable would comport with the estimation of a cost function where delivery time is a function of the various types of mail and delivery points. From a theoretical viewpoint the use of the density variable is wrong; it should be eliminated from the equations. [4]

Witness Bradley's computation of the density variable as reported for each ZIP code is also incorrect: the purported density variable does not measure density. In reporting mail deliveries, the number of routes reporting mail delivery for a specific ZIP code varied from day-to-day. In computing density for a ZIP code, one would use the land area in a ZIP code divided by the maximum number of delivery points in the ZIP code.

However, witness Bradley defines density differently. In aggregating to the ZIP code level he sums variables over all of the routes in the ZIP code. However, land is a constant for all routes within a ZIP code; instead, one should take the mean. Accordingly, his definition of density has a larger value than it should. Witness Bradley disagrees:

> Not confirmed. While it is certainly true that "land" is a constant within a ZIP CODE over a two week period, it is not true that the number of routes included in each ZIP CODE observation is constant. As the number of included routes varies, so does the number of delivery points. I did not have information about the square miles associated with the included and excluded routes. Thus, I used a crude method to attempt to account for the variation in number of routes included. If you look at the above code, you will see the phrase "n=nrts." This implies that the number of underlying observations included in each ZIP CODE-day is the number of routes. By summing the "land" variable over the underlying observations, I thus weighted the "land" variable by the number of routes included in that

[4] Equations with the density value are included in Table 1 for points of comparison. I do not advocate their use.

ZIP CODE—day observation. This provides a rough variation in the land variable as the number of reported routes varied.

Even though the number of routes included in a ZIP Code-day observation may vary, density should not vary: otherwise something other than density is being measured.

Witness Bradley's "weighted average" is strongly dependent on the number of routes reporting, and reported density becomes a function of the number of delivery points for which data were reported on a specific day. Density should measure the degree of proximity of delivery points, possibly providing information on congestion or carrier route miles to be driven. Such a variable does not vary by day. Accordingly, witness Bradley's definition of density appears to vary as a function of number of routes reporting as well as delivery points and square miles. It appears that the "density" variable does not measure density. One would expect density to be a function of square miles and maximum number of delivery points. This variable would take on various values depending on the degree of congestion in a ZIP Code; however, the variable should be based on the maximum number of delivery points, not the number of delivery points for which data were collected on a given day.

C. The Underlying City Carrier Model Appears to be of an *Ad-Hoc* Equation Estimation Nature Rather than Being Based on an Economic Analysis.

In reviewing an economic model one generally expects to see the maximization or minimization of something subject to constraints. This does not occur in the case of witness Bradley's model; rather, data are fitted to a quadratic function. Further specification or explanation of an economic model would be appropriate.

IV. COMMENTS ON THE EQUATIONS

Table 1 summarizes witness Bradley's recommended Carrier Cost model plus variations of the model. The relevant SAS programs are in OCA-LR-L-4. Each of the equations will be discussed.

