Rural Carrier Cost System Documentation

I. PREFACE

A. Purpose and Content

USPS-FY15-35 documents the statistical design of the Rural Carrier Cost System (RCCS). It contains documentation of the statistical design and the programs used to develop volume estimates for classes, products, and price categories of mail collected and delivered on rural routes. Also included are proportions, coefficients of variation (CVs), and confidence intervals for the estimates.

B. Predecessor Document

USPS-FY14-NP23, USPS-FY14-35, and Docket No. R2006-1, USPS-LR-L-12.

C. Corresponding Non-Public or Public Document

A nonpublic version of this document is provided as USPS-FY15-NP23.

D. Methodology

Documentation for the RCCS provided in USPS-LR-L-12, Docket No. R2006-1 included complete programs and descriptions for sample frame development and sample selection. Those programs and descriptions have incurred no substantive changes and are not reproduced herein. The estimation programs and output formats have incurred no substantive changes and are described in the RCCS System Documentation section below.

The FSS distribution keys for Standard Regular, ECR Basic, and ECR HD continue to be adjusted for the first three quarters, by using FY13 RCCS distribution key proportions to redistribute within those products to correct for FSS-marked pieces that do not have the proper ECR markings. Quarter 4 has not been adjusted due to the corrections to ECR marked pieces as a result of the Postal price change effective on May 31, 2015.

E. Input/Output

Volume estimates from the RCCS rely on no input data. Outputs from the RCCS are used as inputs to:

USPS-FY15-19	FY 2015 Delivery Costs By Shape
USPS-FY15-32	FY 2015 CRA "B" Workpapers (Public Version)
USPS-FY15-NP3	FY 2015 International Cost Segments and Components Report
USPS-FY15-NP14	FY 2015 CRA "B" Workpapers (Nonpublic Version)

II. ORGANIZATION

The relevant source code and outputs from the RCCS are provided on the accompanying CD-ROM. The 'RCCS_ReadMe_FY15(Public)' file describes the contents of the CD-ROM, which includes an Excel file containing proportions, coefficients of variation (CVs), and confidence intervals for RCCS estimates. Additionally, an overview of the statistical design and descriptions of the estimation processes are described in the System Documentation section below.

III. SYSTEM DOCUMENTATION

A. Overview

The RCCS is a continuous, ongoing statistical study, or probability sample of rural carrier route-days. Approximately 6400 RCCS samples are scheduled each Fiscal Year. For each selected route-day, a sample of mail is selected, and for each selected mailpiece, the class, product, compensation category, and shape of mail is recorded directly into a portable microcomputer using the Computerized On-Site Data Entry Systems (CODES) software.

The RCCS gathers data for distributing major portions of carriers' salaries, benefits and related costs to the various categories of mail for postal ratemaking purposes. Accrued carrier costs, available from payroll data, are total amounts and are not generally associated with any particular class of mail or service. Therefore, special methods are needed to determine the costs associated with the various products and services.

Rural carrier activity consists of delivering mail to and collecting mail from delivery receptacles or customers located on rural routes. In addition it includes certain activities such as providing Extra Services, collecting postage, and selling stamps. A rural carrier conducts almost all of the activities of a post office.

Rural delivery is organized and operated in terms of individual routes. Rural routes are divided into two broad categories, depending on the way the carrier is paid. Most rural routes are evaluated routes, that is, the rural route is evaluated in terms of time standards, and the carrier is paid a salary based on the evaluated time. Evaluated routes include the H, J, and K route types. The evaluated time is developed from route factors such as route length, boxes served, and quantity of mail delivered. Other routes include mileage routes, type M, paid on the basis of mileage on the route, and auxiliary routes, type A, paid on the basis of hours worked.

B. Use of RCCS Data in Cost Attribution

Total accrued costs for rural carriers are summarized in Cost Segment 10 (CS 10). The costs are divided into separate components for evaluated routes and other routes, based on payroll records.

The route factors are measured during the National Rural Mail Count, which is usually conducted annually. During the National Rural Mail Count, all mail for a large proportion of the rural routes is counted, and time measurements for other factors are evaluated. Therefore, factors related to volume (volume variable cost drivers) and factors independent of volume (fixed cost drivers) are measured during the National Rural Mail Count.

