

UNITED STATES OF AMERICA
POSTAL REGULATORY COMMISSION
WASHINGTON, D.C. 20268-0001

Periodic Reporting
(Proposal Thirteen)

Docket No. RM2015-7

CHAIRMAN'S INFORMATION REQUEST NO. 1

(Issued January 6, 2015)

To clarify the Postal Service's petition to consider changes to analytical principles, filed December 11, 2014, the Postal Service is requested to provide a written response to the following questions.¹ The responses should be provided by January 12, 2015.

1. Please refer to USPS-RM2015-7/1 "Report on the City Carrier Street Time Study" (Report). On page 76, the Postal Service states: "Another issue that bears investigation is the relatively high marginal time for FSS mail.... The normal way to proceed with this investigation would be to separately estimate the regular delivery equation for FSS and non-FSS ZIP Code days. However, there is a problem with this approach because of the relatively small number of FSS ZIP Code days. With just 967 observations for FSS zones, multicollinearity becomes a serious problem. Even with the reduced-variable model, twelve of twenty-seven estimated coefficients have low t-statistics in an FSS-only model."

¹ See Petition of the United States Postal Service for the Initiation of a Proceeding to Consider Proposed Change in Analytical Principles (Proposal Thirteen), December 11, 2014.

- a. Please provide the results of the separately estimated delivery equation for FSS and non-FSS ZIP Code days, and the t-statistics and other diagnostic results used to evaluate it.
 - b. Please discuss how the low t-statistics of the separately estimated equation for FSS and non-FSS ZIP Code days limit its use as compared to the broader model approach. Specifically, please detail what t-statistic results would need to be achieved for the benefits of this approach to outweigh the concerns about the level of significance of the variables.
2. On page 34 of the Report, the Postal Service states: "Prudence dictated that imputation was considered only for ZIP Code days for which at least 80 percent of the routes reported volume."
 - a. Please describe how this 80 percent threshold was determined.
 - b. Please detail the impact of this choice of threshold.
3. Please discuss how Segment 7 cost pool data will be updated in subsequent fiscal years. If route evaluations will be used to form Segment 7 cost pools, will they be formed using the same criteria discussed in Section D of the Report (pages 9-14)?
4. Please refer to page 10 of the Report, where the Postal Service states that there were 37 route evaluations reporting gross street time over 12 hours and that since this outcome is not possible, these evaluations were dropped. Were these 37 routes examined to determine if any of them included pivot time from another route, which could suggest that more than 12 hours may have been a possible outcome?
5. Please refer to the excel files contained in the USPS-RM2015-7/1, WEBTOOL_Excel_Files_Parcel_Accountable_Study folder, and the variables listed in USPS-RM2015-7/1, Deviation_Acct_Variabilities_Model.SAS and the In_Receptacle_Variabilities_Model.SAS.

- a. Please define and identify the mail contained in the following column headings or variables:
 - i. PP_EXPRESS
 - ii. PP_PRIORITY
 - iii. PP_FC
 - iv. PP_OTHER
 - v. ON_DEMAND_PICKUP_STOPS
 - vi. ON_DEMAND_STOPS_MADE
 - vii. ON_DEMAND_EXPRESS ON_DEMAND_PRIORITY
 - viii. ON_DEMAND_FC
 - ix. ON_DEMAND_OTHER
 - x. ACCOUNTABLE DEVIATION IN_RECEPTACLE
 - xi. ON_DEMAND_EXPRESS
 - xii. ON_DEMAND_FC
 - xiii. ON_DEMAND_OTHER
 - xiv. ON_DEMAND_PICKUP_STOPS
 - xv. ON_DEMAND_PRIORITY
 - xvi. ON_DEMAND_STOPS_MADE
 - xvii. PP_EXPRESS PP_FC PP_OTHER
 - xviii. PP_PRIORITY PP_STOPS_MADE
- b. Please identify which of the variables listed above are carrier pickup items.
- c. Please confirm that carrier pickup mail or time associated with this mail are not included in the estimation of variabilities for the regular delivery model, the parcel/accountable model or the in-receptacle model.
- d. Please explain what role, if any, the variables listed above played in these models.
- e. Please identify the cost pools to which each of the variables listed above are assigned.

