

City Carrier Cost System (CCCS) Statistical and Computer Documentation (Public Version) (Source Code and Data on CD-ROM)

I. PREFACE

A. Purpose and Content

USPS-FY08-34 documents the statistical design of the City Carrier Cost System (CCCS). It contains printed and electronic 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 city letter routes.

B. Predecessor Documents

USPS-FY07-28 and Docket No. R2006-1, USPS-LR-L-11.

C. Methodology

Beginning with data collection in Quarter 1, Fiscal Year 2008, CCCS has been redesigned to align with the results of the City Carrier Street Time Study (CCSTS), which was reviewed by the PRC in Docket No. R2005-1. As a result of this redesign, the CCCS no longer collects mailpiece data by stop type. The new design eliminates the second-stage sample (by "Trip"), and it changes the third-stage sample from a clustered design (by "Stop") to a systematic design that samples mailpieces on a "skipped" basis (See Library Reference USPS-LR-L-11, page 2, filed in Docket No. R2006-1). The sample design of the CCCS is now similar to the design of the Rural Carrier Cost System (See Library USPS-LR-L-12, page 2, filed in Docket No. R2006-1).

Documentation for the CCCS provided in USPS-LR-L-11, Docket No. R2006, 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. However, as a result of the CCCS redesign, new estimation programs and output formats were developed similar to those from the Rural Carrier Cost System and are described in the System Documentation section below.

D. Input/Output

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

USPS-FY08-2 FY 2008 Public Cost Segments and Components
USPS-FY08-19 FY 2008 Delivery Costs By Shape
USPS-FY08-NP3 FY 2008 International Cost Segments and Components Report
USPS-FY08-NP12 FY 2008 Nonpublic Cost Segments and Components Report

II. ORGANIZATION

The relevant source code and outputs from the CCCS are provided on the accompanying CD-ROM. The 'CCCS_ReadMe_FY08' file describes the contents of the CD-ROM. 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

Documentation for the CCCS provided in USPS-LR-L-11, Docket No. R2006, 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 CCCS is a continuous, ongoing cross-sectional statistical study, or probability sample of city carrier route-days. For each selected route-day, a sample of mail is selected, and for each selected mailpiece, the class, product, and other characteristics are recorded directly into a portable microcomputer using the Computerized On-Site Data Entry Systems (CODES) software.

The CCCS gathers data for distributing major portions of carriers' salaries, benefits and related costs to the categories of mail for postal rate-making and related USPS management 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 mail categories.

City delivery is organized and operated in terms of individual routes. Because of their different operating characteristics, routes are divided for cost development into two groups: letter routes and special purpose routes. Letter routes account for more than 95 percent of street activity costs. The CCCS considers only regular letter routes.

B. Use of CCCS Data in Cost Attribution

Total accrued labor costs for city carriers are prorated between office activity Cost Segment 6 (CS 6) and street activity Cost Segment 7 (CS 7) on the basis of time proportion estimates obtained from the In-Office Cost System (IOCS). The data from CCCS are used for apportioning street activity costs to categories of mail. Carrier street activity consists primarily of delivering mail to customers located within the zones served by city delivery. In addition it includes certain other street-related carrier activities such as delivering relays, making collections and pickups, and moving mail to and from post offices and other postal facilities.

Data from the CCCS are used to distribute volume variable costs across classes, products – including extra services, and price categories. The delivery portion of the CCCS (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

2846 portion of the CCCS provides mail category data for the distribution of volume variable mail collection costs.

C. STATISTICAL STUDY DESIGN

The universe under study in CCCS is all mail being delivered on city letter routes. A stratified, two-stage sample design is used for CCCS. The details for each of the stages are listed below.

