

RECEIVED
SEP 24 4 54 PM '01
POSTAL RATE COMMISSION
OFFICE OF THE SECRETARY

**DOMESTIC REVENUE, PIECES & WEIGHT (DRPW)
STATISTICAL AND COMPUTER DOCUMENTATION**
(Source Code and Data on CD-ROM)

USPS-LR-J-16/R2001-1

Table of Contents

This document is a Category 1 library reference. It updates and replaces USPS-LR-I-21/R2000-1, USPS-LR-I-22/R2000-1, USPS-LR-I-23/R2000-1, USPS-LR-I-24/R2000-1, USPS-LR-I-27/R2000-1, USPS-LR-I-29/R2000-1, USPS-LR-I-31/R2000-1, USPS-LR-I-32/R2000-1, USPS-LR-I-33/R2000-1, USPS-LR-I-34/R2000-1, USPS-LR-I-65/R2000-1, USPS-LR-I-66/R2000-1, USPS-LR-I-67/R2000-1, and USPS-LR-I-68/R2000-1. It provides statistical and computer documentation for the Domestic Revenue, Pieces, and Weight System, which is a subject of the testimony of witness Pafford, USPS-T-3.

I. Overview	3
II. Statistical Study Design.....	4
1. First Stage Sample	4
2. Second Stage Sample	4
3. Third and/or Fourth Stage Samples	5
III. Creating the Sampling Frame.....	5
1. First Stage Panel of Finance Numbers.....	5
2. Second Stage MEPs and COUs	6
IV. Stratifying and Selecting the Sample.....	8
1. Stratification and Sample Allocation.....	8
A. First Stage	8
B. Second Stage	8
2. Sample Selection	13
3. Joint Scheduling	13
4. Sample Partition	14
V. Administering Tests in the Field.....	14
1. Downloading the Sample	14
2. Administrative Functions on Base Unit Computers.....	15
3. Collecting the Data	15
4. Transmitting the Data to San Mateo	16
VI. Data Validation and Editing.....	16
1. Creating the Check-in File.....	16
2. Editing, Validating and Processing of Data on the Mainframe.....	16
3. Influential Test Detection and Adjustment.....	17
VII. Estimation.....	18
1. Expanding to the Test Level	19
2. RPW Report Ratio Estimator.....	19
3. Variance and Confidence Limits.....	20
Appendix I: File Formats.....	22
1. Postal Accounting File with CAG	23
2. District, Site and Service Center Mapping File	23
3. Finance Number to District Mapping File	23
4. Postal Accounting File with CAG	23
5. MEP and CAG K&L Frame.....	24
6. Estimated Historic ODIS Proportions to Create Stratification Variables.....	25
7. Sample File.....	25

8. RPW Panel Offices.....	26
9. Labels for Coefficient of Variation.....	26
10. Historic Coefficients of Variation	26
11. CAG K&L Offices Out-of-Operation	26
12. Quarterly Test Date File.....	26
13. RPW Reference Volumes by MEP.....	27
14. Special MEP Frame for Stratification	27
15. Pre-Stratification Predicted Mail Volumes	28
16. Mail Category Rate File	28
17. Mail Category Master File.....	28
18. Multiple Special Service Mapping File	28
19. RPW Checkin Output File.....	29
20. File of SAS Code to Correct Data Errors Called in by Service Centers	29
21. RPW Panel Office File.....	30
22. Sampled Vs Total Panel Offices by CAG File.....	30
23. Mail Category to Summary Category File.....	30
24. ODIS Quarterly Sample Data.....	30
25. ODIS Sample File	31
26. RPW Sample File	31
27. ODIS Frame File for Reference Volumes	31
28. ODIS Letter and Flat Volumes.....	31
Appendix II: Flowcharts	32
1. Building the Sampling Frame Flowchart.....	33
1a. First Stage Panel of Finance Numbers.....	33
1b. Second Phase MEPs and COUs	34
2. Sample Selection Flowchart	35
3. Domestic RPW Processing Flowchart	36
Appendix III: Tables	41
Table 1: First-Stage Counts of Finance Numbers – Population and Sample Size	42
Appendix IV: Lists of Source Code and Data on CD-ROM.....	43

I. Overview

The Domestic Revenue, Pieces, and Weight (DRPW) system is a continuous, ongoing statistical sampling system. It produces estimates of revenue, pieces and weight for mail categories that are not available from the Postal Service's revenue accounting system and where data are not available from postage statements. The probability sampling system samples mail as it exits the postal system, with the exception of registered, insured, COD and Certificates of Mailing, for which a continuous probability sample of mail entering the postal system is administered.

All mail is partitioned into frame units, as part of a sampling frame. The frame unit for mail exiting the postal service is a mail exit point, or MEP, while the frame unit for mail entering the postal service is the combined originating unit or COU. The basic sampling unit is a randomly selected MEP (COU) and day within the Postal Quarter, or MEP-day (COU-day). A MEP-day is usually a 24-hour period over which mail either exits the postal system (MEP-day) or enters the postal system (COU-day). For each MEP-day (COU-day) in the sample, a subsample of the mail is selected. This would involve a census, mailpiece-skip interval, or container-skip interval process.

The primary sampling unit (PSU) is the finance number, which in many cases is a post office. In larger geographic areas, a finance number would include stations and branches. The finance number is the first stage of sampling. Finance numbers are stratified into Cost Ascertainment Groups (CAGs). A panel of finance numbers is maintained with the 11 CAG groups. CAGs A, B, and C is a certainty stratum, defined independently in each district. All MEP-days are eligible to be sampled in these strata. CAGs D through L are non-certainty first-stage strata, and only MEP-days within the panel of RPW offices are eligible for sampling. At the second-stage, for non-certainty first-stage strata, MEPs are stratified within each finance number. Beginning FY2000 the RPW design changed to a multi-stage cluster design. Prior to this, the design was a multi-stage phase sampling design. Within each second-stage MEP stratum a random sample of MEPs is selected, and a random test date assigned. The basic sampling unit is a MEP-day. Two additional levels of sampling can occur below the MEP-day level. A third stage of sampling occurs when mailpiece-skip subsampling is applied. This is a systematic random selection of mailpieces applied by the data collector to the mail. A third and fourth stage of sampling occurs when a container-skip interval is applied to containers, and then a mailpiece-skip interval within selected containers.

For selected mailpieces, revenue, volume and weight by mail category are recorded on laptop microcomputers. Weight is usually captured automatically using electronic scales connected to the laptop computers. The data are reviewed at the laptop level and again after it is uploaded to the district office in

its Base Unit computer system. After review, the data are transmitted electronically to a mainframe computer, where it is again checked-in and edited. The data are then weighted, checked for influence, and adjusted (re-weighted) if necessary. Finally, the data are aggregated to a reporting level. Estimates are provided by mail category level for revenue, pieces and weight.

II. Statistical Study Design

The universe under study consists of all mail exiting the postal system, and all mail entering the postal system for registered, insured, COD and Certificates of Mailing. DRPW is a multi-stage cluster design. The sampling frame is the list of finance numbers and MEPs and COUs within finance number. MEPs and COUs are stratified within first-stage panel finance numbers. A random sample of MEPs is selected within each stratum, within each finance number, each Postal Quarter (PQ), and a date randomly assigned for conducting the test: usually a 24-hour period. For MEPs with mail volumes larger than can be reasonably recorded within a data collectors work schedule, subsampling occurs to avoid delays in delivering the mail.

1. First Stage Sample

The first-stage sampling unit is the finance number. The first-stage sample frame is a list of all finance numbers. A given finance number corresponds to a post office in many cases. Finance numbers are stratified into Cost Ascertainment Groups (CAGs) strata based on total revenue receipts for the most recent fiscal year. CAG stratum 1 (CAGs A-C) is a certainty stratum; all finance numbers are eligible for sampling at the second stage. In the remaining CAG strata, a set of finance numbers form a panel that remains relatively stable from year-to-year. CAG D and CAG E panels were reselected through probability sampling, effective PQ 1 FY 1999. CAG F through CAG L panels were similarly reselected effective PQ 1 FY 2001. The panel was originally chosen in approximate proportion to the total revenue receipts for all offices in the CAG.

2. Second Stage Sample

The second phase sampling frame is the list of Mail Exit Points (MEPs) within the selected finance numbers. The MEPs from the selected first stage finance numbers are stratified differently depending on the CAG strata. For certainty CAG stratum 1 (CAGs A – C), all MEPs within a district are stratified using multivariate cluster algorithms. For non-certainty strata (CAGs D – L), MEPs are stratified within each finance number based on expected average daily volume by shape and class. Within each second-stage stratum, a random sample of MEPs is selected. Test dates are randomly assigned through an algorithm that jointly considers other statistical programs' test schedules. Details

of the random date assignment process are contained in LR-USPS-J-17/R2001-1.

3. Third and/or Fourth Stage Samples

When a selected MEP has a large volume of mail on the test day, a subsample of the mail may be selected to facilitate counting and recording the mail without causing delivery delays. Subsampling involves either 1) container-skip subsampling - systematic random selection of containers (third stage) and then mailpieces within selected containers (fourth stage), or 2) mailpiece-skip subsampling - the systematic random selection of mailpieces across all containers (third stage). Data collectors are instructed to sample all mailpieces on the test day as the first option (Census), conduct a mailpiece-skip subsampling as the next option, and container-skip subsampling as the last option. Container-skip subsampling gives the largest variance and therefore is the least preferred method. Detailed procedures are described in Chapter 3.5 through 3.7 of the Handbook F-75, Data Collection User's Guide for Revenue, Volume and Performance Measurement Systems, USPS-LR-J-23.

III. Creating the Sampling Frame

The sampling frame of finance numbers is established prior to the postal fiscal year. The RPW panel of offices from the prior year is compared against a postal accounting file containing a list of all finance numbers in the nation and their Cost Ascertainment Group (CAG) value. CAG is updated and district staff notified of changes in the panel of RPW offices for the upcoming fiscal year.

Once the RPW panel is developed, the data base of all MEPs and COUs maintained in the district office (panel and non-panel offices) is captured or 'frozen' to begin the process of creating a sampling frame.¹ The process begins by checking the database of MEPs and COUs for such things as COUs for every finance number, and second-phase stratification information present for every frame unit. Updates to the database are made if necessary.

1. First Stage Panel of Finance Numbers

The sampling frame for the first-stage is initially established prior to the beginning of the postal fiscal year and updated each quarter. Seven weeks prior to the beginning of quarter one, a snapshot of the MEP and COU database is taken using program SNAPSHOT. Program CAGKL1A generates printouts that are sent to each district office. These identify offices whose MEPs must

¹ The MEP database is maintained by the MEP DBMS System, documented previously in MC 96-3, SSR-53, SSR-54, and SSR-55.

potentially be dropped and added to the MEP database. The districts have until five weeks prior to the beginning of the quarter to make changes before the frame is 'frozen'. Information on potential drops and adds comes from a comparison the districts MEP and COU database (input file HSFRANT.MASTER.TEST.FYyy) and an official postal service accounting file (input file HSF.ADGNYN.HQ070D01.FYyy). The accounting file provides information about each finance number, such as the most recent Cost Ascertainment Group or CAG value. Offices 'advancing' to CAG A, B, or C (i.e., CAG D or lower) and not in the RPW panel become RPW panel offices. This is because CAG A, B, and C are, by definition, certainty strata. Offices 'relegating' to CAG K or L (i.e., CAG J or higher) must have their MEPs removed from the MEP database.²

After frame freeze, program PANELA creates the RPW panel office file using the RPW panel from the previous year (HSF.HQMAND07.SMPOFFyy-1.PQ4), the frozen frame (HSFRANT.MASTER.TEST.FYyy01), and the new CAG values obtained from input file HSF.ADGNYN.HQ070D01.FYyy. Program PANELB establishes the population and sample counts for the first stage expansion factors. Program CAGKL2 provides a file of CAG K and L offices used in sample selection.

CAGKL1A is re-run for postal quarters two, three and four using the same input file HSF.ADGNYN.HQ070D01.FYyy as was used for quarter1, and printouts mailed to the districts. In this way, the district offices' MEP maintenance is continually checked.

2. Second Stage MEPs and COUs

Five weeks prior to the beginning of the Postal Quarter the sampling frame is 'frozen'. Frozen means that each district is locked out of making any further changes to their MEP DBMS. The output file produced for the frozen frame is HSFRANT.PS765T01.MASTER.FYyyyy. GENVSAM produces a copy of this frame for RPW processing.