Table 1: Initial Carrier Cost Model plus Variants

	Letters	Flats	Sequenced Mail	Collection Volume	Small Pkg	Deliv Pts	Density	Packages	Accountables	Deliv Pts
CC1: Witness Bradley Recommended Case										
CC1A: Full Quadratic										
Volume Variability	0.24069	0.11606	0.00700031	0.04707	-0.019978	0.67068	-0.0777	0.28451	0.24998	0.51162
Marginal Cost	1.53	2.26	0.46	2.22	-12.29			37.8	80.56	
CC1: Witness Bradley Recommended Case										
CC1B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.22268	0.0712	0.012912	0.088135	0.015836	0.66888	-0.0825			
Marginal Cost	1.39	1.36	0.82	4	9.56					
CC2: Recommended Case, density modified			**Density redefined as dens =**		**units/sqm;**					
CC2A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.25654	0.13852	0.015301	0.039501	-0.022269	0.68459	-0.0551			
Marginal Cost	1.62	2.68	0.99	1.81	-13.6					
CC2: Recommended Case, density modified			**Density redefined as dens =**		**units/sqm;**					
CC2B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.24101	0.06943	0.0067851	0.080097	0.0047302	0.70117	-0.0556			
Marginal Cost	1.52	1.34	0.43	3.68	2.89					
CC3: Witness Bradley Recommended Case without the Density Variable										
CC3A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.23729	0.09846	0.00961802	0.055261	-0.005995	0.67683				
Marginal Cost	1.53	1.94	0.63376	2.58541	-3.73382					
CC3: Witness Bradley Recommended Case without the Density Variable										
CC3B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.1905	0.06018	0.015632	0.097381	0.034415	0.65003				
Marginal Cost	1.21	1.17	1.02	4.49	21.12					
CC4: DPS Case: Letters contain only DPS mail. Flats include all other letter mail plus flats.										
CC4A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.1463	0.18436	0.010861	0.070447	-0.005367	0.67569	-0.0611			
Marginal Cost	1.38	1.74	0.7	3.2	-4.25					
CC4: DPS Case: Letters contain only DPS mail. Flats include all other letter mail plus flats.										
CC4B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.22404	0.15523	0.010843	0.083474	0.0006248	0.61091	-0.0763			
Marginal Cost	2.11	1.46	0.69	3.77	0.38					
CC5: DPS Case, No Density. Letters contain only DPS mail. Flats include all other letter mail plus flats.										
CC5A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.16827	0.16695	0.014871	0.065867	0.0052223	0.66961				
Marginal Cost	1.62	1.6	0.96645	3.03943	3.20827					
CC5: DPS Case, No Density. Lette Letters contain only DPS mail; flats include all other letter mail plus flats.										
CC5B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.23492	0.12524	0.013848	0.092313	0.013357	0.57798				
Marginal Cost	2.24	1.19	0.89491	4.23555	8.15939					
CC6: Witness Bradley Recommended Case--No Density and No Collection Volume.										
CC6A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.25106	0.09922	0.0081284		0.029457	0.65476				
Marginal Cost	1.64	1.98	0.54		18.6					
CC6: Witness Bradley Recommended Case--No Density and No Collection Volume.										
CC6B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**			
Volume Variability	0.20613	0.08324	0.015632		0.069213	0.64571				
Marginal Cost	1.31	1.61	1.02		42.5					

Table 1: Initial Carrier Cost Model plus Variants, Continued

CC7: DPS Case, No Dens No CV	Letters contain only DPS mail; flats include all other letter mail plus flats.						
CC7A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**
Volume Variability	0.17773	0.18642	0.013452		0.041029	0.64495	
Marginal Cost	1.72	1.8	0.88139		25.4107		

CC7: DPS Case, No Dens No CV	Letters contain only DPS mail; flats include all other letter mail plus flats.						
CC7B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Small**	**Deliv Pts**	**Density**
Volume Variability	0.25447	0.14466	0.013891		0.046062	0.57234	
Marginal Cost	2.43	1.37	0.9		28.1		

CC8: All Deliverables Combined:	Packages Include Small and Large Packages							
CC8A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.21501	0.11317	0.00128914	0.046896	0.032839	0.62188	-0.0638	0.009175486
Marginal Cost	1.5	2.5	0.09	2.3	15.3			36

CC8: All Deliverables Combined:	Packages Include Small and Large Packages							
CC8B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.23199	0.06046	0.0068126	0.081036	0.027866	0.65235	-0.0579	-0.005035105
Marginal Cost	1.62	1.3	0.48	3.9	12.7			-19

CC9: All Deliverables Combined	No Density							
CC9A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.23085	0.09846	0.00392398	0.052749	0.042791	0.63202		0.002243181
Marginal Cost	1.7	2.2	0.28	2.64	20			8.8

CC9: All Deliverables Combined	No Densi	No Density						
CC9B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.21545	0.05174	0.00830111	0.088895	0.046993	0.64664		-0.019314
Marginal Cost	1.52	1.12	0.59	4.4	21.6			-74

CC10: All Deliverables Combined, DPS			Letters contain only DPS mail; flats include all other letter mail plus flats.					
CC10A: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.11792	0.19564	0.00155636	0.064863	0.036315	0.64488	-0.0518	-0.006575585
Marginal Cost	1.24	2.1	0.11	3.2	16.8			-25.4

CC10: All Deliverables Combined, DPS			Letters contain only DPS mail; flats include all other letter mail plus flats.					
CC10B: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.21879	0.18268	0.00334832	0.073233	0.0095664	0.59846	-0.0569	-0.020921
Marginal Cost	2.28	1.94	0.24	3.56	4.35			-79.7