First Stage Sample

The first stage sample is a stratified random sample of route-days. Every city letter route is assigned to one of four strata based upon whether the route is a business or residential route, and also on the size of the route's post office (CAG A-E or F-L). 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 (Mailpiece)

The second 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 "rth" piece, and every "sth" 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 CCCS produces two types of estimates—volumes and distribution keys (ratios). A description of the estimates is provided in the overview. Estimates are computed on a quarterly and annual basis. The annual volume estimates are the sum of the four quarterly estimates. This section provides the formulas used for FY 2008 to calculate the volumes, distribution keys, and the coefficients of variation associated with those estimates.

Notation:

<i>y</i>	variable of interest
<i>w</i>	weighting factor
<i>h</i>	postal quarter
<i>i</i>	shape domain
<i>j</i>	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 (second stage weight)
\hat{Y}	estimate of the total volume
\hat{R}	estimate of the distribution key
Cov	estimate of the covariance
V	estimate of the variance
CV	estimate of the coefficient of variation

The weight applied to each record consists of three parts. First is the number of delivery days, d_h , in each quarter. Second is the first stage weight, indicated by N_{hk}/n_{hk} . 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:

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

Variates are defined as follows:

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

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

The quarterly volume for the intersection of the i^{th} and j^{th} domains is

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

The quarterly volume for the i^{th} domain is

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

The quarterly distribution key for the intersection of the i^{th} and j^{th} domains is

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

The annual volume for the intersection of the i^{th} and j^{th} domains is

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

The annual volume for the i^{th} domain is

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

The annual distribution key for the intersection of the i^{th} and j^{th} 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 i^{th} and j^{th} domains is

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

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

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

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

The estimated variance for the quarterly volume for the intersection of the i^{th} 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 i^{th} 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 i^{th} domain is

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

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

The estimated stratum variance for the quarterly volume for the i^{th} 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 i^{th} domain is

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

The estimated variance for the annual volume for the i^{th} 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 i^{th} and j^{th} domains is

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

where

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

The estimated covariance between the quarterly volumes for the intersection of the i^{th} 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 i^{th} 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 i^{th} and j^{th} domain is

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

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

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

E. Quarterly Volume Estimates and Distribution Keys

Once the city 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.

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. The quarterly estimation programs are as follows:

ALDRAN.FY2008Qq.CITY.CNTL(ALD299) is run to concatenate monthly files to form the quarterly file. The input files are the edited monthly files.

INPUTS:

Validated Monthly Data Files – DSN=ALDRAN.SHAPE.CCS08mm

Example for FY 08 month 10: ALDRAN.SHAPE.CCS0810

Only those tests that actually belong in the quarter (indicated by the first digit of the testid) are used for estimation. Below is a list of the months that should be used as inputs for the estimation for each quarter:

PQ1 includes months 10, 11, and 12.

PQ2 includes months 01, 02, and 03.

PQ3 includes months 04, 05, and 06.

PQ4 includes months 07, 08, and 09.

The program is generally run more than once, so various global analyses and edits may be performed. Additionally, weights for the second stage of sampling are applied to the data.

OUTPUTS:

The SAS dataset DSN = ALDRAN. CITY.SASDNS.SHAPE.FILE.FY2008Qq with SAS members RAWSHP and TESTCNT. RAWSHP contains all of the raw data records for the quarter and TESTCNT includes a listing of all test identification numbers for the quarter (used for weight development).

ALDRAN.FY2008Qq.CITY.CNTL(CKEYA1) produces first-stage weights to be applied to the data received from the ALD299 program. It executes the SAS code in DSN=ALDRAN.FY2008Qq.RURAL.PARMLIB(ALD750JZ) that calculates the first-stage weights applied to all weighted volume estimates.