PROG02 subsets the frozen frame to just those MEPs and COUs for the RPW panel offices and adds the headquarters maintained list of CAG K and L frame units. This program creates three subset MEP files: CENFRAME, SPCFRAME, and CAGFRAME. The CENFRAME file contains all CAG A through CAG C MEPs. The CENFRAME MEPs are processed through the multi-variate clustering stratification programs that allow for strata to be formed based on the predicted class volumes. The SPCFRAME file includes originating, APO/FPO, predominantly Standard Mail (A) and Periodical MEPs (MEP Type 5), and

² CAG K and L MEPs and COUs are not maintained in the MEPs database, but instead are included in the sampling frame by headquarters through the postal accounting file – HSF.ADGNYN.HQ070D0.FYyy.

unstable MEPs (MEP Type 6). The CAGFRAME file contains all CAG D through CAG J MEPs that are not special MEP types. These SPCFRAME and CAGFRAME MEPs are assigned special strata codes, and are NOT processed through the stratification programs which group MEPs based on the multivariate distances between predicted mail class volumes (MEPs from the CENFRAME file). PROG02 reads various input files. The HSFRAN.PS001D01.ZIPTABLE.ORFEO file is needed to assign the customer service district code. The HSDHQN.PS001D01.ODIS.FINFILE is needed to assign the customer service district codes for APO/FPO MEPs in the NJ Bulk and JFK districts (assigned by finance number). The HSF.MEM.CAG.KL.FY2000 file provides the CAG K and L MEPs and COUs since the field offices do not maintain these on their MEP DBMS (they come from Minneapolis Accounting at the beginning of the Fiscal Year). The HSF.MEP.CAG.KL.DELETE.Fyyyyy file provides 'out-of-business' finance numbers. In addition to these processes, the program performs various edits on the data, such as deleting MEPs that are strictly defined as incoming registered MEPs, and correcting for nonnumeric MEP ZIP Codes. This program is also where problems with frame units are corrected. Fixes for such things as duplicate MEPs, MEPs with incorrect reference volume, and incorrect MEP Types are included in SAS data step MEPS3B.

PROG02AL creates a frame file that includes MEPs in RPW panel offices, and non-panel offices. The output file is used to later apportion estimated volumes (at the district level) and ultimately target coefficients of variation (CVs) later used in the sample allocation programs. SAS Code in SAS data steps MEPS3A and MEPS3B from PROG02 are manually copied to the same SAS data steps in PROG02AL.

PROG03 checks the frame for inconsistencies such as duplicate MEPs, large changes in the numbers of MEPs in a district from a previous time period, and finance numbers without MEPs and COUs. This information, along with information gathered from similar checks in the ODIS sample selection system, is used to generate SAS code to fix identified problems. This SAS code is inserted into PROG02. PROG02 is then re-run to generate a final file of the MEP and COU frame for RPW sampling.

The number of MEPs by MEP Type and district are summarized and compared with the previous Postal Quarter in PROG04. Comparisons are made on total MEPs, COUs, CAG K&L MEPs, and APO/FPO MEPs.

IV. Stratifying and Selecting the Sample

1. Stratification and Sample Allocation

A. First Stage

The first-stage sample of finance numbers is stratified into nine Cost Ascertainment Group (CAG) strata based upon the annual revenue of each office as reported under the Accounts Reporting System. FY2000 sampling used the CAG value as of FY 1998. Section III.1 describes program execution steps for updating the CAG value.

All offices that were in CAG A, CAG B, or CAG C at the close of Fiscal Year (FY) 1999 make up CAG stratum 1. Panel offices that change CAG are assigned to their new CAG stratum. Due to the general tendency over time for finance numbers to move upward in CAG, particularly in CAGs K and L, periodic reselection or replenishment of CAGs is made by random selection. CAG D and CAG E panel was rotated at the start of FY 1999, and CAGs F through L were rotated at the start of FY 2001. Table 1 in Appendix III shows the first-stage strata and sample size by Postal Quarter.

B. Second Stage

Within each finance number selected in the first-stage sampling, the list of all COUs and MEPs is obtained from the MEPs database. All possible mail exit points as well as all possible COU entry points for registered, COD, Certificates of Mailing, and insured mail are identified. Separate strata are defined and samples drawn each Postal Quarter within each district for each of the following special MEP types: APO/FPO, special delivery, COU, CAG K&L, and unstable. Separate strata are also defined and samples drawn each Postal Quarter within each first-stage panel office for CAG D through CAG J MEPs. For the remaining MEPs (CAG A through CAG C), strata are computationally defined within district using multivariate clustering algorithms. Second stage universe and sample sizes, and strata definitions are documented in Appendix IV.

Stratification begins with the execution of PROG06. This program computes the predicted First Class letter mail class volume for use as a stratification and allocation variable in later programs involving the CENFRAME file (all destinating CAG A through CAG C MEPs except special MEP types). It also prepares other stratification variables, with the total set including First Class letters, parcels, flat/PPs, priority, and BRM/Accountables. To arrive at predicted First Class letter volume, letter shape reference volumes from the MEP DBMS are multiplied by the proportion of letter volume that is First Class. This is done for each MEP. The estimated proportions, that is, the proportion of letters that

are First Class are derived from historic estimates. Parcel, flat/IPP, Priority and BRM/accountable mail reference volumes are taken straight from the MEP DBMS.

PROG06AL is the same as PROG06 except that the entire frame (RPW panel and non-panel offices) is used. That is, stratification variables are defined for the entire frame.

PROG07 and PROG08 employ the use of multivariate clustering methods to stratify the MEP sampling frame (for all MEPs except special MEP types) from the CENFRAME file. In PROG07, strata are determined using stratification variables. These include predicted First Class mail letter volume, Priority mail reference volume, BRM/Accountable Mail reference volume, flat/IPP reference volume, parcel reference volume, and combined RPW test and travel time (what is actually used is a value of 0.5 if combined time is under 2 hours, 1.0 if combined time is between 2 and 6 hours, and the one-half the ceiling of two times the combined time divided by 6). In CLUS1CSD, each variable is normalized to a mean of zero and a variance of one. A k-means clustering algorithm is applied to the multivariate normalized data to create a large number of clusters. Fifty clusters are chosen, and this is done independently in every district. In PROG08 Ward's clustering algorithm is used to combine clusters created in PROG07 into strata. Ward's algorithm is a multi-step, agglomerative clustering procedure that starts with n clusters and at each step selects two clusters to combine, so that after n steps there is only one cluster. The two clusters that the algorithm selects at each step to combine are those that minimize the resulting within-cluster variability. When the Ward's algorithm is applied to clusters of MEPs, each step of the algorithm produces a candidate stratification. The one selected is based on the estimated variances resulting from a fixed-cost optimal allocation in the associated strata.

Once the clusters are collapsed, strata values are defined separately for each district. Strata values range between 400-499.

PROG08 outputs four basic files. The POP file contains stratum standard deviations for each item. There are 5 records (one for each mail volume stratifier) for each allocation strata. The STRATA file contains stratum identifiers, stratum sizes, and average costs per stratum. The ITEMS file contains item (i.e., estimate) labels, target coefficients of variation, and item estimates. The FREEZE file contains stratum identifier and sample size for each stratum with a fixed (i.e., frozen) sample size. "Cost" is input in the DOMAIN file. This file contains the maximum cost for each cost-constraint domain (i.e., District). Cost is defined as the number of delivery days in the AP times the average historic cost per day. Cost does not change from postal quarter to postal quarter. It represents the test workload in a District.

PROG09 computes the predicted First-Class mail letter volume used in stratification and allocation for special MEP types coming forward from the SPCFRAME file (see program documentation for PROG02) . It also prepares other stratification variables, with the total set including First Class letters, parcels, flat/IPP, priority, and BRM/Accountables. To arrive at predicted First Class letter volume, letter shape reference volumes from the MEP DBMS are multiplied by the proportion of letter volume that is First Class. This is done for each MEP. The estimated proportions, that is, the proportion of letters that are First Class is derived from historic estimates. Parcel, flat/IPP, Priority and BRM/accountable mail reference volumes are taken straight from the MEP DBMS.

Based on the stratification volumes, MEP types, and CAG, strata codes are assigned in the following manner. Strata 350 through 355 include APO/FPO MEPs based on First Class predicted and parcel volume. Strata 370 and 371 are CAG A – C COUs with total reference volume of less than 100 pieces, and greater than or equal to 100 pieces. Strata 380 and 381 are CAG K and L MEPs and COUs. Stratum 392 is assigned to MEPs with MEP Type 5 (predominantly Standard Mail (A) and Periodicals), while stratum 395 is assigned to MEPs with MEP Type 6 (unstable MEPs).

Strata 500-599 are assigned to CAG D through CAG J MEPs within each finance number within each district (from the CAGFRAME file). There can be up to eight strata within each finance number within each district. However, offices in these CAGs typically have only one, two, or three MEPs, resulting in few strata within each office. The strata include: accountables (80% of reference volume is accountable mail); 1-C letters (over 90% of reference volume is predicted 1-C letters); Std (A) flats (90% predicted Std (A) flats); mixed mail (90% of reference volume mixed letters and flats); priority (over 1000 priority pieces and 80% of reference volume priority); small 1-C letters; small Std (A) flats, and small mixed MEPs.

PROG09 outputs three basic files (required for the CHROMY sample allocation algorithm). The POP file contains stratum standard deviations for each item. There are 5 records (one for each mail volume stratifier) for each allocation strata. The STRATA file contains stratum identifiers, stratum sizes, and average costs per stratum. The ITEMS file contains item (ie., estimate) labels, target coefficients of variation, and item estimates. The STRATIFY file contains the MEPs with the attached strata values.

PROG10 classifies the NJ Bulk and JFK districts APO/FPO MEPs into strata codes ranging from 350-359. Strata are pre-assigned here as in PROG09. PROG10 outputs three basic files. The POP file contains stratum standard deviations for each item. There are 5 records (one for each mail volume stratifier) for each allocation strata. The STRATA file contains stratum identifiers, stratum sizes, and average costs per stratum. The ITEMS file contains item (ie.,

estimate) labels, target coefficients of variation, and item estimates. The STRATIFY file contains the MEPs with the attached strata values.

PROG10A computes the proportion of RPW frame volume to total frame volume by district for each stratification variable. The proportions are output to a file and used in PROG10B. This proportion is the amount of volume accounted for by the MEPs from the CENFRAME file (or that, which is sampled using the CHROMY algorithm).

PROG10B creates an output file of estimates and CVs by district, which are input into the sample allocation program - PROG17. CVs are set in this program, but are later over-ridden in PROG17. Estimated district level average daily volumes (usually average daily ODIS volume calculated over several quarters) is input from an external file. Total estimated AP volume is calculated by multiplying by the number of days in the AP. Using the proportions established in PROG10A, the volumes are reduced by that amount (by stratification variable). This is necessary, because one of the constraints in the CHROMY sample algorithm is the target variance (ie., the square of the product of the target CVs times the estimated AP volume).

PROG11 merges the POP, STRATA, and ITMNUM files from PROG08 through PROG10. It also merges an external file of coefficient of variations (CVs) and estimated mail volumes. Output files include a merged POP, STRATA, ITEMNUM, and ITEMS (CVs and volumes).

PROG12 prints a frequency distribution of originating MEPs and small panel office MEPs, predominantly Standard Mail (A) and Periodical MEPs (MEP Type 5), and unstable MEPs (MEP Type 6) by District. The total number of originating MEPs must be input into PROG13 to properly allocate originating MEP samples among originating MEP strata.

PROG13 begins the process of sample allocation. This program derives the sample allocation for COU strata 370 – 372, and CAG D – J MEP strata 500-599. For COUs, the population count is entered in a macro parameter, and is obtained from a report out of PROG12. A target national sample size of 500 COUs per AP is typically specified. The COU sample size in each district is derived as a proportion to the number COUs, and then this sample is spread to the strata 370 – 372 in proportion to the number COUs in the strata. Samples are fixed (frozen) by district and strata, and output to the COU freeze file. For CAG D – J strata 500-599, either one, two or three tests are allocated per AP to each strata depending on the frame size. Strata counts are fixed (frozen) and output to the CAG freeze file. The sample sizes for the predominantly Standard Mail (A) and Periodical MEP stratum, and the unstable MEP stratum are set in PROG18.

PROG14 is used to add fix, or “freeze” the sample size in specified strata. Data step A is edited in the SAS code for this purpose. The program then takes these and merges them to the originating and small panel office frozen strata sample sizes and outputs a new freeze file. This program may be run iteratively after CHROMY to arrive at the desired frozen strata. However, on the first run of this program, the additional frozen strata sample sizes are “commented out”, allowing just those fixed sample sizes from the originating and small panel office MEP strata to go forward.