6. Please refer to the Report, Table 6 on page 14, rows entitled “Parcel & Accountable Delivery” and “Relay.”
 - a. Please confirm that the Parcel & Accountable Delivery cost pool as a share of the total increases at a faster rate than Relay between FY 2009 and FY 2011. If not confirmed, please explain.
 - b. Please confirm that that the Parcel/Accountable Delivery cost pool percent continues an upward trend from FY 2011 to FY 2013, while Relay declines during this period. If not confirmed, please explain.
 - c. Please confirm that the Network Travel cost pool has continuously declined as a share of the total between FY 2009 and FY 2011. If not confirmed, please explain.
 - d. Was a reduction in relay time or network travel time a goal of route reconfiguration, or were these reductions unintended consequences?
7. Please refer to the Report, Table 7, page 17.
 - a. Please confirm that the Column entitled “Ratio of Total Street Time to Attributable Street Time” is the ratio of (Direct + Indirect Street Time)/(Direct Street Time). If not, please explain.
 - b. Please confirm that this ratio is applied to both in-receptacle and deviation parcels. If not, please explain.
8. Please refer to the Report, Table 9, page 19.
 - a. Please confirm that cost pool shares for the activities listed in the table are based on direct plus allied street time shares, adjusted for in-receptacle parcels. If not, please explain.
 - b. Please explain whether the Segment 7 cost pool shares and/or the street time used to estimate shape variabilities in R2005-1 were based on direct street time for the activities listed in Table 9.
9. This question seeks information on the formation of three SAS datasets in USPS-RM2015-7/1: package_study_volume_masked_zips.sas7bdat, scanrouteday_masked_zips.sas7bdat, and carrier_scans_pa_study.sas7bdat.

- a. Please identify which files were created from raw scan data imported using SAS or another program, such as excel.
 - b. If the (SAS) files created from raw data were modified to create the datasets listed above, please provide electronic copies of the raw data, the SAS program(s) which created the final SAS dataset created from raw scan data, and the associated SAS logs. Please also explain each of the initial variables read in and describe the process of creating the final dataset.
 - c. Please provide a description of the variables EVENT_DATE1, EVENT_TIME1, EVENT_CODE, TRACK_BARCODE_ID, TRDV_DEV_ID, barcode2, and barcode1.
 - d. Please confirm there were 19 Event Codes: 01, 02, 03, 04, 05, 06, 12, 13, 14, 15, 23, 45, 51, 52, 53, 54, 55, 56, and 59.
 - i. If confirmed, please explain each component or action of the event, from start to finish, for each of the 19 event codes.
 - ii. In not confirmed, please list all event codes and explain the nature of each code.
 - e. Please confirm that barcode1 is the beginning of the activity associated with an event code and barcode2 is the end of the activity associated with that event code. If not, please explain.
10. Please refer to the Report, page 22, which describes the variables included in the regular delivery model.
- a. Please confirm that unless a small parcel was cased in the office, it would be considered an in-receptacle parcel for this study. If there were exceptions to this, please explain.

- b. For motorized routes or route segments, are small parcels that are not cased in the office typically mixed in with large parcels, or are they typically put in a tray, container, or other location in the vehicle separate from large parcels? If not, how are large and in-receptacle parcels placed in the vehicle?
 - c. For routes or route segments in which the carrier walks to the box to deliver the mail, are small parcels that are not cased in the office typically placed in carrier satchels? If not, please explain.
11. Please refer to USPS-RM2015-7/1, Collection_Study_Exhibit_2 - Instructions for Local study Coordinators, which states that “No Carrier Pickup parcels are included in the study.” Please confirm that carrier pickup parcels are not considered to be collected customer parcels. If not, please explain.
12. Please refer to USPS-RM2015-7/1, Report, page 28, which states that a “sample size of 300 ZIP Codes was determined to be the largest sample...” and “the collection study utilized a stratified systematic sample from a frame of 10,720 ZIP Codes.” Table 10 shows six strata, the number of ZIP Codes per strata in the sample, and the Share of delivery time per strata in the sample.
- a. Please add eight columns to Table 10 so that it also shows the following information:
 - i. the number of ZIP Codes in each strata in the sample frame;
 - ii. the number of ZIP Codes in each strata in the population;
 - iii. the share of street time in each strata in the sample frame;
 - iv. the share of street time in each strata in the population;
 - v. the average street time in each strata of the sample;
 - vi. the average street time in each strata of the population;
 - vii. the variance of street time in each strata of the sample; and
 - viii. the variance of street time in each strata of the population.