INPUTS:

City Master frame for universe counts

DSN= ALDRAN.HQ059T01.CITYEXTR.PQ&PQ.FY&FY

Date file for number of delivery days in the quarter

DSN=ALDRAN.FY2008.PARMLIB(DATEPQq)

Data file for number of tests returned

DSN=ALDRAN.CITY.SASDNS.SHAPE.FILE.FY2008Qq

Sample file for stratum designation

DSN=ALDRAN.PS400T01.CITY.PQqFY08

File containing validated collection tests

DSN=ALDRAN.FY08.QqDET.CCS.EDIT

Flat file containing all previously calculated weights

DSN=ALDRAN.CITY.WEIGHTS(FY2008Qq)

SAS file with weights to be used later

DSN=ALDRAN.CCS2008.PQq.YTDWGT.DATA

OUTPUTS

SAS file with weights for processing data

DSN=ALDRAN.CCS2008.PQq.YTDWGT.DATA

Flat file containing weights for processing data

DSN=ALDRAN.CITY.WEIGHTS(FY&CC&FY.Q&PQ)

ALDRAN.FY2008Qq.CITY.CNTL(CKEYA2) processes collection mail. It executes SAS code in DSN=ALDRAN.FY2008Qq.CITY.PARMLIB (ALD750X7) that calculates the weighted volumes for collection mail data.

INPUTS:

File containing validated collection tests

DSN=ALDRAN.FY08.QqDET.CCS.EDIT

SAS file with collection weights used in estimation

DSN=ALDRAN.CCS2008.PQq.YTDWGT.DATA

OUTPUTS

Quarterly collection volumes for Key Distribution

DSN=ALDRAN.LOTUS.CITY.FY2008.PQq.COLL

ALDRAN.FY2008Qq.CITY.CNTL(CKEYB1) 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.

INPUTS

File with weights

DSN=ALDRAN.CCS2008.PQq.YTDWGT.DATA

File with mail category information for the mailcode output file

DSN=ALDRAN.FY2008Qq.SORTED.MAILCODE(CITYV2)

File with mail category information for the CRA bucket output file

DSN=ALDRAN.SASAUTOS.CTYMACRO.LIB2008(FORMATSF)

City quarterly data file (SAS file)

DSN=ALDRAN.CITY.SASDSNS.SHAPE.FILE.FY2008Qq (member

RAWSHP)

Sample file for strata

DSN=ALDRAN.PS400T01.CITY.PQqFY08

OUTPUTS

Weighted data for each mailcode (Layout 002)

DSN=ALDRAN.FY08.CITY.Qq.MCODE

Weighted data for each CRA bucket (Layout 003)

DSN=ALDRAN.FY08.CITY.Qq.CRABKT

ALDRAN.FY2008Qq.CITY.CNTL(ZFILE2) reproduces sections of the ALD299 and, CKEYA1 programs to reproduce data by testid, mailcode, and skip. The resulting SAS data file ALDRAN.CITY.Z.DATA.FY2008Qq is created for each postal quarter, converted from mainframe to PCSAS, and concatenated into one annual SAS data file. The SAS data set extension is RAWSHp.

INPUTS

City quarterly data file (SAS file)

DSN=ALDRAN.CITY.SASDNS.SHAPE.FILE.FY2008Qq (member
RAWSHp)

File with weights

DSN=ALDRAN.CCS2008.PQq.YTDWGT.DATA

Sample file

DSN=ALDRAN.PS400T01.CITY.PQqFY08

File with mail category information

DSN=ALDRAN.FY2008Qq.SORTED.MAILCODE(CITYV2)

OUTPUTS

Quarterly Z File

DSN=ALDRAN.CITY.Z.FY2008Qq

F. Annual Estimates

Annual volume estimates are used to distribute costs to categories of mail. The volumes are calculated by summing the quarterly volumes. The annual volumes program is executed from the following file: ALDRAN.FY2008.CITY.CNTL. Two members are utilized to produce the annual volume estimates.

ALDRAN.FY2008.CITY.CNTL(SMICOLL) is used to produce annual city collection mail volumes.

INPUTS:

The quarterly volumes files:

DSN=ALDRAN.LOTUS.CITY.FY2008.PQ1.COLL

DSN=ALDRAN.LOTUS.CITY.FY2008.PQ2. COLL

DSN=ALDRAN.LOTUS.CITY.FY2008.PQ3. COLL

DSN=ALDRAN.LOTUS.CITY.FY2008.PQ4. COLL

OUTPUT:

Annual volume report files for collected mail volume estimates:

DSN= ALDRAN.LOTUS.CITY.FY2008.COLL.DATA

ALDRAN.FY2008.CITY.CNTL (SMIMCOD) is used to produce annual city delivery mail volume estimates by mailcode.

