Postal Regulatory Commission
Submitted 3/11/2021 12:36:34 PM
Filing ID: 116270
Accepted 3/11/2021

PROPOSAL ONE: TECHNICAL APPENDIX

This appendix provides statistical documentation for statistical design changes and the replication study that was conducted with the Quarter 2, FY2020 ODIS-RPW sample data to measure the impact on the coefficients of variation (CV) on estimated revenue and pieces. It begins with a brief review of the ODIS-RPW sampling program.

ODIS-RPW Probability Sampling Program

The ODIS-RPW system is a probability-based destinating mail sampling system used to support the Postal Service's many and varied business needs for revenue and pieces information. ODIS-RPW primarily supplies official RPW estimates of revenue, piece, and weight for single-piece stamped and non-PC Postage metered mail and services.

The ODIS-RPW sampling and estimation is divided into two segments; non-digital and digital as described below.

ODIS-RPW Program – Non-digital:

ODIS-RPW data collectors travel to randomly selected Mail Exit Points (MEPs) on randomly selected days and randomly sample mail as it arrives at the delivery units. Container and mailpiece skip-sampling procedures are applied to the mail containers. Data collectors record mail characteristics from sampled mailpieces, including revenue, pieces, weight, mail class, subclass, extra services, and indicia type.

ODIS-RPW Program – Digital:

Letters and cards processed through Delivery Barcode Sorter (DBCS) machines are sampled digitally. Currently, approximately 89 percent of all Delivery Point Sequence (DPS) letters and cards, and approximately 66 percent of all letters and cards are sampled under the digital framework. Images for randomly selected MEP days and randomly selected mailpieces are chosen according to a sampling plan as they are processed by DBCS machines. These images are sent to ODIS-RPW. Exact daily volumes of mailpieces processed are available for all MEPs. Data collectors observe digital images of selected mailpieces and record mail characteristics, including revenue, mail class, subclass, extra services, and indicia type.

Digital Population

Proposed Changes: Two minor changes to the current digital sampling design are proposed. First, a stratum is added to the current five Business Delivery Stops (BDS) stratification. The addition is intended to circumvent the effects of a technical constraint that sets the minimum skip interval of the second-stage subsampling of digital images at fifty. We target a subsample of 175 digital images per test to control the within-test variation. Because 5,000 divided by fifty equals one hundred, it follows that for Mail Exit Points (MEPs) with average daily volume of less than 5,000 pieces, the number of digital images sampled are one hundred or less due to the minimum skip constraint of fifty. With day volume variation, we often see 50 images or less from those MEPs. Subsampling of 50 images doubles the within-test standard deviation that would have been obtained from 175 images. The additional stratum, by keeping small volume

MEPs with a higher sampling variation in their own stratum, prevents these MEPS from influencing other large volume MEPs in the expansion process.

Second, sample areas (geographical stratification) are consolidated from 189 to ten. The leading digit of the ZIP Code is used to form the ten geographical strata. A consolidation is necessary to avoid creating empty strata[1], and each stratum has many representations in a sample after the sample reduction. Geographical stratification does not provide much homogeneity in strata beyond what is provided by mail-characteristic stratification, but simulations showed that even a coarse level of geographical stratification reduces sampling variability, mainly because the influence of outliers can be kept within their respective strata.

Except for changes in definitions of sample area and BDS stratification, there is no change to the current ratio estimator.

Indices are defined as follows:

$$i = Business\ Delivery\ Stops, \quad (i = 1, \dots, 6)$$
$$j = Consolidated\ Sample\ Area\ j, \quad (j = 1, \dots, 10)$$
$$k = MEP\ day\ for\ stratum\ (i,j), \quad (k = 1, \dots, n_{ij}, \dots, N_{ij})$$
$$l = Mailpiece\ for\ MEP\ day\ k\ in\ stratum\ (i,j), (l = 1, \dots, n_{ijk}, \dots, N_{ijk})$$

Define

$$N_{ij} = \ the\ number\ of\ MEP\ days\ in\ strata\ (i,j)$$
$$n_{ij} = \ the\ number\ of\ MEP\ days\ sampled\ in\ strata\ (i,j)$$
$$x_{ijk} = the\ known\ volume\ in\ MEP\ day\ k\ in\ strata(i,j)$$
$$n_{ijk} = the\ number\ of\ mailpieces\ sampled\ in\ MEP\ day\ k\ in\ strata\ (i,j)$$
$$y_{ijkl} = \begin{cases} revenue, pieces\ or\ weight\ for\ mailpiece\ l\ for\ the\ product\ of\ interest, \\ 0, \quad otherwise \end{cases}$$

The digital estimator $\hat{t}_R$ adjusted for two minor changes is

[1] Empty strata are defined as statistical sampling strata with no usable tests.