PROG15 sets up the final files that are read into CHROMY. Because originating and unstable MEP sample sizes, and strata 500-599 samples sizes are fixed in earlier programs, the associated strata level records and domain costs are removed from the output files that are used in PROG17 (CHROMY program). More specifically, in order for CHROMY to allocate samples against a domain cost (cost per district), the cost for originating and unstable MEPs, and strata 500-599 must be subtracted from the total domain cost. This function is performed by outputting an adjusted cost file (cost per day per domain or district). The same type subtraction is done for the POP, STRATA, and FREEZE files, except what is deleted are the strata records associated with originating and small panel office MEPs. Strata allocation sample sizes for originating and unstable MEPs, and small panel offices are output and later combined in PROG18 with sample sizes from the CHROMY program (PROG17).

PROG17 is often called CHROMY, in honor of the Dr. Chromy who wrote the sample selection algorithm. This program determines the sample size in each stratum with each district. Target CVs for the Accounting Period (district level) are set at 20% for letters and priority, 50% for parcels, 40% for flats/IPPs, and 70% for BRM/Accountables. The program implements an algorithm for finding the minimum-cost stratified sample that satisfies multiple precision constraints. The program is iterative. The output includes a file of sample sizes by district and strata. Also output is a temporary SAS data set containing the domain (district) and cstrat value. This is printed as part of a table in the FRMSMP program.

PROG18 is setup to make final minor changes to the district/strata sample sizes. It is to be run first “commenting out” the ADJ SAS data step. This step is used later to read in a flat file of adjusted sample sizes. After commenting out the ADJ SAS step and running the program the first time, use the tables to compare the district sample sizes to the maximum domain sample sizes. Also look at the sample sizes at the strata level and mark any adjustments where needed. A “flat” file of adjusted sample sizes is edited that overlays the strata sample sizes output from CHROMY and the special MEP type sample allocations. PROG18 is then rerun and the sample sizes are analyzed to see if the adjustments had the desired effect. Each iteration leads to new adjustments to arrive at a final adjusted sample size data set. The sample size at the district level must equal

the domain cost, or the preferred sample allocation (which is fixed and unchanging for each district).

A flow chart of those programs is shown in Appendix II-2.

2. Sample Selection

PROG19 is the sample selection program for MEPs and COUs within the RPW panel offices. The RPW frame is input along with the sample allocations. Within each stratum a sample of size n is chosen by systematic sample selection. Let k = the frame size divided by the sample size. Then, beginning with a random start between 1 and k a sample of size n is chosen systematically. Before the selection, the frame is sorted by stratum, district, finance number, facility ZIP Code, and then headquarters code. The program outputs a sample selection file without the test Ids and test dates. These are added in PROG21. Also output is a file with all RPW frame units and their associated strata numbers.

PROG20 calculates and prints the average number of MEPs per finance number both for the previous and current Postal Quarters. Sample sizes at the finance number level changing more than 10% are flagged in the printout. This program also provides a printout of panel offices that were not sampled through the sample selection process in PROG19.

PROG21 assigns a testid and testdate to each sampled MEP. For a description of the procedure for attaching the dates and testids refer to R94-1 documentation, G-65, Section 14, page 03. See that reference for a detailed program description.

A flow chart of those programs is shown in Appendix II-2.

3. Joint Scheduling

Statistical Programs data collection staffing is relatively fixed within a district, with a certain number of full time and part-time data collectors. Because of this, the number of tests assigned per day is a critical concern. There are several major statistical systems testing going on at the same time; ODIS, RPW, TRACS, IOCS, etc. Selecting test dates for each system independently causes some days with substantially more tests than other days. To help alleviate this peak load problem, tests for ODIS, RPW, RCCS and CCCS are jointly scheduled, or 'smoothed'. Aside from maintaining the Sunday and holiday test schedule, the joint scheduling process re-assigns test dates to achieve a good temporal distribution.

Specifically, for a typical postal quarter with three accounting periods, any three consecutive records in the randomized list of dates will have one date from each accounting period, and any six consecutive records will have one date from each pay period. Districts have an option between smoothing tests to achieve approximately the same number of tests per ODIS area, the same number of tests per program (RPW, ODIS, etc...), or not be smoothed.

Sample files for the four systems are concatenated, and Sunday and holiday tests are excluded from joint scheduling. Within a district, tests are sorted by ODIS area (almost a one-to-one correspondence with 3-digit ZIP Codes), and statistical system, or visa-versa depending on the choice of the MSP. Dates from the randomized date file are sequentially assigned. For each system, the output sample file is the same format as the input sample files, and contains the same records, except for the smoothed testdate.

The program name, inputs, outputs, and file layouts are documented in USPS-LR-J-17/R2001-1.

4. Sample Partition

After jointly scheduling, the Statistical Programs Service Center executes a c-list which partitions the sample file for each base unit computer (typically, one per district). The sample files are then made available for each CODES site to download. The sample partitioning procedures, JCL, COBOL programs, etc... are documented in Section 3 of Library Reference USPS-LR-J-22.

V. Administering Tests in the Field

Within each district, the Manager of Statistical Programs (MSP) is responsible for downloading the sample file, assigning tests to data collectors, reviewing the data after tests are conducted, and ensuring timely transmission of the data to the mainframe computer system at San Mateo.

1. Downloading the Sample

The CODES base unit contains a function that allows each site to download its sample. Site personnel utilize functions in the COMMUNICATION portion of the base unit software to download the sample. Once the site's sample is downloaded to its base unit, site personnel can download individual sample Testid information to disks, which allow a test to be performed on the CODES laptop computer.

2. Administrative Functions on Base Unit Computers

Administrative tracking of tests is performed on the base unit computer(s) in each district. Administrative functions available on the base unit computer include the following: preparing program/data disk, so the tests can be conducted, RPW data transfer for uploading test data to the mainframe, RPW system management, and RPW data management. RPW data management involves the review of an RPW test data file, keying a test as zero volume, cancel, or reschedule, and editing the test header information.

The RPW base unit software used to perform these administrative functions is documented in Section 2 of RPW-CODES Computer System Documentation filed as Library Reference USPS-LR-J-22.

3. Collecting the Data

RPW tests are conducted by data collectors knowledgeable in determining the subclass or rate category at which a piece was mailed, based on the indicia, postage, markings and endorsements on the mail piece. Prior to the test, the data collector contacts the test facility to notify them of the test, and seeks their cooperation in isolating the appropriate test mail. The data collector arrives on the test date prior to the beginning test cutoff time. The sample unit is the MEP-day. MEPs are usually defined around a mail processing stream, and the data collector would apply either census, mailpiece-skip or container-skip subsampling to select all or a portion of the MEP-days mail volume.

When applying counted-skip subsampling procedures (mailpiece-skip or container-skip), the skip interval used is based on the expected number of containers and/or mailpieces, and the expected average number of pieces per container. Detailed procedures for subsampling are described in the Data Collection User's Guide for Revenue, Volume, and Performance Measurement Systems, USPS-LR-J-23.

Data are recorded on a laptop computer, using the Computerized On-Site Data Entry System (CODES) for RPW. As data are entered, the CODES software has several automated checks to prevent data collectors from entering improper or invalid information. RPW data collection procedures are detailed further in Chapter 3 of Handbook F-75, Data Collection User's Guide for Revenue, Volume, and Performance Measurement Systems, filed as Library Reference USPS-LR-J-23. The CODES-RPW software, used on laptop computers to record the data, is documented in Section 1 of Library Reference USPS-LR-J-22.

4. Transmitting the Data to San Mateo

Each week, the MSP transmits all tests from the base unit computer to the mainframe computer at the San Mateo ISSC. A backup copy of the data are made as they are transmitted, in case of transmission errors or loss of data on the mainframe computer system. This process is more fully documented in Section 3 of USPS-LR-J-22.

VI. Data Validation and Editing

1. Creating the Check-in File

Each Monday through Wednesday morning, CODES Support executes a pre-check-in program to verify that all CODES sites have transmitted data for all tests scheduled for the previous week. They submit the JCL procedure HSD8350T. Once satisfied that all data has been received, the check-in program to do various basic data validation edits is executed by submitting the JCL procedure HSD8500P.

The JCL procedures and COBOL programs used in the Check-in process is documented in Section 3 of USPS-LR-J-22.

2. Editing, Validating and Processing of Data on the Mainframe

In addition to edits performed as data are entered by the data collector into the CODES software, RPW uses a series of programs on the mainframe to validate and edit data, and do the remaining processing outlined below. A flow chart of those programs is shown in Appendix II-3. The programs, in the order in which they are run, are as follows:

1. CRTMNSRx-Creates a file of average volumes by strata within district which will be used by the processing to identify influential tests.

2. EDITRx-Performs editing beyond the simple editing performed by the Checkin process. The Checkin process checks for valid date values, dates with the processing period, numeric fields having numeric values, and performing administrative functions for tests, i.e. cancel or reschedule tests, etc. The EDITRx program does more extensive edits which can be based on combinations of values or specific classes or types of mail. Mail classes that have rates based on zones are checked for valid values for zone. The mail class is used to determine whether the rates are ounce or pound rated, the type of rounding, if any, is done and the weight limits that exist. Dimensional data is checked to determine balloon, oversize, machinability, weight and dimensional data to determine standard or non-standard. Using the criteria necessary, the rates are looked up, compared to the revenue entered by the data collector, and short paid and over paid revenue is calculated.

3. MAKEAPRx-Takes the weekly edited files from the EDITRx program for enough weeks (4 for all AP's except for the last AP in the Postal Quarter which has 5 weeks) to make an Accounting Period file. The weekly files are concatenated, sorted, and duplicates identified and eliminated.

4. MAKEPQRx-Takes three AP files from the MAKEAPRx program and concatenates them to make a Postal Quarter file. The same duplicate record identification and elimination process is utilized as in the program MAKEAPRx.

5. EXPPQRx-Expands the revenue, volume and weight by the first stage blowup (panel office), second stage blowup (MEP days divided by the number of samples), skip factor and number of delivery days in the Postal Quarter.

6. MCATRx-Totals mail category data to represent summary category data for National estimates.

7. D2SUMRx-Provides revenue, piece and weight data by frame unit.

8. DSKEYTRx-Provides a distribution key for registered and certified special service data.

9. SPYRx-Reports RPW revenue, piece and weight data by summary category code compared with the same period last year.

10. OSPYRx-Reports ODIS revenue, piece and weight data by summary category code compared with the same period last year.

11. ODISPPRx-Produces a file of ODIS volumes by MEP of all fourth class, parcel post, and parcel shape mail.

12. OTLxRx-Produces a report and file of influential tests by summary mail category code.

13. OTLIRx-Produces a report and file of influential tests for First Class single piece mail by letter, flat and irregular parcel shapes.

14. PRINTRx-Reports First Class single piece reference volumes by shape, MEP and test.

15. LETFLTRx-Reports letter and flat test volumes greater than a specified percentage of the reference volumes.

16. BLDADJRx-Builds a file of adjustments to be applied to tests based on a selection from a file of influential tests.

17. SPWGTRx-Reports volumes by weight step and zone within selected Mail Categories.

3. Influential Test Detection and Adjustment

The RPW process detects influential tests, and adjusts the sampling weights through a simple weight reduction technique. Influential tests may individually contribute a large percentage to the national estimated revenue, pieces, and/or weight, typically through unusual volume test days with large sampling weights. The detection methods include 1) looking at a test's contribution to the national estimate, and 2) distributionally comparing it to other tests in its stratum. In addition, estimates are charted in a time series analysis, often utilizing independent survey data (i.e., ODIS volumes). Where inconsistent time trends are apparent, additional influential tests may be identified. A simple weight reduction technique is employed. A discussion of influential data point

detection and treatment can be found in Chapter 26, Business Survey Methods, Editors Cox, Binder, et al, Wiley Series, 1995.

Clist options 8, 9, and 10 are executed that produce an initial comparison of current to previous (SPLY) estimates, and potential influential test reports. These reports show test level distributions by category for revenue, pieces, weight, weight per piece and revenue per piece. Influential tests are identified using c-list option 11. C-list option 12 creates an adjustment factor by rate category, with values between 0 and 1. The adjustment factor is:

$$a_{hijkl} = \begin{cases} \frac{m_j * d_j}{M_j} & \text{if influential} \\ 1 & \text{otherwise} \end{cases}$$

VII. Estimation

Revenue, pieces and weight are collected at the four-digit mail category level, weighted, and then aggregated to the summary RPW report level. File HIS.RPW.MCODES.FYyy.DATA provides a list of four-digit mail categories and their mapping to three-digit summary codes (Appendix 1, Item 23). Raw data are first weighted by the container-skip and mailpiece-skip interval values, then by the second stage MEP weights and finally by the first-stage finance office expansion factors. The adjustment factor is applied in the expansion. Simple expansion estimates are ratioed to estimated residual trial balance revenue and multiplied by residual trial balance revenue. Sampling and estimation is conducted quarterly, and the estimates and variances summed across the four postal quarters.