INPUTS:

The quarterly volumes files:

DSN=ALDRAN.FY08.CITY.Q1.MCODE

DSN=ALDRAN. FY08.CITY.Q2.MCODE

DSN=ALDRAN. FY08.CITY.Q3.MCODE

DSN=ALDRAN.FY08.CITY.Q4.MCODE

OUTPUT:

The annual volume file for city delivered mail volume estimates by mailcode.

DSN=ALDRAN.LOTUS.CITY.FY2008.MCODE.DATA

City Z File Layout - 001

The variable names and explanations follow.

<u>SAS Variable Name</u>	<u>Description</u>
BKTCHAR	Letter Character
BKTNUM	Bucket Number
COMPLETE	Total number of completed delivery tests in the quarter
DELDAYS	Delivery days in the quarter
DELWGT	The first stage weight
F2846	Total number of collection mail forms completed in the quarter
F28WGT	Weight assigned to collection mail strata
MAILCODE	Mailcode for the record
MASTER	Stratum universe count of routes
NAME	Description of mailcode
SKIP	Skip interval for record (second stage weight)
STRATUM	Stratum in which the route (testid) exists
NOPIECES	Total mailpieces for the entry weighted by the skip interval
TESTID	Identification number for test
WGT	DELWGT/1000

City Mcode File – Layout 002

<u>Position</u>	<u>Description</u>
1 - 15	Volume
18 - 23	Mailcode
25 - 27	Bucket number
30 - 80	Mailcode description

Bucket Descriptions Layout - 003

Bucket	Description
001	'FIRST-CLASS MAIL
111	' SINGLE PIECE LETTERS
112	' SINGLE PIECE FLATS
113	' SINGLE PIECE PARCELS
121	' PRESORT LETTERS
122	' PRESORT FLATS
123	' PRESORT PARCELS
141	' SINGLE PIECE CARDS
151	' PRESORT CARDS
189	' TOTAL FIRST CLASS MAIL
190	'PRIORITY MAIL
200	'EXPRESS MAIL
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	'
361	' ECR SATURATION LETTERS
362	' ECR SATURATION FLATS
363	' ECR SATURATION PARCELS
370	' TOTAL ECR
380	'
390	' NOT FLAT MACHINABLE (NFM)
395	'
399	' TOTAL STANDARD
400	'
401	'PACKAGE SERVICES
410	' PARCEL POST SINGLE PIECE
420	' PARCEL POST BULK
430	'
442	' BOUND PRINTED MATTER FLATS
443	' BOUND PRINTED MATTER PARCELS
450	' MEDIA AND LIBRARY
460	'

Bucket Descriptions Layout - 003

490	' TOTAL PACKAGE SERVICES
600	'
610	'U.S. POSTAL SERVICE
620	'FREE MAIL - - BLIND & HNDC
630	'
700	'INTERNATIONAL SURFACE MAIL
710	' LETTERS/CARDS/AO
720	' PARCEL POST
730	'
740	'
800	'INTERNATIONAL AIRMAIL
810	' LETTERS/CARDS/AO
820	' PARCEL POST
830	' EXPRE PRIORITY
840	' EXPRESS
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 OTHER
920	' USPS PFS
990	'OTHER MAIL CLASS
999	'COMPETITIVE PRODUCTS

City Collection File Layout - 004

- 1 - 2 Line Number**
- 4 - 23 Rate Category**
- 26 - 36 Customer Outgoing Letter and Flat Volumes**
- 38 - 48 Customer Outgoing Parcel Volumes**
- 50 - 60 Collection Box Letter and Flat Volumes**
- 62 - 72 Collection Box Parcel Volumes**
- 86 - 96 Total by Rate Category**