C11: Combined with DPS noDensity			Letters contain only DPS mail; flats include all other letter mail plus flats.					
C11: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.133	0.19662	0.00538559	0.059631	0.053977	0.6409		-0.006556601
Marginal Cost	1.41	2.13	0.38	2.96	25.08	-25.5		

C11: Combined with DPS noDensity			Letters contain only DPS mail; flats include all other letter mail plus flats.					
C11: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.22855	0.16386	0.00511854	0.080923	0.024411	0.58294		-0.036383
Marginal Cost	2.41	1.76	0.36	3.98	11.22	-140		

C12: Combined noDensity NoCV								
C12: Full Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.22562	0.10738	-0.00050955		0.063389	0.63693		0.000618476
Marginal Cost	1.62	2.37	-0.036		29.7			2.4

C12: Combined noDensity NoCV								
C12: Restricted Quadratic	**Letters**	**Flats**	**Sequenced**	**Collection**	**Packages**	**Deliv Pts**	**Density**	**Accountables**
Volume Variability	0.21922	0.07146	0.00786595		0.081127	0.63797		-0.000723733
Marginal Cost	1.6	1.6	0.56		37.4			-2.8

A. Twenty-four Alternatives Were Considered: A Full Quadratic and Restricted Quadratic for Each Equation.

CC1. Witness Bradley Recommended Case: The results reproduce witness Bradley's model presented in R2005-1. The full quadratic version had a sign problem in

the case of small packages; the restricted quadratic is witness Bradley's recommended model.

CC2. Recommended Case, Density Modified: The variable "density"" as used in witness Bradley's model has been modified to redefine density = units/sq.mile; the data were obtained from the density file in his study. Again there is a sign problem with small packages in the Full Quadratic in the computation of volume variability and the underlying regression equation. The Restricted Quadratic does not have a sign problem.

CC3. Witness Bradley Recommended Case without the Density Variable: The density variable was omitted; the full quadratic has a sign problem in the case of small packages. The restricted quadratic does not have a sign problem in the value of volume variability.

CC4. DPS Case: Letters contain only DPS mail. Flats include all other letter mail plus flats: Witness Bradley briefly discussed modeling Delivery Point Sequenced Mail (DPS) separately from other types of mail. However, such an approach was not his preferred case. DPS mail is received by the carrier presorted and goes to the street as a separate bundle. Other letters and flats received by the carrier are cased by the carrier and go to the street as a second bundle along with other mail. Based on my understanding of Postal Service delivery practices, the equation seems to model actual carrier activities more closely. In the equation, I substitute DPS mail for "let" and combine all other letter mail and flat mail under the variable "cf". In the case of the full quadratic, there is again a sign problem with small packages. The restricted quadratic case does not have a sign problem in the values for volume variability.

CC5. DPS Case, No Density, Letters contain only DPS mail. Flats include all other letter mail plus flats: The density variable is eliminated from the DPS analysis. All computed volume variabilities are positive in both the restricted and non restricted cases.

CC6. Witness Bradley Recommended Case—No Density and No Collection Volume: Collection volume was eliminated to see the effect on the other regressors. Any analysis of the DOIS data will, at least at this point, need to omit collection volume, for the variable is unavailable.

CC7. DPS Case, No Density, No CV: This is the DPS case with both density and collection volume omitted.

CC8. All Deliverables Combined in One Equation: Witness Bradley estimated volume variabilities for large packages and accountables separately from other types of mail in the City Carrier study. The delivery of accountables and packages generally requires that the carrier make a separate trip from the van to the delivery point with the parcel and/or accountable after delivering other mail. In assembling the database the Postal Service scanned delivery time from the carrier starting to deliver the parcel/accountable until the delivery had been completed. Accordingly, there was no scan of time for large parcels and accountables without there being a need to deliver a large parcel or an accountable. Witness Bradley modeled two equations. In this combined equation, all time and all delivered mail were estimated in a single equation. Packages were combined to include both large and small packages. Although the restricted quadratic has a sign problem, the full quadratic does not have a sign problem. The use of a density variable eliminates the usefulness of this equation.

CC9. All Deliverables Combined, No Density: This equation is similar to the above case but without the density variable. There are no sign problems for the full quadratic, but the marginal costs do not comport with *a priori* assumptions of reasonableness.