$$\hat{t}_R = \sum_{i=1}^{6} \hat{t}_{R,i} \overset{\text{def}}{=} \sum_{i=1}^{6} \frac{X_i}{\hat{X}_i} \hat{t}_i$$

where

$$\hat{t}_i = \sum_{j=1}^{10} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} \frac{x_{ijk}}{n_{ijk}} \sum_{l=1}^{n_{ijk}} y_{ijkl}$$
$$\hat{X}_i = \sum_{j=1}^{10} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{hi}} x_{ijk}$$
$$X_i = \sum_{j=1}^{10} \sum_{k=1}^{N_{ij}} x_{ijk}$$

$\hat{t}_R$ is a ratio estimator, whose estimates are constructed from the inverse of the probabilities of selection and then adjusted by ratios of total volumes to estimated volumes. The first-order Taylor approximation variance of the ratio estimator is

$$V(\hat{t}_R) = \sum_{i=1}^{6} V(\hat{t}_{R,i})$$

with

$$V(\hat{t}_{R,i}) = \left(\frac{X_i}{\hat{X}_i}\right)^2 \left\{\sum_{j=1}^{10} N_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right) \frac{S_1^2}{n_{ij}} + \sum_{j=1}^{10} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{N_{ij}} x_{ijk}^2 \left(1 - \frac{n_{ijk}}{x_{ijk}}\right) \frac{S_2^2}{n_{ijk}}\right\}$$

where

$$S_1^2 = \frac{1}{N_{ij} - 1} \sum_{k=1}^{N_{hj}} \left(t_{ijk} - R_i x_{ijk}\right)^2$$
$$R_i = \frac{t_i}{X_i}$$
$$S_2^2 = \frac{1}{N_{ijk} - 1} \sum_{l=1}^{N_{hjk}} \left(y_{ijkl} - \bar{y}_{ijk}\right)^2$$
$$\bar{y}_{ijk} = \frac{1}{N_{ijk}} \sum_{l=1}^{N_{hjk}} y_{ijkl}$$

Non-Digital Population

<u>Proposed Sample Design:</u>

The current sample design uses a two-layer hierarchical stratification. First, MEPs are stratified to 189 sample areas. Then ten to fifteen mail-characteristic strata, based on reference volumes of letters, flats, and parcels, are defined independently for each sample area. Independently defined mail-characteristic stratification precludes any sort of aggregation of data across sample area. One drawback of the hierarchical stratifications is that, when there are test cancellations, there is no quick solution to fill statistical sampling strata with no useable tests.

This issue is addressed in the proposed methodology by removing the hierarchical structure of the stratification process like what was done for the digital population. We propose a ten geographical by twenty mail-characteristic cross-stratification, which results in total of 200 strata nationally. In this proposal, all MEPs are stratified based on the transformed reference volumes using the k-means algorithm. Defining a consistent mail-characteristic stratification allows for flexible choices of estimators, e.g., aggregate (or collapsed) data across sample areas or mail-characteristic strata.

Proposed Estimator:

MEP-level total reference volume, used as an auxiliary variable computed from historical data, provides us with a tool to re-gain some of the lost precision resulting from the proposed sample reduction. A ratio estimator is more efficient than the current expansion estimator given that the auxiliary variable is moderately correlated with the response variables, revenue, pieces, and weight. This point was discussed in Docket No. RM2016-1.[2] Developed through the historical data, total reference volume and

[2] *See* Docket No. RM2016-1, Proposal Eleven (October 7, 2015) at page 8 and Appendix at pages 10-11.

actual mail volume are highly correlated, e.g., a correlation of 0.793 based on 20,500 non-digital tests conducted in Quarter 2, FY2020. Due to a random selection, there is always a chance of selecting a non-representative sample, e.g., selected MEPs have small volumes relative to the rest of the MEPs in a stratum. The ratio estimator corrects for volume imbalances in a sample and makes it more representative.

Indices are defined as follows:

$$h = Strata\ based\ on\ Reference\ Volumes, (h = 1, \dots, 20)$$
$$j = Consolidated\ Sample\ Area, (j = 1, \dots, 10)$$
$$k = MEP\ day\ for\ stratum\ (h,j), (k = 1, \dots, n_{hj}, \dots, N_{hj})$$
$$l = Mailpiece\ for\ MEP\ day\ k\ in\ stratum\ (h,j), (l = 1, \dots, n_{hj}, \dots, N_{hjk})$$