- b. Please describe the characteristics of the sampling frame – *e.g.*, the percentage of city carrier routes included, from which database(s) the sampling frame is constructed and how it is constructed, the representativeness of the sampling frame to the population of city carrier routes.
13. Please refer to USPS-RM2015-7/1, *Collection_Study_Exhibit_2 - Instructions for Local Study Coordinators*, which states that the sources of collection mail are 1) “[m]ail collected by regular and auxiliary carriers during delivery, [2] l]etter and flat mail accepted by clerks at the retail window; letter and flat mail collected by clerks from boxes and lobby drops; letter and flat mail dropped off by customers at the dock of Function 4 offices; [and 3)] mail collected by city carriers on dedicated collection runs.”
 - a. Please identify which of these sources were included in the study performed in Docket No. R2005-1.
 - b. Please identify which, if any, of these sources were previously included in general or express collections.
 - c. Please confirm that mail from the second source (collected by clerks) is not mail collected by city carriers. If confirmed, please explain whether this mail was included in the collection mail variable in the regular city carrier delivery model. If not confirmed, please explain.
 - d. Please reconcile the description of collection mail sources found in USPS-RM2015-7/1, *Collection_Study_Exhibit_2 - Instructions for Local Study Coordinators*, and the sources of collection mail described in USPS-RM2015-7/1, *Report*, at 30.
14. Please refer to USPS-RM2015-7/1, *Report* at 34, which states that “there were a sufficient number of reported collection volume days so that it was legitimate to use mean values over the days for which the route reported volume as the imputed values.”

- a. Please provide the mean and median values for each imputed value by route.
 - b. Please discuss the circumstances under which the median value would be a more appropriate imputation value than the mean.
 - c. Please discuss whether any other imputation methods were considered or employed. If any other imputation methods were tested, please provide the data, programs, and logs for these tests. In the response, please include a discussion of the following potential alternatives:
 - i. Simple random imputation, which randomly imputes missing values based on the observed data for the variable with missing values
 - ii. Matching, which, in this case, would impute values using observations with similar delivery characteristics (*e.g.*, uses the same delivery technology and has a similar business delivery ratio).
 - iii. Multiple Imputation (*See e.g.*, the Stata MI procedure).
 - d. Was an alternative analysis performed with a dataset that excludes incomplete records instead of imputing values for them? If so, please provide the data, programs, and logs from this alternative. If not, please detail the impact of using imputed data in the analysis.
15. Please compare the Report at 38, where the Postal Service states that “[t]he mean number of pieces collected per day per ZIP Code was 3,520,” and Table 14 (on page 39), which shows the mean collection volume as 173.6 pieces.
- a. Please reconcile these two measures.
 - b. Please provide a version of Table 14 using collection volumes per ZIP code.
16. Please refer to the Report at pages 43-44, which states that “[o]n any given day, the actual allied time may differ from the systematic allied time because of random factors, such as more or less traffic on the route, or a variation in a carrier’s personal needs time.”

- a. Please confirm that allied time is calculated from Form 3999, or route evaluation data from the sample period Form 3999 data. If not, please explain.
 - b. Please identify the source of allied time in Table 16 (on page 42). Please also identify:
 - i. The number of ziproute days used to calculate allied time;
 - ii. The date ranges used calculate allied time, by activity if they are not the same; and
 - iii. Whether or not these allied time values are limited to the study period and route days used in the regular delivery model before or after observation reduction.
17. Please refer to the Report at 73, which states that the Postal Service removed “some of those higher order terms, including cross-products... that are not statistically significant,” as a method of dealing with moderate multicollinearity.
 - a. Please discuss whether other methods of ameliorating the negative effects of multicollinearity were considered, and if so, why they were not employed. Please include a discussion of the following methods:
 - i. Centering variables around their respective means,
 - ii. Using a logarithmic transformation of variables,
 - iii. Variable Selection Methods such as:
 1. Sequential Regression
 2. Best Subset Selection
 3. Shrinkage Models
 - iv. Newey-West Robust Estimation (See Econometric Foundations, Mittelhammer, Judge, and Miller, at 392).