CC10. All Deliverables Combined in One Equation, with Letters denoted as DPS: There is a sign problem with accountables in both the full and restricted quadratic.

CC11. All Deliverables Combined in One Equation with DPS and no Density Variable: There is a sign problem in the estimation of volume variability in the case of accountables for both the full quadratic and restricted quadratic cases.

CC12. All Deliverables Combined in One Equation, No Density, No CV: This is not a DPS case; there is a sign problem with Sequenced Mail in the calculation of volume variability, and for the full Quadratic and with Accountables for the Restricted Quadratic.

B. The Recommended Equation is CC5.

In conclusion, a wide variety of possible equations were run. Table 2 presents the preferred equation. In general, the full quadratic cases generally had sign problems; this is probably a result of the collinearity of the underlying data, as also evidenced by the VIF values. The restricted quadratic equations without a sign problem and without a density variable but including collection volume were CC3 and CC5. The restricted version of CC5 appears to be slightly better than witness Bradley's case, based solely on the breakout between DPS and other letters. However, the full quadratic version of equation 5 is more general and is the equation recommended.

Table 2: Recommended Equation from Analysis of R2005-1 Data

Definiton of Variables	Variable	Regression CC5 Estimated Coefficient	Regression CC5 Consistent Covariance	HC Std Error	HC t statistic
Intercept	**Intercept**	-1757.2563	1183370.12	1087.828	-1.61538
DPS Mail	**let**	0.97149	0.21728279	0.466136	2.084134
DPS Mail Squared	**let2**	-1.352E-05	2.70E-11	5.19E-06	-2.60426
Flats, Other Letter	**cf**	0.93031	0.24449505	0.494464	1.88145
cf Squared	**cf2**	-2.366E-05	2.27E-07	0.000476	-0.04971
Sequenced Mail	**seq**	3.17986	0.91990358	0.959116	3.315407
Seq Mail Squared	**seq2**	0.00001823	2.98E-10	1.73E-05	1.055779
Collection Volume	**cv**	1.64581	0.49716195	0.705097	2.334161
Coll. Vol Squared	**cv2**	-1.143E-05	1.91E-10	1.38E-05	-0.826
Small Packages	**spr**	-22.05133	765.780293	27.67274	-0.79686
Sm Pkgs Squared	**spr2**	-0.05185	2.59E-04	0.016089	-3.2227
Delivery Points	**dp**	16.42127	3.14803609	1.774271	9.255223
Del. Pts Squared	**dp2**	-0.000529	1.58E-08	0.000126	-4.21099
let*cf	**lf**	-8.49E-06	2.95E-11	5.44E-06	-1.56197
let*seq	**lse**	0.00004532	8.98E-10	3E-05	1.512644
let*cv	**lcv**	-5.607E-05	1.21E-09	3.47E-05	-1.61518
let*spr	**lspr**	-5.46E-06	2.33E-07	0.000483	-0.01131
let*dp	**ldp**	0.00017071	1.35E-09	3.67E-05	4.651638
cf*seq	**fse**	-1.983E-05	1.18E-09	3.44E-05	-0.57682
cf*cv	**fcv**	0.00010929	7.92E-10	2.81E-05	3.88384
cf*spr	**fspr**	0.00069982	2.14E-07	0.000463	1.511988
cf*dp	**fdp**	0.00013418	1.70E-09	4.12E-05	3.256258
seq*cv	**scv**	-0.0003055	4.26E-09	6.52E-05	-4.68354
seq*spr	**sspr**	0.00019702	2.48E-06	0.001576	0.125028
seq*dp	**sdp**	-0.0001589	6.69E-09	8.18E-05	-1.94266
cv*spr	**cspr**	-0.00145	1.09E-06	0.001046	-1.38678
cv*dp	**cdp**	0.00019503	1.47E-08	0.000121	1.610407
spr*dp	**spdp**	0.00569	8.61E-06	0.002935	1.938674

The analysis effort has illustrated the problems of collinearity associated with the appearance of unexpected signs and high VIF values. A negative volume variability is meaningless. Whether the effort was also hampered by an inadequate model is unknown. However, given the problems of the underlying database as evidenced by the types of results obtained it appears that the Carrier Cost analysis presented in Docket No. R2005-1 is flawed and that additional analysis is needed. Pending additional analysis, I recommend the use of the unrestricted variability case of CC5.