The volume variable costs of rural carrier work hours are determined by a variability analysis developed in accordance with the evaluated time and factors of workload

derived from the rural routes participating in the National Rural Mail Count. Volume variable costs are determined for each of the evaluated and other route components of Cost Segment 10.

Data from the RCCS are used to distribute volume variable costs across classes, products – including Extra Services, and price categories. The delivery portion of the RCCS (data collected via the CODES data collection system) provides the mail category data for the distribution of volume variable mail delivery costs. The PS Form 2848 portion of the RCCS provides the mail category data for the distribution of volume variable mail collection costs.

C. STATISTICAL STUDY DESIGN

The universe under study in RCCS is all mail being delivered on rural routes. A stratified, three stage sample design is used for RCCS. The details for each of the stages are listed below.

First Stage Sample (Route)

The first stage sample is a stratified random sample of route-days. Every rural route is assigned to one of two strata based upon whether the route is in a district having 20 or fewer rural routes or if it is in a district with more than 20 rural routes. Within each stratum, routes are geographically ordered, and a systematic random sample of routes is selected. Possible delivery dates (every Monday through Saturday, excluding holidays) are randomized, and systematically assigned to selected routes, to determine the route-days, or first stage sample units to be enumerated. This selection process ensures both geographic and temporal dispersion of the sampled route-days, and helps control workload at the district level.

Second Stage Sample (Office)

The second stage sample unit is the office, when the selected route serves one or more intermediate offices. In that case, one office served by the route is randomly selected. However, most (about 96%) of the routes do not serve intermediate offices, and are completely enumerated at the second stage.

Third Stage Sample (Mailpiece)

The third stage sampling unit is a mailpiece. Parcels and accountables are usually sampled with certainty. A systematic sample of letters and flats is selected. The data collector determines the skip interval (“s”) to be used – typically 10 – and the CODES software generates a random number “r”, between one and “s”. The data collector selects the “r^{th}” piece, and every “s^{th}” piece thereafter. The recommended skip interval is 10. Data collectors are allowed to change skip intervals as the need arises. The skip interval used is stored on each mailpiece record.

D. ESTIMATION AND VARIANCE

The RCCS produces two types of estimates—volumes and distribution keys (ratios). Estimates are generally computed on a quarterly and annual basis, and the annual volume estimates are the sum of the four quarterly estimates. This section provides the formulas used for FY 2015 to calculate the volumes, distribution keys, and the coefficients of variation associated with those estimates.

Notation:

y	variable of interest
w	weighting factor
h	postal quarter
i	compensation category domain
j	product or rate category domain
k	stratum
l	route-day
N	universe count – the number of routes in the stratum
n	completed tests in the stratum
d	delivery days in the postal quarter
s	skip utilized on a record (third stage weight)
r	intermediate office weight (second stage weight)
$\hat{Y}$	estimate of the total volume
$\hat{R}$	estimate of the distribution key
Cov	estimate of the covariance
$\hat{V}$	estimate of the variance
CV	estimate of the coefficient of variation

The weight applied to each record consists of four parts. First is the number of delivery days, d, in each quarter. Second is the first stage weight, indicated by N_{hk}/n_{hk}. Third is the intermediate office weight r (second stage), which is applied on a testid basis and is almost always 1. Finally there is the skip interval, s, which is applied to each record in a test. Dividing by 1000 causes the estimates to be reported in thousands. This weighting process yields unbiased estimates of mail volumes assuming any missing tests are missed at random.

The weighting factor is:

$$w_{hk} = \left(\frac{d_h \times N_{hk} \times r \times s}{n_{hk} \times 1000} \right)$$

Variates are defined as follows:

$$y'_{hijkl} = \begin{Bmatrix} y_{hijkl} \text{ if the unit is in the i}^{\text{th}} \textit{ and } j^{th} \text{ domains} \\ 0 \text{ otherwise} \end{Bmatrix}$$

$$x'_{hikl} = \begin{Bmatrix} x_{hikl} \text{ if the unit is in the i}^{\text{th}} \text{domain} \\ 0 \text{ otherwise} \end{Bmatrix}$$