Notation:

h	CAG stratum ; h=1 (CAG A-C), h=2 (CAG D), , h=9 (CAG L)
i	finance number (post office)
j	MEP stratum
k	MEP
l	mailpiece
N _h	number of finance numbers in CAG stratum h
n _h	number of sampled finance numbers in CAG stratum h
M _j	number of MEPs in stratum j
m _j	number of usable MEPs (sampled minus delinquents and administratively canceled) in stratum j
D _j	number of delivery days in postal quarter in stratum j
y _{hijk}	sample revenue, pieces or weight for the rate category of interest (zero otherwise) in CAG stratum h, post office i, MEP stratum j, MEP k

b_{hijk}	sample total revenue in CAG stratum h, post office l, MEP stratum j, MEP k
a_{hijkl}	adjustment factor (values between 0 and 1) for a rate category for mailpiece l
B	residual trial balance revenue

1. Expanding to the Test Level

Container-skip and mailpiece-skip interval values are stored in the data record for each mailpiece. For census sampling, the container-skip and mailpiece-skip interval values equal '1'. For mailpiece-skip interval subsampling, the container-skip interval value is '1' and the mailpiece-skip interval values greater than '1'. For container-skip subsampling, the container-skip and mailpiece-skip interval values are greater than '1'.

The estimated test-day volume for a mail category is computed as:

$$y'_{hijk} = \sum_l \text{container skip} * \text{mailpiece skip} * y_{hijkl} \quad (1)$$

Total revenue is also calculated from the test data. It is the sum of all revenue corresponding conceptually to that mail which makes up the residual trial balance. The estimated total revenue for a test-day is computed as:

$$b'_{hijk} = \sum_l \text{container skip} * \text{mailpiece skip} * b_{hijkl} \quad (2)$$

The estimated residual trial balance revenue is described in USPS-LR-J-21.

2. RPW Report Ratio Estimator

Test level revenue, pieces and weight for the RPW Report category obtained from equations (1) and (2) in the previous section are multiplied by an adjustment factor (values between 0 and 1), expanded within second-stage MEP strata, and summed across strata within first-stage finance numbers. Estimates of first-stage finance numbers are expanded within first-stage CAG strata, and summed across CAGs. These are simple expansion estimates for revenue, pieces, and weight for the rate category and for total revenue in equation (4) and (5), respectively. The ratio of these two estimates is applied to the residual trial balance revenue (B) to construct the RPW Report ratio estimator shown in (3). A description of the derivation of the residual trial balance revenue is described in Revenue, Pieces, and Weight Adjustment System (ARPW) Documentation and Source Code, USPS-LR-J-21. Equation (3) is this book revenue adjusted estimate for a RPW Report category for revenue, pieces or weight for a postal quarter. For fiscal year estimates, quarterly estimates are summed across the government fiscal quarters.

At the first-stage of selection, the method of estimation assumes that the sample of offices within CAGs D through L constitutes an equal probability sample. The estimation methodology also assumes that non-response (i.e., delinquent and cancelations for units no longer existing) is random, or independent of what is being estimated, through simple reduction in sample size.

The book revenue adjusted estimate for a particular rate category for a postal quarter is:

$$y''' = B * \frac{y''}{b''} \quad (3)$$

where,

$$y'' = \sum_h \frac{N_h}{n_h} \sum_i \sum_j \frac{M_{hij} * d_i}{m_{hij}} \sum_k y'_{hijk} \quad (4)$$

$$b'' = \sum_h \frac{N_h}{n_h} \sum_i \sum_j \frac{M_{hij} * d_i}{m_{hij}} \sum_k b'_{hijk} \quad (5)$$

The estimate for the fiscal year is the sum of y''' over the four postal quarters.

3. Variance and Confidence Limits

The variance of the book revenue adjusted estimate in (3) for a rate category and postal quarter is the weighted average of the variances of CAG A-C (class I) and CAG H-L (class II).³ A weighted average is appropriate since this CAG stratum is a certainty stratum ($n_h/N_h=1$). The rel-variance of y''' ⁴ is:

$$v^2_{(y''')} = \frac{y_I''^2 v^2_{Iy''} + y_{II}''^2 v^2_{IIy''}}{y''^2} + \frac{b_I''^2 v^2_{Ib''} + b_{II}''^2 v^2_{IIb''}}{b''^2} - 2 \frac{y_I'' b_I'' v_{Iy''b''} + y_{II}'' b_{II}'' v_{IIy''b''}}{y'' b''} \quad (6)$$

where, $v^2_{Iy''}$, $v^2_{IIy''}$, and $v_{Iy''b''}$ are the appropriate terms in the estimate of the relative variance for class I (CAGs A-C), $v^2_{IIy''}$, $v^2_{IIb''}$, and $v_{IIy''b''}$ are the appropriate terms in the estimate of the rel-variance for class II (CAGs D-L).

³ See Hansen, Morris H, William Hurwitz and William Madow, Sample Survey Methods and Theory (1953), Vol. 1, Chapter 9, Eqn 29.1, John Wiley & Sons

⁴ Rel-variance is the square of the coefficient of variation.

Also, $y'' = y''_I + y''_{II}$ and $b'' = b''_I + b''_{II}$. Equation (6) is a conservative and unbiased estimate of the rel-variance, often referred to as the ultimate cluster variance estimator (see Hansen, Hurwitz & Madow, Chapter 9, Equation 27.1).

The variance of y''' is then,

$$\text{var}(y''') = v^2_{y''} * y''' \quad (7)$$

The variance for the fiscal year is the sum of the variances from the postal quarters.

The coefficient of variation (CV) is then $v_{y''}$. For normally distributed random variables, the upper and lower 95% confidence limits of a fiscal year estimate (x) are:

$$x(1 \pm 1.96 \times CV(x)) \quad (8)$$

Variance computation begins with program RCODE that defines additional summary category codes. Programs VR00Q0, VR00Q2, VR00Q3, and VR00Q5 are then run that generate standard errors by postal quarter using equation (3) above. Category codes input records that are estimated from the PERMIT system through BRPW, or other accounting systems are set to zero. Program VR0PQSUM is executed that combines the data across the four postal quarters. The last job in the sequence, program VR0COMP, merges DRPW, BRPW, and other source data to compute final RPW coefficients of variation. Appendix IV lists references for source code and data.

Appendix I: File Formats

1. Postal Accounting File with CAG

HSF.ADGNYN.HQ070D01.FYyy

POSITIONS	LENGTH	FORM	DESCRIPTION
1-2	2	NUM	REGION/DIVISION
3-8	6	EBC	FINANCE NUMBER
9-9	1	EBC	CHECK DIGIT
10-12	3	EBC	SUBLOCATION
13-14	2	EBC	WORK CENTER
15-15	1	EBC	NEW CAG
16-43	28	EBC	POST OFFICE NAME
44-45	2	EBC	STATE
46-50	5	EBC	ZIP CODE
51-62	12	EBC	REVENUE
119-119	1	EBC	OLD CAG

2. District, Site and Service Center Mapping File

ZIPTABLE HSFRAN.PS001D01.ZIPTABLE.ORFEO

POSITIONS	LENGTH	FORM	DESCRIPTION
1-3	3	NUM	3-DIGIT ZIP CODE
5-25	20	EBC	CSD NAME
26-27	2	EBC	STATE
29-31	3	NUM	CSD CODE
35-37	3	NUM	SITE CODE
72-72	1	NUM	SPSC CODE
73-79	7	EBC	SPSC NAME

3. Finance Number to District Mapping File

HSDHQN.PS001D01.ODIS.FINFILE

POSITIONS	LENGTH	FORM	DESCRIPTION
1-6	6	NUM	FINANCE NUMBER
7-9	3	NUM	SITE CODE
11-13	3	NUM	CSD CODE
21-40	20	EBC	CSD NAME

4. Postal Accounting File with CAG

HSFRANT.RPW.DOMAIN.PqqFYyy

POSITIONS	LENGTH	FORM	DESCRIPTION
1-10	10	NUM	DOMAIN
11-13	3	EBC	STRING - 'C= '
14-16	3	NUM	CSD CODE
31-40	10	NUM	MAXCOST

5. MEP and CAG K&L Frame

HSFRAV.VS765D01.MASTER, HSFRANT.PS765T01.MASTER.Fyyyyq,
HSFRANT.MASTER.Fyyyyq, HSF.MEP.CAG.KL.FYyyyy

POSITIONS	LENGTH	FORM	DESCRIPTION
1-3	3	NUM	SITE CODE
4-9	6	NUM	FINANCE NUMBER
10-14	5	NUM	MAIN OFFICE ZIP CODE
15-19	5	NUM	FACILITY ZIP CODE
20-20	1	EBC	MEP TYPE
21-21	1	EBC	MEP GROUP
22-25	4	EBC	HQ CODE
26-55	30	EBC	FACILITY NAME
56-56	1	NUM	LAST ACTIVITY CODE
57-64	8	NUM	CREATE DATE
65-72	8	NUM	UPDATE USER
73-80	8	NUM	UPDATE DATE
81-86	6	NUM	UPDATE TIME
87-89	3	NUM	ORIGINATING SITE
90-93	4	EBC	ORIGINAL HQ CODE
94-96	4	EBC	NEW SITE CODE
97-97	1	EBC	TRANSACTION CODE
98-99	2	EBC	REGION/AREA
100-100	1	EBC	CAG
101-130	30	EBC	FACILITY-STREET
131-160	30	EBC	FACILITY-CITY
161-162	2	EBC	FACILITY-STATE
163-165	3	NUM	UNIVERSE CODE
166-166	1	FILLER	
167-246	80	EBC	MEP DESCRIPTION
247-251	5	NUM	MEP ZIP CODE
252-258	7	NUM	TOTAL PIECES
259-265	7	NUM	LETTERS AND CARD VOLUME
266-272	7	NUM	FLAT VOLUME
273-279	7	NUM	IPP VOLUME
280-286	7	NUM	PARCEL VOLUME
287-293	7	FILLER	
294-300	7	NUM	PRIORITY VOLUME
301-307	7	FILLER	
308-314	7	NUM	ACCOUNTABLE MAIL VOLUME
315-317	3	NUM	ODIS TRAVEL TIME
318-320	3	NUM	RPW TRAVEL TIME
321-323	3	FILLER	
324-326	3	NUM	ODIS TEST TIME
327-329	3	NUM	RPW TEST TIME
330-335	6	FILLER	
336-338	3	NUM	SAMPLE GROUP CODE
339-339	1	NUM	MEP-CSD-CHNG-IND
340-340	1	NUM	MEP-OTH-CHNG-IND
341-341	1	NUM	MEM-CHNG-DNL-IND
342-344	3	NUM	ODIS AREA
345-374	30	EBC	MAIN OFFICE NAME
375-377	3	NUM	CSD CODE
378-400	23	FILLER	

6. Estimated Historic ODIS Proportions to Create Stratification Variables

HSFRANT.RPW.STRAT.PARAMETR.FY952

<u>POSITIONS</u>	<u>LENGTH</u>	<u>FORM</u>	<u>DESCRIPTION</u>
1-3	3	NUM	CSD CODE
7-12	6 (4 DECIMALS)	NUM	PROPORTION OF LETTERS THAT ARE FIRST CLASS
13-18	6 (4 DECIMALS)	NUM	PROPORTION OF FLATS THAT ARE FIRST CLASS
19-24	6 (4 DECIMALS)	NUM	PROPORTION OF IPPS THAT ARE FIRST CLASS
25-30	6 (4 DECIMALS)	NUM	PROPORTION OF PARCELS THAT ARE FIRST CLASS
127-132	6 (4 DECIMALS)	NUM	PROPORTION OF LETTERS THAT ARE PARCEL POST
133-138	6 (4 DECIMALS)	NUM	PROPORTION OF FLATS THAT ARE PARCEL POST
139-144	6 (4 DECIMALS)	NUM	PROPORTION OF IPPS THAT ARE PARCEL POST
145-150	6 (4 DECIMALS)	NUM	PROPORTION OF PARCELS THAT ARE PARCEL POST
151-156	6 (4 DECIMALS)	NUM	PROPORTION OF LETTERS THAT ARE OTHER FOURTH
157-162	6 (4 DECIMALS)	NUM	PROPORTION OF FLATS THAT ARE OTHER FOURTH
163-168	6 (4 DECIMALS)	NUM	PROPORTION OF IPPS THAT ARE OTHER FOURTH
169-175	6 (4 DECIMALS)	NUM	PROPORTION OF PARCELS THAT ARE OTHER FOURTH