V. COMMENTS ON DOIS DATA

The Postal Service has provided DOIS data and Table 3 summarizes some of the characteristics of the DOIS data received.[5]

Table 3: DOIS Data Summary

	R2005-1 Database	DOIS Database
Time Frame	1 Quarter	16 Quarters
Initial Observations	47,531.00	739,396
Final Observations	30,087.00	492,097
Number of ZIP Codes	145.00	125
Number of ZIP Code Days	1,545.00	21,700

Data Sources
(1) PAVOL database
(2) work.all database
(3) work.all database
(4) poolr database
(5) volume database in DOIS programs
(6) volume after processing in DOIS programs
(7) poolr database in DOIS programs
(8) poolr database in DOIS programs

Section 2 of the Library Reference OCA-Lr-l-4 details the initial processing of the DOIS database in preparation for use in this study. The database has only been available for a short time, and significantly more time would be required for a thorough analysis. Due to the limited amount of time, I have been able to apply minimal quality control procedures and have not yet made full use of all of the data.

[5] USPS-LR-L-160—DOIS Data Provided in Response to OCA/USPS-T14-8. The data were received on July 21, 2006.

A. Equations Based on the Approach Used in Witness Bradley's Analysis are Reported in Table 4.

Section 3 of OCA-LR-L-4 contains the relevant SAS equations.

Table 4: Equations from DOIS Data

	Letters	Flats	Seq	Parcels	DelivPts	Dens
ND1 Full Quadratic						
Volume Variability	0.22691	0.13338	0.00502997	-0.007179	0.60598	-0.022037
Marginal Cost	2.21	3.3	0.54	-13		
ND1 Restricted Quadratic						
Volume Variability	0.23142	0.13405	0.010221	0.00783	0.55965	-0.033233
Marginal Cost	2.24	3.33	1.09	14.14		
ND2 Full Quadratic						
Volume Variability	0.24157	0.11475	0.00834467	0.012114	0.5699	
Marginal Cost	2.37	2.9	0.9	22.13		
ND2 Restricted Quadratic						
Volume Variability	0.2216	0.12675	0.011466	0.027676	0.56552	
Marginal Cost	2.16	3.17	1.23	50.26		
Redefining DPS, with all other letters and Flats Combined in "Flats"						
ND3 DPS Full Quadratic						
Volume Variability	0.15657	0.20874	0.00437129	-0.006452	0.60343	-0.023468
Marginal Cost	2.26	2.84	0.47	-11.7		
ND3 DPS Restricted Quadratic						
Volume Variability	0.20707	0.19435	0.00929949	0.002188	0.53745	-0.027448
Marginal Cost	2.98	2.63	0.99	3.95		
ND4 DPS Full Quadratic, No Dens						
Volume Variability	0.1792	0.19124	0.00785935	0.009587	0.58482	
Marginal Cost	2.6	2.61	0.84	17.44		
ND4 DPS Restricted Quadratic, No Dens						
Volume Variability	0.21663	0.17681	0.010641	0.017295	0.53854	
Marginal Cost	3.12	2.4	1.14	31.33		

	Letters	Flats	Seq	Paracels	Curb	Central	NDCBU	Other
ND5TechnologyDelivPtsDPS Full Quadratic								
Volume Variability	0.18522	0.14973	0.016781	0.03893	0.10858	0.082005	0.027997	0.40155
Marginal Cost	2.72	2.08	1.82	71.93				
ND5TechnologyDelivPtsDPS Restricted Quadratic								
Volume Variability	0.18757	0.16177	0.014685	0.022148	0.12097	0.088523	0.027111	0.39358
Marginal Cost	2.76	2.24	1.6	40.93				
ND6 Technologydelivpts Full Quadratic								
Volume Variability	0.19372	0.12727	0.016056	0.035732	0.11285	0.083708	0.032864	0.40691
Marginal Cost	1.92	3.2	1.75	66.03				
ND6 Technologydelivpts RestrictedQuadratic								
Volume Variability	0.20222	0.12929	0.014729	0.022519	0.12411	0.086873	0.034912	0.40074
Marginal Cost	2	3.3	1.6	41.66				

Review of DOIS Equations from Table 4

ND1: This equation is modeled after witness Bradley's Full Quadratic as reported in Table 1. There is again a sign problem with parcels in the computation of volume variability in the case of the full quadratic.