The quarterly volume for the intersection of the ith compensation category domain and jth product is

$$\hat{Y}_{hij} = \sum_k \sum_l w_{hk} y'_{hijkl}$$

The quarterly volume for the ith compensation category domain is

$$\hat{X}_{hi} = \sum_k \sum_l w_{hk} x'_{hikl}$$

The quarterly distribution key for the ith compensation category domain and jth product is

$$\hat{R}_{hij} = \frac{\hat{Y}_{hij}}{\hat{X}_{hi}}$$

The annual volume for the ith compensation category domain and jth product is

$$\hat{Y}_{ij} = \sum_{h=1}^{4} \hat{Y}_{hij}$$

The annual volume for the ith compensation category domain is

$$\hat{X}_i = \sum_{h=1}^{4} \hat{X}_{hi}$$

The annual distribution key for the intersection of the ith and jth domains is

$$\hat{R}_{ij} = \frac{\hat{Y}_{ij}}{\hat{X}_i}$$

Variance

In computing the sampling error on the estimates, an ultimate cluster variance estimator is used. An assumption is made that the sampling error within routes is very small relative to the overall sampling error. Therefore, the variance formula used is similar to a single-stage total or ratio estimate, except that it omits the finite population correction (*fpc*) factor.

The estimated stratum mean by postal quarter for the intersection of the ith and jth domains is

$$\bar{y}'_{hijk} = \frac{\sum_l y'_{hijkl}}{n_{hk}}$$

$$\hat{S}^2_{hijk} = \frac{\sum_l \left(y'_{hijkl} - \bar{y}'_{hijk}\right)^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the intersection of the ith and j^{th} domains is

$$V(\hat{Y}_{hijk}) = \frac{w^2_{hk}\hat{S}^2_{hijk}}{n_{hk}}$$

The estimated variance for the quarterly volume for the intersection of the ith and j^{th} domains is

$$V(\hat{Y}_{hij}) = \sum_k V(\hat{Y}_{hijk})$$

The estimated variance for the annual volume for the intersection of the ith and j^{th} domains is

$$V(\hat{Y}_{ij}) = \sum_h V(\hat{Y}_{hij})$$

The estimated stratum mean by postal quarter for the intersection of the ith domain is

$$\bar{x}'_{hikl} = \frac{\sum_l x'_{hikl}}{n_{hk}}$$

$$S_{hik}^2 = \frac{\sum_l \left(x'_{hikl} - \bar{x}'_{hik}\right)^2}{n_{hk} - 1}$$

The estimated stratum variance for the quarterly volume for the ith domain is

$$V(\hat{X}_{hik}) = \frac{w_{hk}^2 \hat{S}_{hik}^2}{n_{hk}}$$

The estimated variance for the quarterly volume for the ith domain is

$$V(\hat{X}_{hi}) = \sum_k V(\hat{X}_{hik})$$

The estimated variance for the annual volume for the ith domain is

$$V(\hat{X}_i) = \sum_h V(\hat{X}_{hi})$$

The estimated stratum covariance between the quarterly volumes for the intersection of the ith and j^th domains is

$$Cov(\hat{Y}_{hijk}, \hat{X}_{hik}) = w_{hk}^2 \hat{S}_{yxhijk}$$

where

$$\hat{S}_{yxhijk} = \frac{\sum_l \left(y'_{hijkl} - \bar{y}'_{hijk}\right)\left(x'_{hikl} - \bar{x}'_{hik}\right)}{n_{hk} - 1}$$

The estimated covariance between the quarterly volumes for the intersection of the ith and j^th domains is

$$Cov(\hat{Y}_{hij}, \hat{X}_{hi}) = \sum_k Cov(\hat{Y}_{hijk}, \hat{X}_{hik})$$

The estimated covariance between the annual volumes for the intersection of the ith and j^th domains is

$$Cov(\hat{Y}_{ij}, \hat{X}_i) = \sum_h Cov(\hat{Y}_{hij}, \hat{X}_{hi})$$

The estimated relative variance (the square of the coefficient of variation) for the quarterly distribution key for the intersection of the ith and j^th domain is

$$(CV)^2(\hat{R}_{hij}) = \left(\frac{V(\hat{Y}_{hij})}{\hat{Y}_{hij}^2} + \frac{V(\hat{X}_{hi})}{\hat{X}_{hi}^2} - \frac{2Cov(\hat{Y}_{hij}, \hat{X}_{hi})}{\hat{X}_{hi}\hat{Y}_{hij}} \right)$$