7. Sample File

HSFRANT.HQ310D01.PqqqFYyy

<u>POSITIONS</u>	<u>LENGTH</u>	<u>FORM</u>	<u>DESCRIPTION</u>
1-3	3	NUM	SITE CODE
4-9	6	NUM	FINANCE NUMBER
10-14	5	NUM	MAIN OFFICE ZIP CODE
15-19	5	NUM	FACILITY ZIP CODE
20-20	1	EBC	MEP TYPE
21-21	1	EBC	MEP GROUP
22-25	4	EBC	HQ CODE
26-28	3	EBC	SAMPLE DAY
29-30	2	EBC	REGION/AREA
31-33	3	FILLER	
34-39	6	NUM	TEST ID
40-47	8	NUM	TEST DATE
48-50	3	NUM	STRATA CODE
51-80	30	EBC	FACILITY NAME
81-81	1	EBC	CAG
82-161	80	EBC	MEP DESCRIPTION
162-166	5	NUM	MEP ZIP CODE
167-172	6	FILLER	
173-202	30	EBC	MAIN OFFICE NAME
203-205	3	NUM	CSD CODE
206-250	45	FILLER	

8. RPW Panel Offices

HSF.HQMAND07.SMPOFFyyPQq

POSITIONS	LENGTH	FORM	DESCRIPTION
1-6	6	NUM	FINANCE NUMBER
7-8	2	ALPHA	CLASS
9-36	28	ALPHA	POST OFFICE NAME
37-41	5	NUM	POST OFFICE ZIP CODE
42-42	1	ALPHA	CAG
43-44	2	ALPHA	STATE

9. Labels for Coefficient of Variation

HSFRANT.RP6.CVS1.PQ2FY95

POSITIONS	LENGTH	FORM	DESCRIPTION
12-30	19	EBC	LABEL
14-16	3	NUM	DISTRICT
21-22	2	NUM	SHAPE

10. Historic Coefficients of Variation

HSFRANT.RPW.CVS.NEWSTRAT.FY963

POSITIONS	LENGTH	FORM	DESCRIPTION
12-30	19	EBC	LABEL
14-16	3	NUM	DISTRICT
21-22	2	NUM	SHAPE
31-40	10 (5 DECIMALS)	NUM	COEFFICIENT OF VARIATION
51-70	20	NUM	ESTIMATE

11. CAG K&L Offices Out-of-Operation

HSF.MEP.CAG.KL.DELETE.Fyyyyy

POSITIONS	LENGTH	FORM	DESCRIPTION
1-6	6	NUM	FINANCE NUMBER

12. Quarterly Test Date File

HSF.HQ000D01.CARDATA(HSF020SA)

POSITIONS	LENGTH	FORM	DESCRIPTION
1-3	3	NUM	SEQUENCE NUMBER
4-4	1	NUM	SUNDAY/HOLIDAY FLAG
5-7	3	ALPHA	MONTH
8-9	2	NUM	DAY
10-11	2	NUM	YEAR
12-14	3	ALPHA	DAY OF WEEK

13. RPW Reference Volumes by MEP

HSFRANT.RPW.STRATIFY.ALL.PQxFYxx

POSITIONS	LENGTH	FORM	DESCRIPTION
1-25	25	ALPHA	MAIL EXIT POINT KEY
48-50	3	NUM	STRATA
51-80	30	ALPHA	FACILITY NAME
81-81	1	ALPHA	COST ASCERTAINMENT GROUP
82-161	80	ALPHA	MAIL EXIT POINT DESCRIPTION
167-173	7	NUM	MEP DBMS TOTAL VOLUME
174-180	7	NUM	MEP DBMS LETTER VOLUME
181-187	7	NUM	MEP DBMS FLAT VOLUME
188-194	7	NUM	MEP DBMS IPP VOLUME
195-201	7	NUM	MEP DBMS PARCEL VOLUME
202-208	7	NUM	MEP DBMS PRIORITY VOLUME
209-215	7	NUM	MEP DBMS ACCOUNTABLE MAIL VOLUME
258-260	3	NUM	CUSTOMER SERVICE DISTRICT

14. Special MEP Frame for Stratification

HSFRANT.RPW.SPECIAL.FRAME.PRED.FYxxx

POSITIONS	LENGTH	FORM	DESCRIPTION
N/A	N/A	NUM	PREDICTED FIRST CLASS VOLUME
N/A	N/A	NUM	PREDICTED PRIORITY VOLUME
N/A	N/A	NUM	PREDICTED PARCEL POST VOLUME
N/A	N/A	NUM	PREDICTED OTHER FOURTH CLASS VOL
N/A	N/A	NUM	PREDICTED ACCOUNTABLE MAIL VOL
N/A	3	ALPHA	SITE
N/A	6	ALPHA	FINANCE NUMBER
N/A	5	ALPHA	MAIN OFFICE ZIP CODE
N/A	5	ALPHA	FACILITY ZIP CODE
N/A	1	ALPHA	MAIL EXIT POINT TYPE
N/A	1	ALPHA	MAIL EXIT POINT GROUP
N/A	4	ALPHA	HEADQUARTERS CODE

15. Pre-Stratification Predicted Mail Volumes

HSFRANT.RPW.PRESTRAT.PQxFYxx.SAS.DATA

POSITIONS	LENGTH	FORM	DESCRIPTION
N/A	25	ALPHA	MAIL EXIT POINT KEY
N/A	N/A	NUM	PREDICTED FIRST CLASS VOLUME
N/A	N/A	NUM	PREDICTED PRIORITY VOLUME
N/A	N/A	NUM	PREDICTED PARCEL POST VOLUME
N/A	N/A	NUM	PREDICTED OTHER FOURTH VOLUME
N/A	N/A	NUM	PREDICTED ACCOUNTABLE MAIL VOL

16. Mail Category Rate File

HSISMN.RPW.EDITS.FYxxVxx.SASDATA

POSITIONS	LENGTH	FORM	DESCRIPTION
N/A	N/A	NUM	MAIL CATEGORY CODE
N/A	N/A	NUM	RATE LB
N/A	N/A	NUM	RATE OZ
N/A	N/A	NUM	RATE 1
N/A	N/A	NUM	RATE 2
N/A	N/A	NUM	ZONE
N/A	4	ALPHA	MAIL CATEGORY CODE
N/A	65	ALPHA	MAIL CATEGORY DESCRIPTION
N/A	1	ALPHA	BALLOON/OVERSIZE INDICATOR

17. Mail Category Master File

HSISMN.RPW.CATMAST.FYxxVxx.DATA(MASTER)

POSITIONS	LENGTH	FORM	DESCRIPTION
1-4	4	NUM	MAIL CATEGORY CODE
1-4	4	ALPHA	MAIL CATEGORY CODE
6-70	65	ALPHA	MAIL CATEGORY DESCRIPTION

18. Multiple Special Service Mapping File

HSISMN.RPW.CATMAST.FYxxVxx.DATA(MULTSS)

POSITIONS	LENGTH	FORM	DESCRIPTION
1-4	4	NUM	MAIL CATEGORY CODE
6-9	4	NUM	SPECIAL SERVICE CODE 1
11-14	4	NUM	SPECIAL SERVICE CODE 2
16-19	4	NUM	SPECIAL SERVICE CODE 3

19. RPW Checkin Output File

HSDHQN.PS840T02.CODESOUT.RPWxxxxx

POSITIONS	LENGTH	FORM	DESCRIPTION
1-6	6	NUM	FINANCE NUMBER
7-12	6	NUM	TEST IDENTIFICATION NUMBER
13-17	5	ALPHA	ZIP CODE
18-19	2	NUM	CENTURY
18-25	8	NUM	TEST DATE (YYYYMMDD)
26-27	2	NUM	VERSION
28-31	4	NUM	TIME
32-35	4	NUM	SKIP FACTOR
36-36	1	NUM	FIRST DIGIT OF CATEGORY CODE
36-39	4	NUM	MAIL CATEGORY CODE
36-39	4	ALPHA	MAIL CATEGORY CODE
40-43	4	NUM	VOLUME
44-52	9 (3 DECIMALS)	NUM	REVENUE
53-55	3	NUM	WEIGHT IN LBS
56-59	4 (1 DECIMAL)	NUM	WEIGHT IN OZS
60-62	3	NUM	AGENCY
63-66	4	NUM	SPECIAL SERVICE CATEGORY CODE
67-69	3	NUM	ZIP CODE
70-70	1	NUM	ZONE
71-75	5 (3 DECIMALS)	NUM	SHORT PAID
76-80	5 (3 DECIMALS)	NUM	OVER PAID
81-81	1	ALPHA	INDICIA
82-82	1	NUM	AUTOMATION COMPATABLE
83-83	1	NUM	SHAPE
84-86	3	NUM	LENGTH IN INCHES
87-88	2	NUM	WIDTH IN INCHES
89-90	2	NUM	HEIGHT IN INCHES
91-92	2	NUM	GIRTH IN INCHES
93-93	1	NUM	SAMPLING METHOD
94-95	2	NUM	SPECIAL SERVICE RECORD NUMBER
96-97	2	NUM	SESSION NUMBER
98-101	4	NUM	RECORD NUMBER
102-104	3	ALPHA	DATA COLLECTOR ID
105-105	1	ALPHA	LAPE FLAG
106-108	3	ALPHA	CUSTOMER SERVICE DISTRICT
109-110	2	ALPHA	ACCOUNTING PERIOD
111-118	8	NUM	RECORD LOCATOR

20. File of SAS Code to Correct Data Errors Called in by Service Centers

HSISMN.RPW.xxxxFIX.DATA

SAS code that is included in the program at run time to enable primary outlier fixes and other edits to be applied to a particular run without requiring the main program to require changes.

21. RPW Panel Office File

HSF.HQMAND07.SMPOFFxx.PQx

POSITIONS	LENGTH	FORM	DESCRIPTION
1-6	6	NUM	FINANCE NUMBER
42-42	1	NUM	COST ASCERTAINMENT GROUP

22. Sampled Vs Total Panel Offices by CAG File

HSISMN.RPW.HQ100D01.EXPAND1.FYxxPQx

POSITIONS	LENGTH	FORM	DESCRIPTION
1-1	1	ALPHA	CHK
5-5	1	ALPHA	COST ASCERTAINMENT GROUP
19-23	5	NUM	NUMBER OF PANEL OFFICES IN POPULATION
35-37	3	NUM	NUMBER OF SAMPLED PANEL OFFICES

23. Mail Category to Summary Category File

HSISMN.RPW.MCODES.FYxx.DATA

POSITIONS	LENGTH	FORM	DESCRIPTION
1-4	4	ALPHA	MAIL CATEGORY
6-55	50	ALPHA	MAIL CATEGORY DESCRIPTION
57-59	3	ALPHA	SUMMARY CATEGORY
61-110	50	ALPHA	SUMMARY CATEGORY DESCRIPTION

24. ODIS Quarterly Sample Data

HSA.HQ350T01.FYxxQTx

POSITIONS	LENGTH	FORM	DESCRIPTION
7-11	5	ALPHA	TEST IDENTIFICATION NUMBER
23-23	1	ALPHA	MAIL SHAPE
37-37	1	ALPHA	MAIL MARKINGS
39-39	1	ALPHA	MAIL CLASS
67-70	4	NUM	MAIL VOLUME

25. ODIS Sample File

HSF.HQ134D01.FYxxx

POSITIONS	LENGTH	FORM	DESCRIPTION
1-25	25	ALPHA	MAIL EXIT POINT KEY
34-38	5	ALPHA	TEST IDENTIFICATION NUMBER
48-50	3	ALPHA	STRATA CODE
162-164	3	ALPHA	DESTINATION ZIP
167-169	3	ALPHA	PROCESSING DISTRIBUTION CENTER

26. RPW Sample File

HSF.HQ330T02.PQxxFYxx

POSITIONS	LENGTH	FORM	DESCRIPTION
1-25	25	ALPHA	MAIL EXIT POINT KEY
4-9	6	NUM	FINANCE NUMBER
15-19	5	ALPHA	FACILITY ZIP CODE
34-39	6	NUM	TEST IDENTIFICATION NUMBER

27. ODIS Frame File for Reference Volumes

HSFRANT.ODIS.FRAME.FYxxxxxx

POSITIONS	LENGTH	FORM	DESCRIPTION
1-25	25	ALPHA	MAIL EXIT POINT KEY
15-19	5	ALPHA	FACILITY ZIP CODE
62-81	20	ALPHA	MEP DESCRIPTION (20 chars)
62-121	60	ALPHA	MEP DESCRIPTION (60 chars)
174-180	7	NUM	LETTER REFERENCE VOLUME
181-187	7	NUM	FLAT REFERENCE VOLUME