ND2: This case eliminates the variable "Density." The sign problem is also eliminated in the values of the volume variabilities. There are no sign problems for the restricted quadratics equation, and the marginal costs appear to be reasonable in relationship to each other.

ND3, a DPS Case: In this case, DPS mail is denoted under the heading "Letters" with all other letter mail being treated as part of "Flats". Priority Mail is consolidated with parcels. There is a sign problem with parcels in the full quadratic. In the case of the restricted quadratic the marginal cost relationships are questionable on an *a priori* basis. This case has a density variable.

ND4: This is a DPS case with the full quadratic having the density variable eliminated.

ND5, technology case, DPS: Witness Bradley consolidated delivery points; that is, he did not differentiate delivery points by type of route in presenting his regressions. However, the database contained data by type of delivery point:

Curbline: A method of city delivery where the letter carrier, walking or in a vehicle, delivers to customer mailboxes located at the curb.

Central delivery: Delivery to several addresses at one delivery point, such as a collection box unit.

NDCBU: NDCBU denotes a centralized unit of more than eight individually locked compartments sized to accommodate the delivery of magazines, merchandise samples, and several days' accumulation of mail.

Other: This would include delivery to the door as well as all other forms of delivery not already covered.

Accordingly, for this regression the four types of delivery points, also referred to as delivery technologies, were used in lieu of consolidated delivery points. On an *a priori* basis, the marginal cost relationships in the results to ND5 are questionable.

ND6, Technology Case: There do not appear to be sign problems, and delivery points are presented in terms of technologies. This is not a DPS case. There is a reasonable relationship among the marginal costs and no negative variables for volume variability.

It is clear that several of the cases could be advocated as alternatives to the variabilities propounded by witness Bradley:

In the case of the Full Quadratic, ND1 and ND3 would be eliminated due to sign problems. ND5 would be eliminated due to concerns over the marginal cost relationships between DPS and other letter/flat mail. ND6 is a more general case in terms of delivery technologies and is the one I advocate.

In the case of restricted quadratic, ND6 also appears to be a reasonable case in terms of disaggregation of delivery points and reasonable relationships between marginal costs.

B. Accordingly, Based on the Limited Analysis of the Database up to this Point, I Advocate the Use of the ND6 Unrestricted Quadratic for Volume Variability.