The estimated relative variance for the annual distribution key for the intersection of the ith and j^{th} domain is

$$(CV)^2(\hat{R}_{ij}) = \left(\frac{V(\hat{Y}_{ij})}{\hat{Y}_{ij}^2} + \frac{V(\hat{X}_{i})}{\hat{X}_{i}^2} - \frac{2Cov(\hat{Y}_{ij}, \hat{X}_{i})}{\hat{X}_{i}\hat{Y}_{ij}} \right)$$

E. Quarterly Volume Estimates and Distribution Keys

Once the rural carrier data for an entire quarter have been validated, quarterly volume estimates and distribution keys are produced. The estimated volumes are compared with the same period from the previous year and with estimates from other statistical systems. Substantial differences between the reports are investigated for additional quality assurance. The quarterly distribution keys programs are as follows:

Quarterly estimation is a five-step process. First, monthly files are concatenated to form the quarterly file. Second, the weights used in the estimation procedures are produced. Third, collection mail volume estimates are calculated. Fourth, delivery volume estimates are calculated. Fifth, the Z file is produced.

ALDRAN.FY2015Qq.RURAL.CNTL(ALD399) is run to concatenate monthly files to form the quarterly file.

ALDRAN.FY2015Qq.RURAL.CNTL(RKEYA1) produces first-stage weights to be applied to the data received from the ALD399 program. It executes DSN=ALDRAN.FY2015Qq.RURAL.PARMLIB(ALD750JZ) that calculates the first-stage weights applied to all weighted volume estimates.

ALDRAN.FY2015Qq.RURAL.CNTL(RKEYA2) processes collection mail counts. It executes SAS code in DSN=ALDRAN.FY2015Qq.RURAL.PARMLIB(ALD750X1) to calculate the weighted volumes for collection mail data.

ALDRAN.FY2015Qq.RURAL.CNTL(RKEYB1) processes delivered mail counts. It merges 1) the stratum from the sample selection file, 2) the weights for each stratum from the weights file, and 3) the mail category information from the mailcode file onto the raw mail counts file. The program then sums up the information to two levels – mailcode, for external use, and CRA Bucket, for internal use. The output format used by RCCS is a 4-character mailcode variable {class, product, compensation category, and DMM shape (A=Letter, B=Flat, C=Parcel)} . Descriptions for each mailcode with estimated volumes are provided in the RCCS output file ALDRAN.LOTUS.RURAL.FY2015.MCODE.DATA.

<u>OUTPUT</u>

ALDRAN.FY15.RURAL.Qq.MCODE – Weighted data for each mailcode (Layout 001)

ALDRAN.FY15.RURAL.Qq.CRABKT – Weighted data for each CRA bucket.

ALDRAN.FY2015Qq.RURAL.CNTL(ZFILE2) reproduces sections of the ALD399 and RKEYA1 programs to reproduce data by testid, mailcode, and skip. The resulting quarterly SAS data files ALDRAN.RURAL.Z.FY2015Qq are combined into an annual

data set and sent to the Postal Regulatory Commission. The SAS data set extension is RAWCNTS.

OUTPUT

Quarterly Z File – ALDRAN.RURAL.Z.FY2015Qq (Layout 002)

F. Annual Estimates

Annual volume estimates are used to distribute costs to mail categories. The volumes are calculated by summing the quarterly volumes. The annual volumes program is executed from the following location: ALDRAN.FY2015.RURAL.CNTL. Two members are utilized to produce the annual volume estimates.

ALDRAN.FY2015.RURAL.CNTL(SMICOLL) is used to produce annual rural collection mail volumes.

OUTPUT

Annual volume report files: DSN=ALDRAN.LOTUS.RURAL. FY2015.COLL.DATA – for collected mail volume estimates.

ALDRAN.FY2015.RURAL.CNTL(SMIMCOD) is used to produce annual rural delivery mail volumes.

OUTPUT

DSN=ALDRAN.LOTUS.RURAL.FY2015.MCODE.DATA for delivered mail volume estimates (Layout 001).