28. ODIS Letter and Flat Volumes

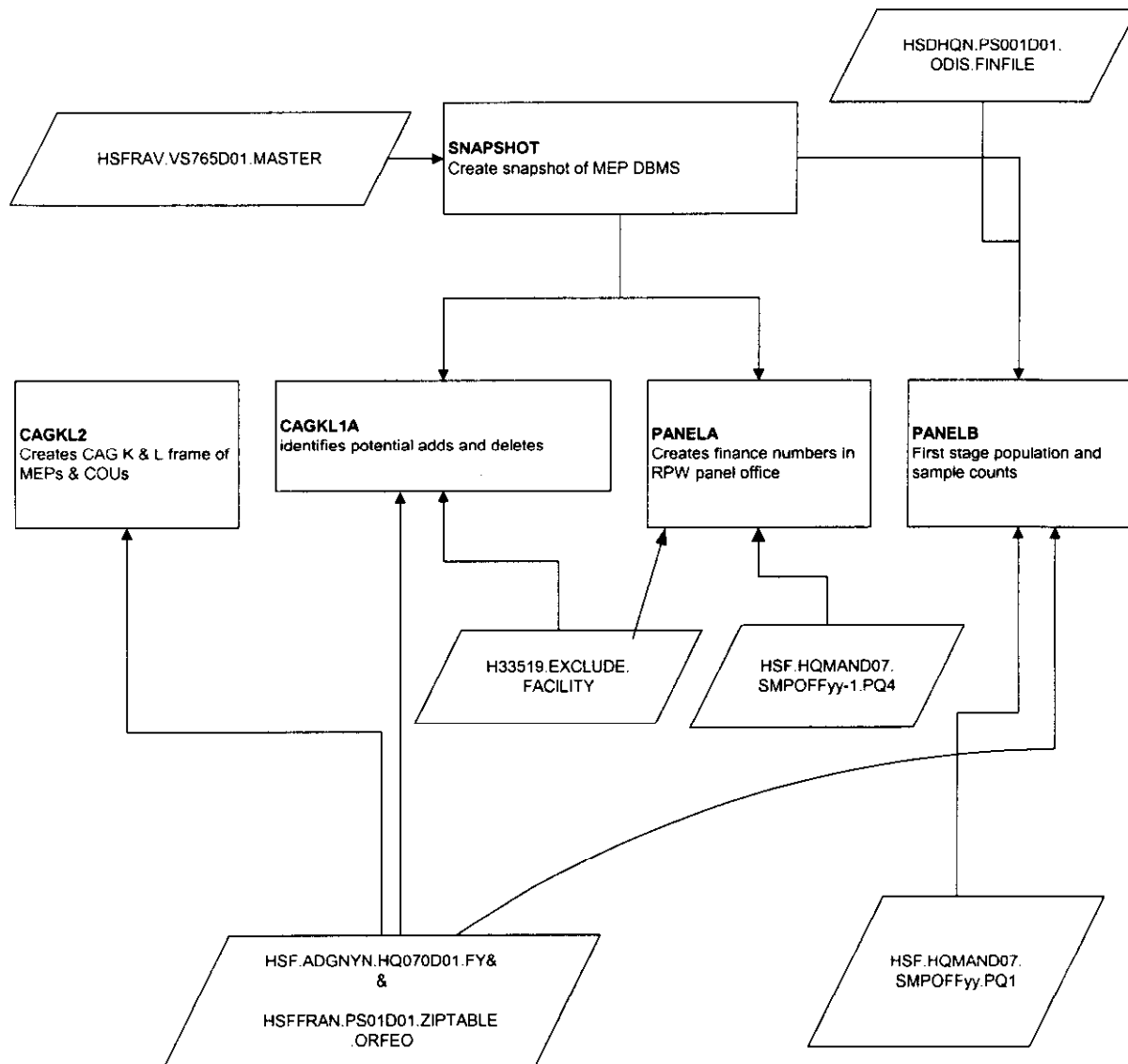
HSISMN.RPW.ODIS.LETFLT.FYxx

POSITIONS	LENGTH	FORM	DESCRIPTION
3-4	2	ALPHA	POSTAL FISCAL YEAR
7-8	2	ALPHA	ACCOUNTING PERIOD
10-34	25	ALPHA	MAIL EXIT POINT KEY
38-46	9	NUM	LETTER VOLUME
48-56	9	NUM	FLAT VOLUME
58-66	9	NUM	SINGLE PIECE LETTER VOLUME
68-76	9	NUM	SINGLE PIECE FLAT VOLUME

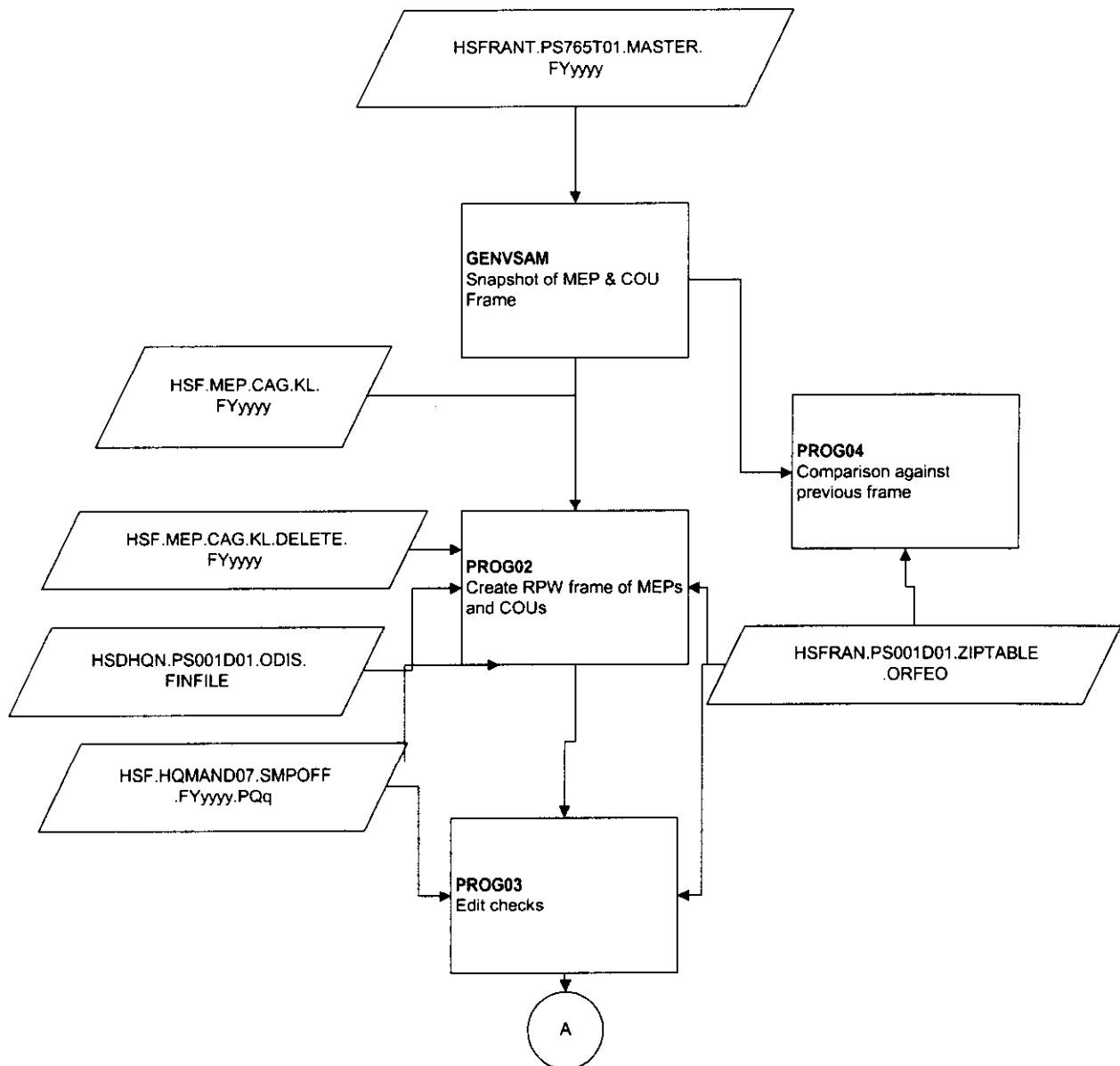
Appendix II: Flowcharts

1. Building the Sampling Frame Flowchart

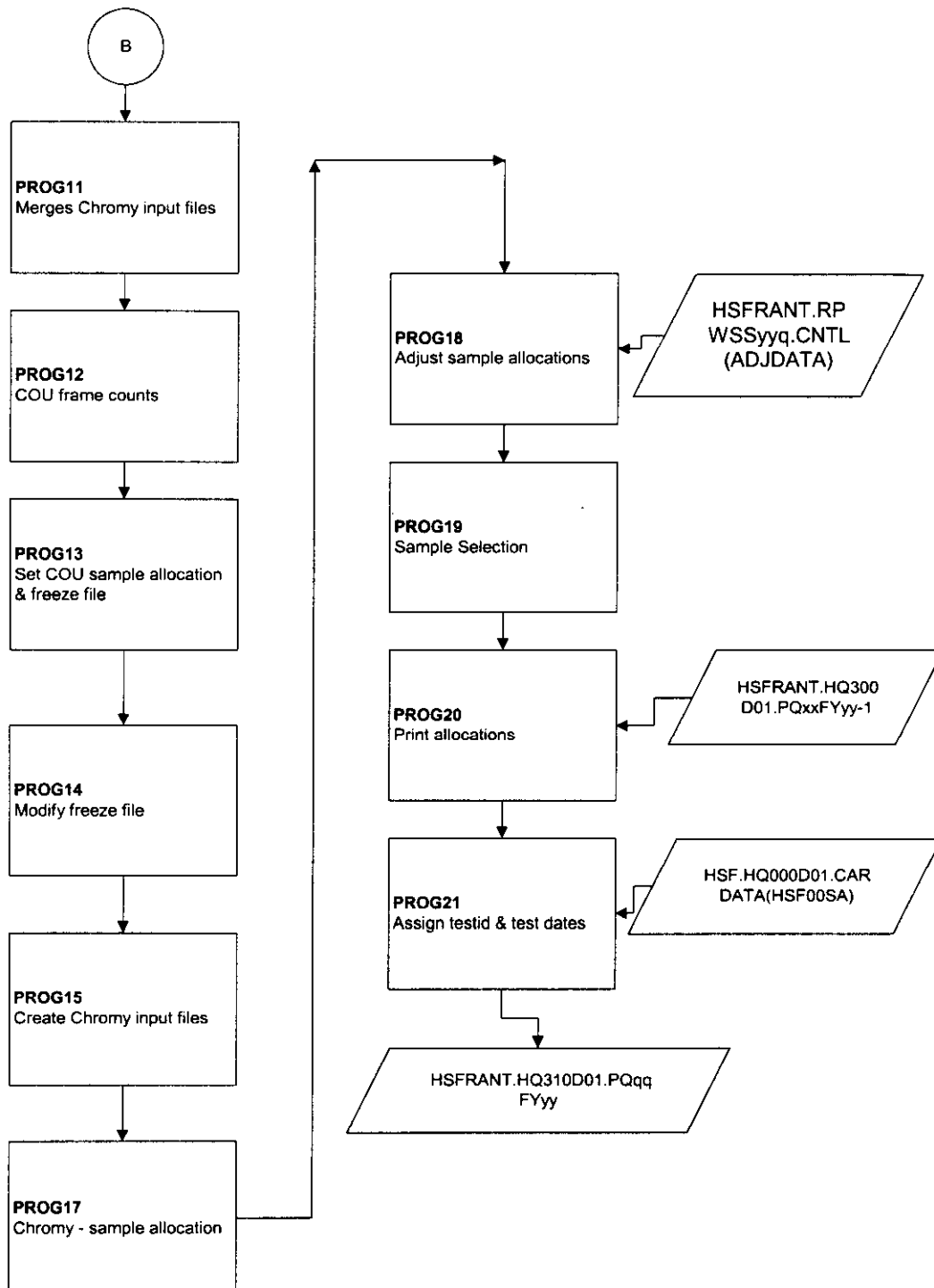
1a. First Stage Panel of Finance Numbers