Table 4: Preferred DOIS Equation

Variable Definition	Regression ND6 Variable Name	Estimated Coefficient	Regression ND6 Consistent Covariance	HC Std Error	HC t statistic
	Intercept	-11197	4359226.43	2087.8761	-5.362866
DPS+autoltrs+casltrs	let	2.31286	0.01315392	0.1146905	20.16609
let*let	let2	-0.0000035	1.48E-12	1.215E-06	-2.881385
Casflats+autoflats	cf	3.9017	0.06150982	0.2480117	15.73192
cf*cf	cf2	-0.00000333	4.21E-11	6.486E-06	-0.513433
Sequenced Mail	seq	2.88021	0.09038381	0.300639	9.580294
Seq Mail Squared	seq2	-0.00001475	1.76E-11	4.191E-06	-3.519386
Small Packages	spr	197.95689	145.280509	12.053236	16.42355
Sm Pkgs Squared	spr2	-0.04056	2.53E-05	0.0050335	-8.057999
let*cf	lf	-0.00001309	1.20E-11	3.465E-06	-3.77779
let*seq	lse	-0.00000352	2.97E-11	5.447E-06	-0.64623
let*spr	lspr	-0.00035885	4.11E-08	0.0002026	-1.771017
cf*seq	fse	-0.00002208	1.24E-10	1.112E-05	-1.985303
cf*spr	fspr	-0.00171	1.61E-07	0.000401	-4.264702
seq*spr	sspr	-0.00142	2.39E-07	0.0004886	-2.906084
bus curb+res curb dp	a	15.13237	8.30E-01	0.9110805	16.60926
bus cntrl +res cntrl dp	b	14.13536	7.02E-01	0.8378591	16.87081
bux ndcbu +res ndcbu dp	c	26.72165	2.00E+00	1.413828	18.90021
bus othr +res othr	d	26.88361	8.00E-01	0.8942271	30.06352
a*a	a2	-0.00053889	6.16E-09	7.845E-05	-6.868793
b*b	b2	-0.00011104	9.66E-10	3.108E-05	-3.572986
c*c	c2	-0.00042078	2.02E-08	0.000142	-2.96416
d*d	d2	-0.00014803	2.95E-09	5.434E-05	-2.724
let*a	leta	0.00010186	4.57E-10	2.137E-05	4.767171
let*b	letb	-0.00003881	1.05E-10	1.023E-05	-3.792423
let*cf	letc	-0.00017013	6.16E-10	2.481E-05	-6.857045
let*d	letd	0.00007091	3.22E-10	1.793E-05	3.953783
cf*a	cfa	0.0002058	1.23E-09	3.504E-05	5.873976
cf*b	cfb	-0.00018459	1.44E-09	3.798E-05	-4.859799
cf*cf	cfc	-0.00072321	3.12E-09	5.582E-05	-12.95533
cf*d	cfd	0.00030126	1.57E-09	3.962E-05	7.603489
seq*a	seqa	-0.00004967	9.10E-10	3.017E-05	-1.646473
seq*b	seqb	-0.00011845	2.31E-09	4.807E-05	-2.464248
seq*c	seqc	0.00031128	1.89E-09	4.353E-05	7.15097
seq*d	seqd	-0.00001268	1.22E-09	3.489E-05	-0.363385
spr*a	spra	-0.00788	3.95E-06	0.0019886	-3.96249
spr*b	sprb	-0.01244	2.53E-06	0.0015896	-7.825778
spr*c	sprc	0.00412	3.40E-06	0.0018443	2.233925
spr*d	sprd	-0.00236	2.22E-06	0.0014901	-1.583761
a*b	ab	-0.0009084	2.69E-08	0.000164	-5.540032
a*c	ac	-0.00057237	2.49E-08	0.0001578	-3.6279
a*d	ad	0.00111	2.92E-08	0.0001709	6.493158
b*c	bc	0.0051	6.98E-08	0.0002641	19.30825
b*d	bd	0.0013	1.92E-08	0.0001386	9.379349
c*d	cd	-0.00126	4.74E-08	0.0002177	-5.787346

VI. LESSONS LEARNED AND COMMENTS

A. The Theoretical Underpinnings of the City Carrier Cost Analysis Continue to be Weak.

Witness Bradley's initial approach advocated the use of a quadratic function with the variable density. The current analysis has advocated a quadratic form consistent with economic theory, involving the elimination of the density variable as well as disaggregation of technologies in terms of delivery points. Future work could consider whether some type of economic model, involving minimization of costs subject to some type of constraint could be developed. I have not yet used or examined all of the variables which could be considered, and whether currently unused variables could be combined with alternative models is an interesting issue.

B. One Variable Missing from the DOIS Database is Collection Volume.

Although witness Bradley's analysis had the collection volume variable, it is not available in DOIS. The nature of collection volume has changed significantly since the data collection effort for the initial database, as delineated in "It's a Pickup."[6] Although one tends to think of collection volume in terms of the carrier removing outgoing letters from a receptacle, the Postal Service now has an online program allowing customers to request that letter carriers pick up packages, including Express Mail, Priority Mail, and Parcel Post when combined with Express Mail or Priority Mail, on their regular delivery routes. This type of pickup appears to require different types of carrier efforts, such as leaving the truck to obtain the package and/or put the package in a different part of the truck, than had previously been the case in simply collecting mail from the mailbox.

[6] *Mailers Companion,* "It's a Pickup," USPS, March 2004.

Accordingly, the collection volume variability developed by witness Bradley appears now to be irrelevant. The development of an approach that accounts for carrier pickup in the estimation of volume variability is an issue for further consideration.

C. There has been Extensive Discussion of Some of the Problems Associated with the Database Used by Witness Bradley.

The problems associated with collinear data appear to have manifested themselves in the Table 1 analysis and to a lesser degree in the DOIS analysis.