Rural Mcode File – Layout 001

Position	Description
1 - 15	Volume
18 - 21	Mailcode
25 - 27	Bucket number (Layout 004)
30 - 80	Mailcode description

Rural Z File – Layout 002

SAS Variable Name	Description
BKTCHAR	Letter Character
BKTNUM	Bucket Number
COMPLETE	Total number of completed delivery tests in the quarter
COUNT	Unweighted count of mailpieces for the record
DELDAYS	Delivery days in the quarter
DELWGT	The first stage weight
F2848	Total number of collection mail forms completed in the quarter
F28WGT	Weight assigned to collection mail strata
INTWT	Intermediate office weight (second stage weight)
MASTER	Stratum universe count of routes
NAME	Description of mailcode
SCODE	Mailcode for the record
SEQ	Sequence number for the record
SKIP	Skip interval for record (third stage weight)
STRATUM	Stratum in which the route (testid) exists
NOPIECES	Total mailpieces for the record weighted to third stage
TESTID	Identification number for test
WGT	DELWGT/1000

Rural Collection File Layout - 003

1 - 2 Line Number
4 - 23 Rate Category
26 – 36 Letter, Card and Flat Volumes
38 – 48 Prepaid Parcels less than or equal to 2lbs Volumes
50 – 60 Other Parcels Volumes
62 – 72 Prepaid Parcels greater than 2lbs Volumes
74 – 84 Certified and Registered Mail Volumes
86 – 96 Carrier Pickup Volumes
98 – 108 Total by Rate Category

Bucket Descriptions Layout - 004

Bucket	Description
001	'FIRST-CLASS MAIL
111	' SINGLE PIECE LETTERS
112	' SINGLE PIECE FLATS
113	' FIRST-CLASS PARCELS (MD)
121	' PRESORT LETTERS
122	' PRESORT FLATS
123	' PRESORT PARCELS (COMP)
141	' FIRST-CLASS PACKAGES
151	' PRESORT CARDS
189	' TOTAL FIRST-CLASS MAIL
190	'PRIORITY MAIL
200	'EXPRESS MAIL-WITH SIGNATURE
201	'EXPRESS MAIL-NO SIGNATURE
210	'PERIODICALS
300	'STANDARD MAIL-REG AND NONPROFIT
311	' STANDARD OTHER LETTERS
312	' STANDARD OTHER FLATS
313	' STANDARD OTHER PARCELS
320	' TOTAL STANDARD OTHER
330	'
331	' ECR BASIC LETTERS
332	' ECR BASIC FLATS
333	' ECR BASIC PARCELS
350	'
351	' ECR HI-DENSITY LETTERS
352	' ECR HI-DENSITY FLATS
353	' ECR HI-DENSITY PARCELS
360	' ECR EDDM-RETAIL
361	' ECR SATURATION LETTERS
362	' ECR SATURATION FLATS
363	' ECR SATURATION PARCELS
370	' TOTAL ECR
380	'
390	'
395	'
399	' TOTAL STANDARD
400	'
401	'PACKAGE SERVICES
410	'
420	' PARCEL SELECT
430	' PARCEL SELECT LIGHTWEIGHT
442	' BOUND PRINTED MATTER FLATS
443	' BOUND PRINTED MATTER PARCELS

450	' MEDIA AND LIBRARY
460	' STANDARD POST
490	' TOTAL PACKAGE SERVICES
600	'
610	'U.S. POSTAL SERVICE
620	'FREE MAIL
630	'
700	'INTERNATIONAL MAIL
710	' LETTERS/CARDS/AO
720	' PARCEL POST
730	' CANADA EXPEDITED PARCEL
740	' CANADA EMS/XPRESSPOST
800	’
810	'
820	'
830	'
840	' INTL EMS NON-CANADA
850	'
880	'TOTAL INTERNATIONAL MAIL
890	'TOTAL DOMESTIC MAIL
900	'TOTAL ALL MAIL
901	' ACCT POSTAGE DUE
902	' ACCT BUSINESS REPLY
903	' ACCT CERTIFIED
904	' ACCT COD
905	' ACCT NUMBERED INSURED
906	' ACCT REGISTERED
907	' ACCT RETURN RECEIPT
908	' ACCT DELIVERY CONFIRMATION
909	' ACCT SIGNATURE CONFIRMATION
910	' ACCT ADULT SIGN. REQUIRED
911	' ACCT ADULT SIGN. RESTRICTED
912	'
919	' ACCT OTHER
920	' PREMIUM FWD SERVICE
990	'OTHER MAIL CLASS
999	‘COMPETITIVE PRODUCTS