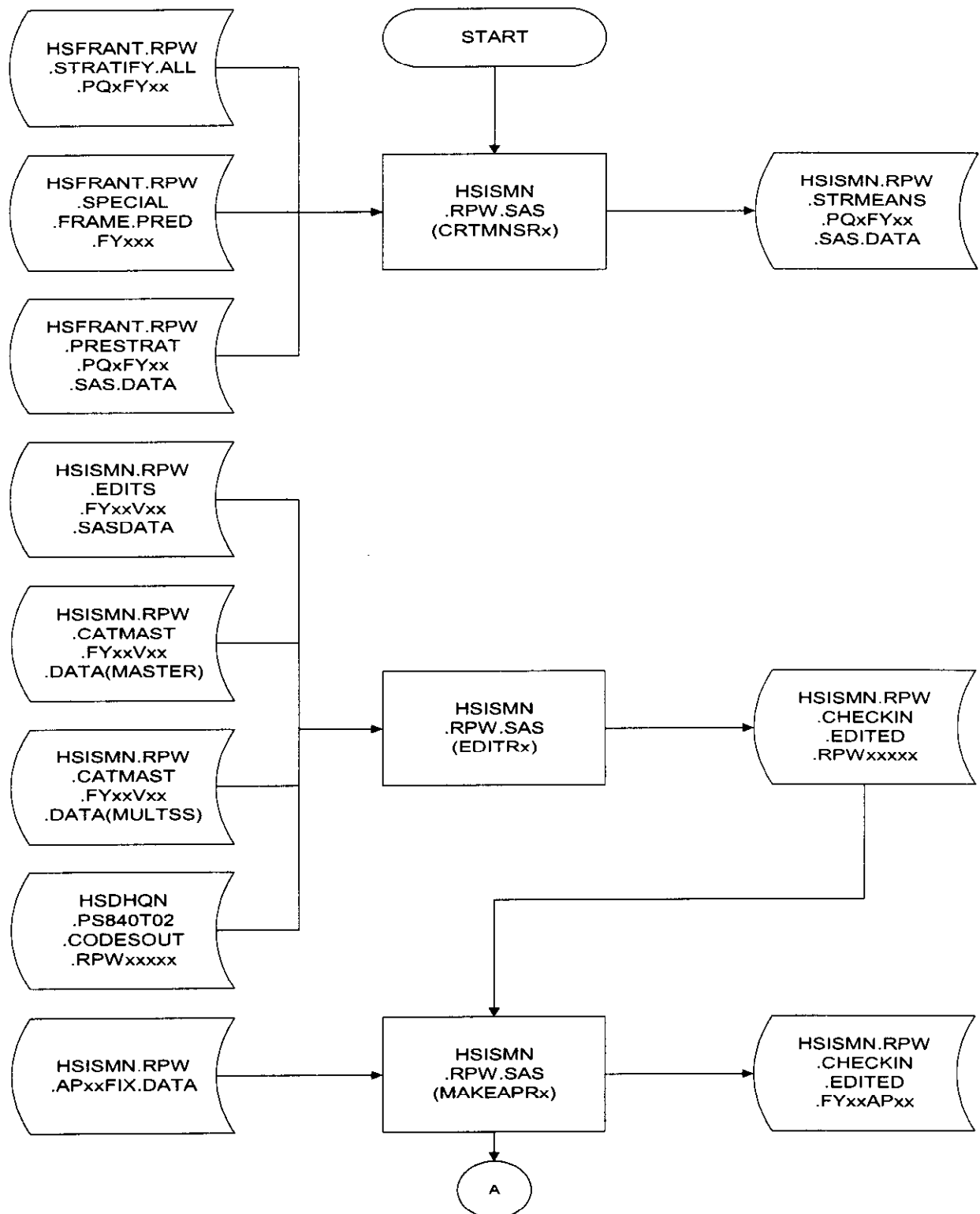
1b. Second Phase MEPs and COUs

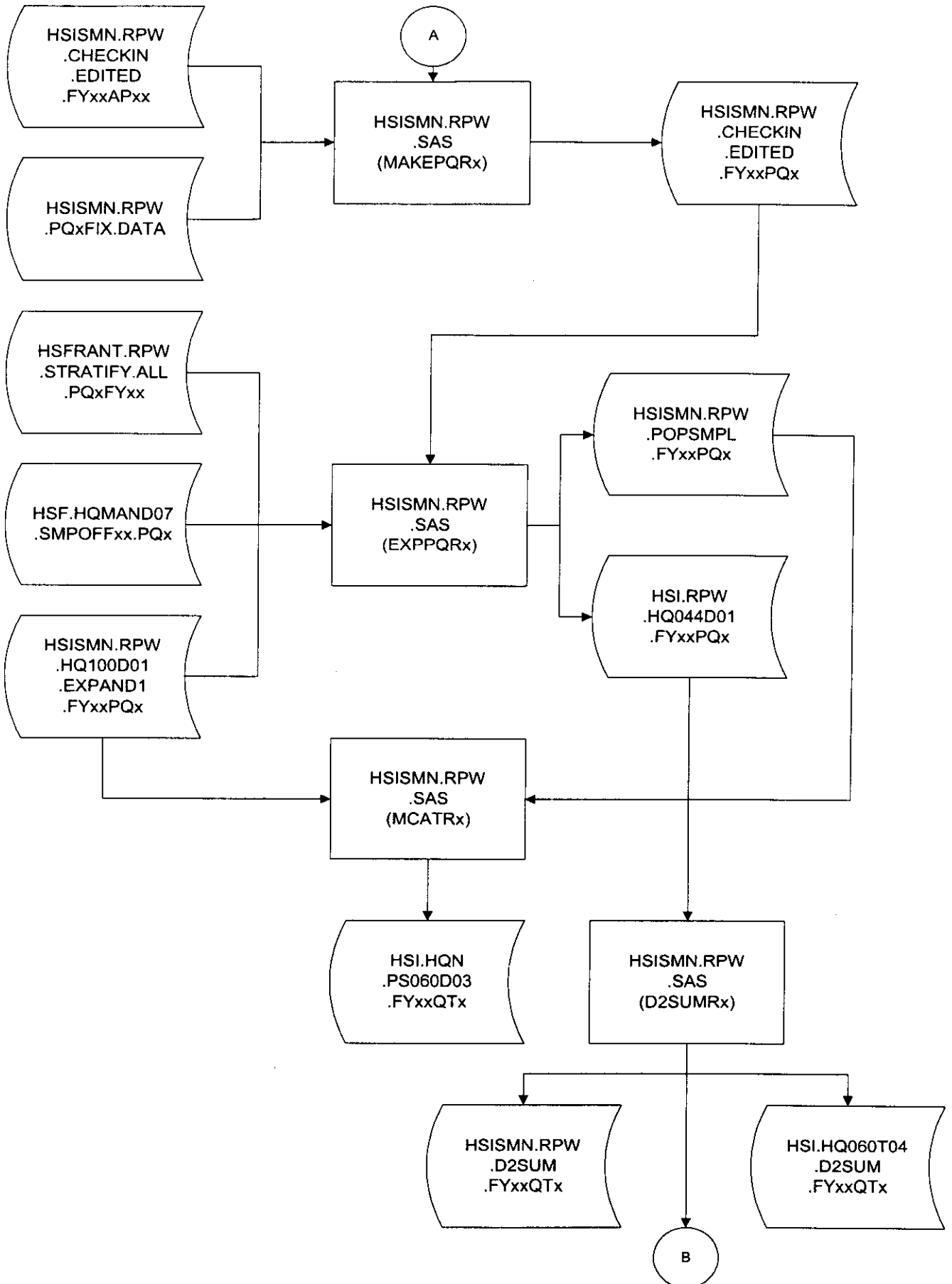


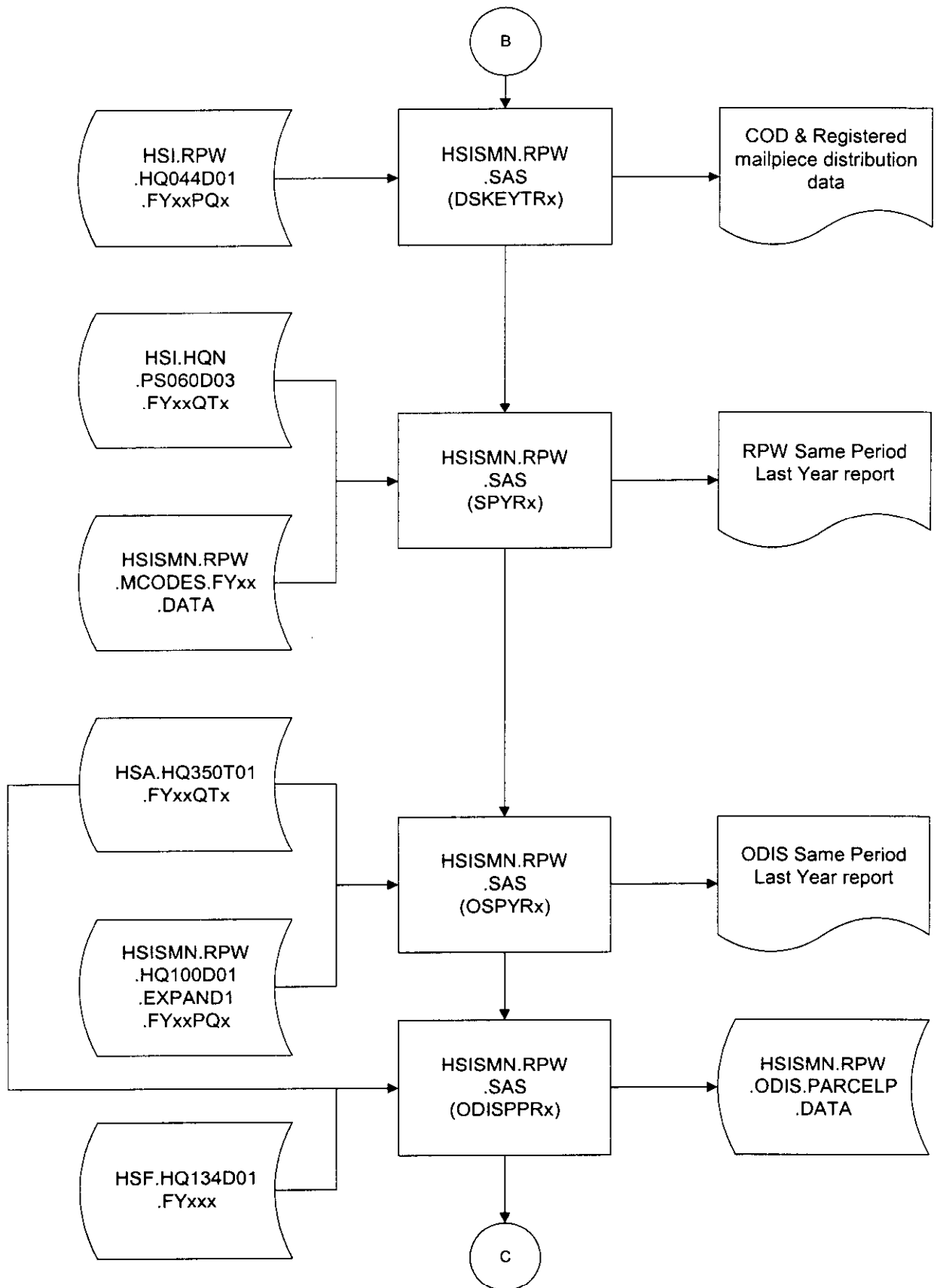
2. Sample Selection Flowchart

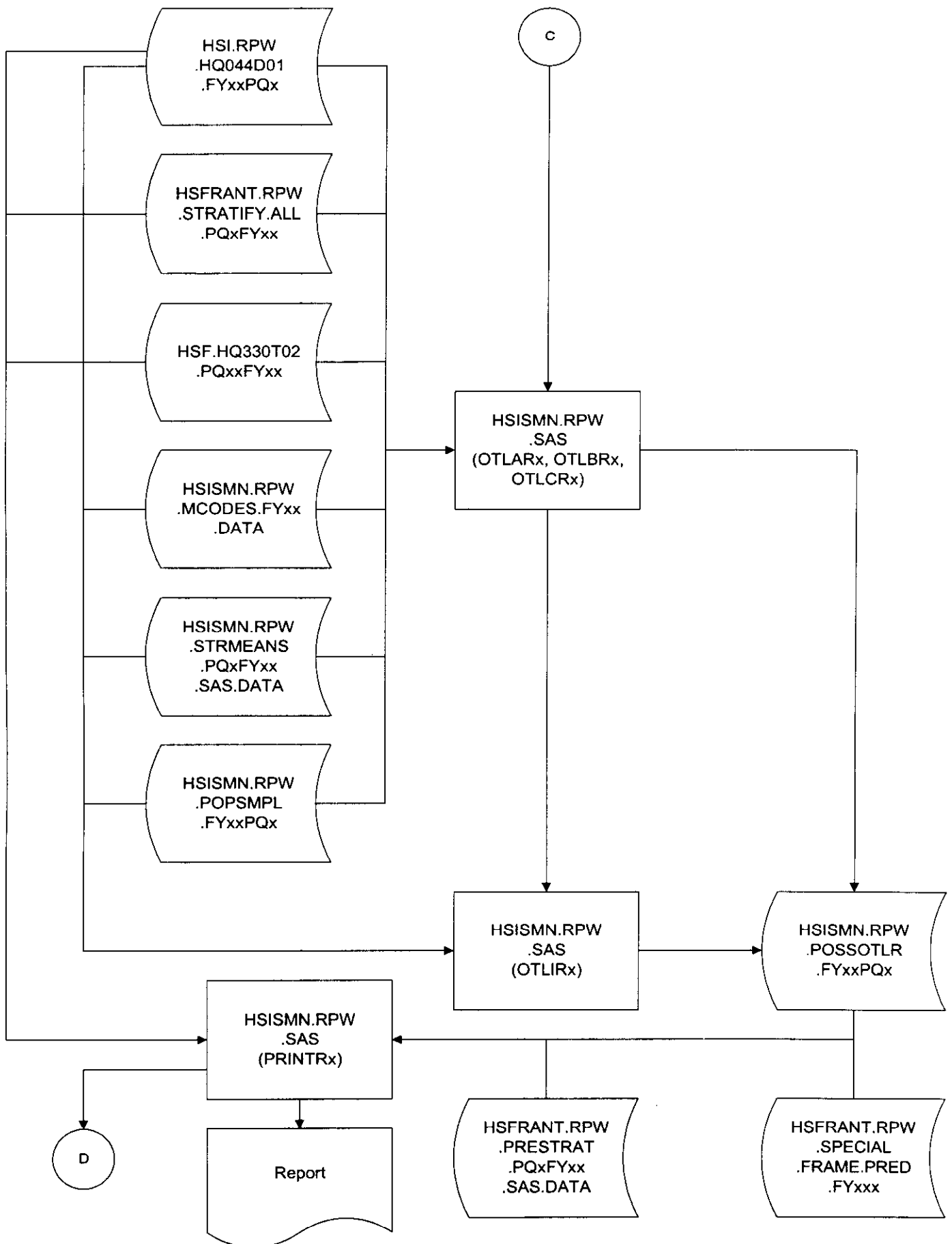


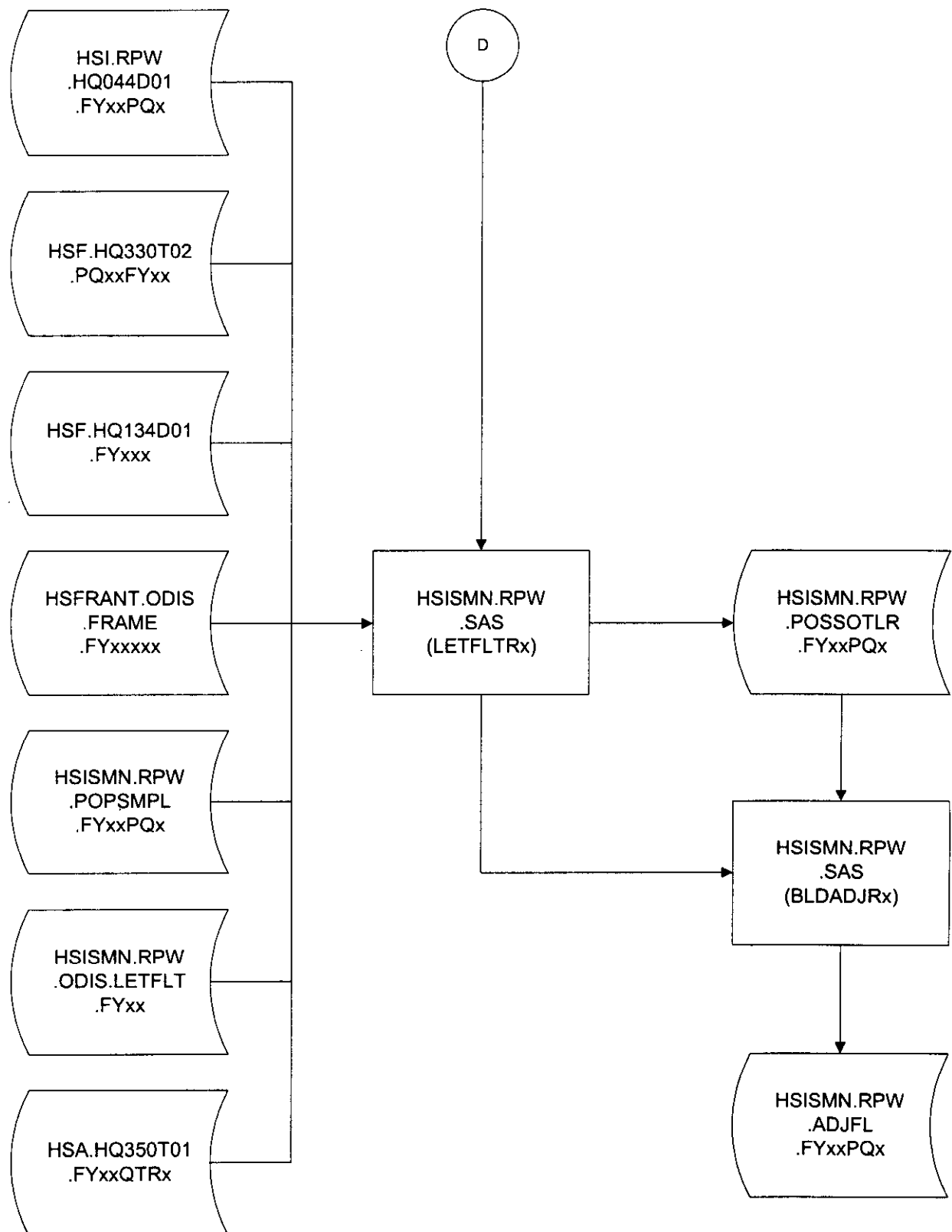
3. Domestic RPW Processing Flowchart











Appendix III: Tables

Table 1: First-Stage Counts of Finance Numbers – Population and Sample Size

	PQ 1 FY 2000		PQ 2 FY 2000		PQ 3 FY 2000		PQ 4 FY 2000	
CAG	Frame Size	Sample Size	Frame Size	Sample Size	Frame Size	Sample Size	Frame Size	Sample Size
A	109	109	109	109	109	109	109	109
B	190	190	190	190	190	190	190	190
C	730	730	730	731	731	731	731	731
D	641	50	641	50	641	50	641	50
E	1574	57	1574	57	1574	57	1574	57
F	2030	47	2030	47	2030	47	2030	47
G	3102	27	3101	26	3101	26	3101	26
H	3766	29	3766	29	3766	29	3766	29
J	4986	26	4986	26	4986	26	4986	26
K	8965	31	8965	31	8965	31	8965	31
L	1457	28	1457	28	1456	27	1456	27

Appendix IV: Lists of Source Code and Data on CD-ROM

<i>Description</i>	<i>Directory</i>	<i>File</i>
SAMPLE SELECTION PROGRAMS	DRPW\DESIGN\ADVREL	SNAPSHOT, CAGKL2, CAGKL1A, PANELA, PANELB,
	DRPW\DESIGN\PQ1	GENVSAM, PROG02, PROG03, PROG04, PROG06, PROG06AL, PROG03, PROG0, PROG06,
	DRPW\DESIGN\PQ2	PROG06AL, PROG07-PROG10, PROG10A, PROG10B, PROG11-
	DRPW\DESIGN\PQ3	PROG15, PROG17-PROG21
	DRPW\DESIGN\PQ4	
SECOND-STAGE UNIVERSE AND SAMPLE SIZES, AND STRATA DEFINITIONS	DRPW\DESIGN\STRDEF	STRCTPQ1, STRCTPQ2, STRCTPQ3, STRCTPQ4, STRATDEF
DATA EDITING, VALIDATION, and PROCESSING SAS PROGRAMS	DRPW\PROCESSING\SAS	ADJPQR1, ADJPQR2, ADJPQR3, ADJPQR4, BLDADJR2, CRTMNSR1, D2SUMR1, DSKEYTR2, EDITR8, EDITR9, EDITRA, EDITRB, EDITRC, EXPPQR2, EXPPQR3, LETFLTR4, LETFLTR5, MAKEAPR1, MAKEPQR1, MCATR2, ODISPPR1, OSPYR4, OSPYR5, OTLAR2, OTLBR3, OTLCR2, OTLIR3, PRINTR1, SPYR2, SPWGTR3
DATA EDITING, VALIDATION, and PROCESSING PQ 1 JCL	DRPW\PROCESSING\JCL\PQ1	ADJPQ, BLDADJPQ, CREATMNS, EDIT011, EDIT012, EDIT013, EDIT014, EDIT021, EDIT022, EDIT023, EDIT024, EDIT031, EDIT032, EDIT033, EDIT034, EDIT035, EXPNDPQ, LETFLTPQ, MAKEAP01, MAKEAP02, MAKEAP03, MAKEPQ, MCATPQ1, MCATPQ2, OTLRPQ1, SPLYPQ1, SPLYPQ2, SPWGT