The DOIS database extends over a period of four years, with data for a few days in each quarter. The time series is not continuous: one has a few days of data in each quarter for 16 quarters. Witness Bradley's analysis was also based on a short time series: two weeks of daily observations. The Durbin Watson statistic for witness Bradley's dataset indicated that the data were autoregressive. The data in the DOIS database are also autoregressive. The DOIS database might most aptly be characterized as similar to witness Bradley's database—covering a few days in a quarter; however, there are sixteen quarters. That is, there are 16 discontinuous sets of observations over a period of four years.

I have not made an adjustment for autocorrelation: a variety of possible adjustments were attempted and yielded unsatisfactory results. An area for future research will be autocorrelation issues.

Based on the preliminary analysis of the DOIS database up to this point, it appears that sign problems are less likely to occur and the expected cost relationships among variables seem more reasonable than has been the case with the CCSTS database.

VII. CONCLUSIONS ON MODELING EFFORTS

Some alternatives to witness Bradley's volume variabilities have been presented. In the case of the analysis using his two-week database, the Equation C5 alternative, which excludes the density variable, was presented; it is the one whose use I advocate if the Commission decides to base its consideration of volume variability on information developed from the CCSTS database.

In performing the modeling effort I considered a wide variety of alternatives to the equation proffered by witness Bradley in the modeling of City Carrier activities. These efforts frequently encountered sign problems, probably due to the underlying deficiencies of the database. Collinearity of the database is a problem, apparently making the application of a full quadratic model very difficult. In the consideration of restricted quadratic models, one frequently obtains relationships among the costs that, on an *a priori* basis, do not appear to be reasonable. Accordingly, I advocate that the Commission view Carrier Cost volume variability as an open question: improvement is needed.

Turning the DOIS analysis, I have made some progress in demonstrating that the database can generate a better analysis. I believe that the more general full quadratic model is preferable to the restricted quadratic, for the full quadratic considers interactions among variables. My recommended equation is Equation ND6, which I believe is the preferable equation for addressing City Carrier costs. The DOIS database has additional data which I have not yet analyzed. I have not yet determined whether additional conclusions can be developed from the database, but further analysis of the DOIS database is an area of inquiry that seems promising. Based on the information available up to this point I recommend the use of ND6.

If the Commission should decide to rely on the DOIS database for the generation of City Carrier Cost results, the information would be based on a database that could be updated on a quarterly basis, since the Postal Service collects the data on an ongoing basis. The ability to have timely updates is a significant advantage in terms of coverage and availability, as contrasted to *ad-hoc* surveys. The drawback to the use of the DOIS database is the lack of information about the collection volume of mail: the database does not have this information. However, witness Bradley's analysis of collection volume is also irrelevant: his data are based on the collection volume procedures in use in 2002, and there has apparently been a major transformation in the collection services offered for Express Mail and Priority Mail. Current collection volume efforts are different from those used when a carrier simply reached into a mail receptacle to extract a few pieces of mail. In conclusion it is clear that substantial progress can be made in the analysis of City Carrier cost volume variabilities.

There is a need for data to become available on a timely basis. Up to this time, we have had available to us in the case of the DOIS only two weeks per quarter over a period of five years. However, the Postal Service should have no problem providing a continuous series of observations for the full five years for the ZIP codes for which they have already furnished data.

Accordingly, I respectfully suggestthat the Commission recommend that DOIS data in the form furnished in USPS-LR-L-160 and as subsequently upgraded with additional ZIP code information be made available with coverage for all delivery days for five years for the relevant routes and ZIP codes, and that such data be updated on an ongoing basis.

Second, I respectfully suggest that the Commission recommend that DOIS data in the form of daily observations summed by quarter (or means in those cases where summation does not make sense) be made available by quarter over a five year time frame for all routes and all ZIP codes, in a manner similar to the data furnished in USPS-LR-L-160.

In asking for daily route and ZIP code data over a five year time frame for a small number of ZIP codes and routes, I am advocating the development of an enhanced database similar to that already furnished. Such a database does not include all ZIP codes and routes, for such a request would require an exceptionally large amount of computing resources. Accordingly, in order to get a complete overview of the Postal Service's City Carrier cost efforts I am also requesting data for all ZIP codes and routes, but summed (or averaged, as appropriate) on a quarterly basis.

I believe that the availability of the data requested would set a basis for an improved understanding of City Carrier costs, and that such data can be relatively easily updated, for it is derived directly from Postal Service records that are regularly maintained.