- b. Did the Postal Service test for model improvement after omitting variables to ameliorate multicollinearity using a measure that could compare goodness of model fit with all variables and without excluded variables, such as the Information Criterion or other measures of model fit?
 - c. If the Postal Service did perform versions of any of the above-mentioned tests of this sort, please provide all data, programs, and associated logs necessary to reproduce the results from each test.
18. Please refer to the Report, pages 84-85, where the Postal Service states that “[r]eview of the data for the 44 observations reveals nothing to suggest that the observations contain data errors or do not come from valid ZIP Codes that perform standard city carrier delivery operations... it is preferred to leave them in the data set when estimating the regression.”
- a. Did the Postal Service run a test to determine whether the F-values for the model with and without the 44 observations in question were statistically different?
 - b. Did the Postal Service run any goodness of fit tests to determine if there was a statistically significant difference in model fit? If not, please explain.
 - c. If the Postal Service did perform a version of the above-mentioned test, please provide all data, programs, and logs associated with necessary to reproduce the results from each test.
19. Please refer to USPS-RM2015-7/1, In_Receptacle_Variabilities_Model.SAS Please confirm that in-receptacle parcels that were accountables were only counted as accountables, and were not include in the model. If not confirmed, please explain.
20. Please refer to the Report at page 87, where the Postal Service states that “the package delivery time equations will include the proportions of the package and accountable deliveries made by mode.”

- a. Please confirm the rationale for this decision is provided on the same page, by the statement “[b]ecause there are so few package and accountable deliveries made on a route, it could well be that the route’s overall delivery type does not reflect the delivery technology for its package and accountable delivery.” If not, please explain.
 - b. Please explain what investigation was made to confirm the statement in part a. Please provide all data, programs, and logs used in the investigation.
 - c. Did the Postal Service run a regression using delivery technologies, rather than delivery mode, in a package delivery model? If so, please provide all data, programs, and logs necessary to reproduce the results.
21. Please refer to the Report at page 99 where the Postal Service states that the “average time per study scan ... [was] 12 seconds.” Please explain whether average scan time was the average of the elapsed time between all before and after delivery scans. If this was not the case, please explain how scan time for each type of delivery was measured.
22. Please refer to the Report at pages 99-100 where there is a discussion of the problem of zero-second scans due to the carrier’s failure to “hit the ‘A’ key for the begin activity scan.”
 - a. Please provide the raw data used to correct for this problem.
 - b. Please describe each variable in the raw data, and provide all data, programs and logs needed to make the adjustments to correct for this problem.

23. Please explain how the time inserting an in-receptacle parcel was separated from the time inserting mail (which was counted in the regular delivery model) jointly inserted into the same receptacle.
 - a. If the time was not separated, please explain.
 - b. If the time was separated, please provide the raw data and SAS or other programs and logs needed to identify scan and insertion time associated with this type of package delivery and to separate it from regular mail delivered at the same time to the same receptacle.
24. Please refer to the Report at pages 101-103, where the Postal Service discusses methods to resolve several data problems that were identified in the parcel/accountable and in-receptacle measurement of delivery time. Please provide all raw data, explain all variables, and provide all programs and logs used to make the data improvements discussed.
25. Please refer to the Report at page 103, where the Postal Service states that in response to observing many ZIP Code days with large package and accountable time, the Postal Service performed “[f]urther investigation of these large ZIP Code days [which] suggested that their cost generating process may be different from the rest of the ZIP Code days.”
 - a. Please describe the further investigation to which this quote refers.
 - b. If the investigation involved quantitative analysis, please provide all data and programs needed to reproduce the results.
 - c. Please also provide all data and programs needed to reproduce the investigation referred to on pages 105-109.

26. Please refer to the Report at pages 108-109 where it discusses the Grid Search Method employed to develop the optimum threshold value for the in-receptacle switching equation.
 - a. Please provide all programs and logs used to derive the optimum threshold for the in-receptacle equation.
 - b. Did the Postal Service use a cross validation method to confirm the significance of the estimated threshold value? If not please explain.
 - c. If a cross validation method was employed, please all raw data, explain all variables, provide all programs and logs used to cross validate.
27. Please refer to the Report at pages 113 where it mentions the Postal Service used the same Grid Search Method employed to develop the optimum threshold value for the equation for the deviation delivery switching equation.
 - a. Please provide all programs and logs used to derive the optimum threshold for the deviation delivery switching equation.
 - b. Did the Postal Service use a cross validation method to confirm the significance of the estimated threshold value? If not, please explain.
 - c. If a cross validation method was employed, please all raw data, explain all variables, provide all programs and logs used to cross validate.
28. This question seeks further information on the calculation of in-receptacle delivery time and its relation to regular delivery time.
 - a. Please confirm in-receptacle packages are delivered with the customer's "regular delivery mail." If not, please explain.
 - b. Was all in-receptacle scan time subtracted from regular delivery street time? If not, please explain.

- c. Please provide all raw data, explain all variables, and provide all programs, quantitative analyses, and logs used to calculate the time associated with the placement of an in-receptacle parcel in a customer's "box." If this was not calculated, please explain why not.

By the Acting Chairman.

Robert G. Taub