DATA EDITING,
VALIDATION,
and PROCESSING
PQ 2 JCL

DRPW\PROCESSING\JCL\PQ2

ADJPQ, BLDADJPQ,
CREATMNS, EDIT041,
EDIT042, EDIT043, EDIT044,
EDIT051, EDIT052, EDIT053,
EDIT054, EDIT061, EDIT062,
EDIT063, EDIT064, EDIT065,
EXPNDPQ, LETFLTPQ,
MAKEAP04, MAKEAP05,
MAKEAP06, MAKEPQ,
MCATPQ1, MCATPQ2,
OTLRPQ1, SPLYPQ1, SPLYPQ2,
SPWGT

DATA EDITING,
VALIDATION,
and PROCESSING
PQ 3 JCL

DRPW\PROCESSING\JCL\PQ3

ADJPQ, BLDADJPQ,
CREATMNS, EDIT071,
EDIT072, EDIT073, EDIT074,
EDIT081, EDIT082, EDIT083,
EDIT084, EDIT091, EDIT092,
EDIT093, EDIT094, EDIT095,
EXPNDPQ, LETFLTPQ,
MAKEAP07, MAKEAP08,
MAKEAP09, MAKEPQ,
MCATPQ1, MCATPQ2,
OTLRPQ1, SPLYPQ1, SPLYPQ2,
SPWGT

DATA EDITING,
VALIDATION,
and PROCESSING
PQ 4 JCL

DRPW\PROCESSING\JCL\PQ4

ADJPQ, BLDADJPQ,
CREATMNS, EDIT101,
EDIT102, EDIT103, EDIT104,
EDIT111, EDIT112, EDIT113,
EDIT114, EDIT121, EDIT122,
EDIT123, EDIT124, EDIT131,
EDIT132, EDIT133, EDIT134,
EDIT135, EXPNDPQ1,
LETFLTPQ, MAKEAP10,
MAKEAP11, MAKEAP12,
MAKEAP13, MAKEPQ,
MCATPQ1, MCATPQ2,
OTLRPQ1, SPLYPQ1, SPLYPQ2,
SPWGT

DATA EDITING,
VALIDATION,
and PROCESSING
PQ 7 JCL

DRPW\PROCESSING\JCL\PQ7

INSURED, SPWGT

DATA EDITING, VALIDATION, and PROCESSING FILES	DRPW\PROCESSING\FILES	ADJFLQ1, ADJFLQ2, ADJFLQ3, ADJFLQ4, EXPNDQ1, EXPNDQ2, EXPNDQ3, EXPNDQ4, EDIT99V2, EDIT00V1, EDIT00V2, FACTORS, MCODE99, MSTR99V2, MSTR00V1, MSTR00V2, MULT99V2, MULT00V1, PGMEDITS
DATA EDITING, VALIDATION, and PROCESSING CLIST	DRPW\PROCESSING\CLIST	ADJ, BUILDADJ, CREATMNS, EDIT, EXPND, LETFLT, MAKEAPPQ, MCAT, MCAT2, POSSOTLR, RPW, RPWR6, SPLY, SPLY2, WEIGHT, CMDRVPM
VARIANCE PROGRAMS	DRPW/VARIANCE	RCODE, VR00Q0, VR00Q2, VR00Q3, VR00Q5, VR0PGSUM, VR0COMP
INPUTS *	HSFRAN/PS001D01/ZIPTABLE HSFRANT/RPW/STRAT/PARAMETR HSFRANT/RPW/DOMAIN HSFRANT/RPW/CVS/NEWSTRAT HSFRANT/RPW/CVS1	
SAS LOGS OF EXPANSION	DRPW\PROCESSING\LOGS	FY00PQ1, FY00PQ2, FY00PQ3, FY00PQ4

* Some input files contain facility specific information such as finance number, and are not included.