Define

$$N_{hj} = the\ number\ of\ MEP\ days\ in\ strata\ (h,j)$$

$$n_{hj} = the\ number\ of\ MEP\ days\ sampled\ in\ strata\ (h,j)$$

$$N_{hjk} = the\ number\ of\ mailpieces\ in\ MEP\ day\ k\ in\ strata (h,j)$$

$$n_{hjk} = the\ number\ of\ mailpieces\ sampled\ in\ MEP\ day\ k\ in\ strata\ (h,j)$$

$$z_{hjk} = the\ total\ reference\ volume\ of\ sampled\ MEP\ day\ (h,j,k)$$

$$y_{hjkl} = \begin{cases} revenue, pieces\ or\ weight\ for\ mailpiece\ l\ for\ the\ product\ of\ interest, \\ 0, \quad otherwise \end{cases}$$

The proposed ratio estimator $\hat{t}_R$ for a given product is defined as

$$\hat{t}_R = \sum_{h=1}^{20} \hat{t}_{R,h} \stackrel{\text{def}}{=} \sum_{h=1}^{20} \frac{Z_h}{\hat{Z}_h} \hat{t}_h$$

where

$$\hat{t}_h = \sum_{j=1}^{10} \frac{N_{hj}}{n_{hj}} \sum_{k=1}^{n_{hj}} \frac{\hat{N}_{hjk}}{n_{hjk}} \sum_{l=1}^{n_{hjk}} y_{hjkl}$$

$$\hat{N}_{hjk} = (Container\ Skip) \times (Mailpiece\ Skip) \times n_{hjk}$$

$$\hat{Z}_h = \sum_{j=1}^{10} \frac{N_{hj}}{n_{hj}} \sum_{k=1}^{n_{hj}} z_{hjk}$$

$$Z_h = \sum_{j=1}^{10} \sum_{k=1}^{N_{hj}} z_{hjk}$$

The first-order Taylor approximation variance of the ratio estimator is

$$V(\hat{t}_R) = \sum_{h=1}^{20} V(\hat{t}_{R,h})$$

with

$$V(\hat{t}_{R,h}) = \left(\frac{Z_h}{\hat{Z}_h}\right)^2 \left\{\sum_{j=1}^{10} N_{hj}^2 \left(1 - \frac{n_{hj}}{N_{hj}}\right)\frac{S_1^2}{n_{hj}} + \sum_{j=1}^{10} \frac{N_{hj}}{n_{hj}} \sum_{k=1}^{N_{hj}} \widehat{N}_{hjk}^2 \left(1 - \frac{n_{hjk}}{\widehat{N}_{hjk}}\right)\frac{S_2^2}{n_{hjk}}\right\}$$

where

$$S_1^2 = \frac{1}{N_{hj} - 1} \sum_{k=1}^{N_{hj}} \left(t_{hjk} - R_h z_{hjk}\right)^2$$

$$R_h = \frac{t_h}{Z_h}$$

$$S_2^2 = \frac{1}{N_{hjk} - 1} \sum_{l=1}^{N_{hjk}} \left(y_{hjkl} - \bar{y}_{hjk}\right)^2$$

$$\bar{y}_{hjk} = \frac{1}{N_{hjk}} \sum_{l=1}^{N_{hjk}} y_{hjkl}$$

Simulation Methodologies

Because ODIS-RPW is a complex survey system with over 2,200 strata, an analytical assessment of the impacts of sample reduction on CVs is difficult. Statistical simulations are used for the purpose. Simulations were done using the most available useable time period data; Quarter 2, FY2020. Quarters 3 and 4, FY2020 were not used as the COVID-19 quarantine affected the completion of some ODIS-RPW tests.

Digital Simulation:

Since daily volumes are known for all MEPs and historical data are more abundant for the digital population, we constructed a pseudo-population to perform various simulations. These simulations included changes in sample size, sample design, and the estimator on the pseudo-population. The primary categories estimated from the digital letter and card population are First-Class Single Piece (FCSP) stamped and meter pieces. Then, because some of those pieces contain Certified and Return Receipt Extra Services, those services are estimated as well.

There are approximately 13,000 MEPs in the digital population in Quarter 2, FY2020, and DBCS daily volumes (pieces) are available for all MEPs. Day volumes provide us with the frame on which we build a pseudo-population; 13,000 MEP times 75 delivery days equals 975,000 MEP days in the population.

Three levels of sampling variation are incorporated in a pseudo-population. Products of interest are FCPS stamp (FCSP-NM), FCSP meter (FCSP-ME), and some special services such as Certified and Return Receipt. USPS mail is also included. The shares of the latter three categories are extremely small, therefore, we aggregated the three and formed the following four categories for the digital population: (1) FCSP-NM, (2) FCSP-ME, (3) SSUSPS (special services plus USPS), and (4) the rest.

The first level of variation is among-MEP variation. That is, the true proportions of products that vary by MEP. Since we consider four product categories in the digital population, proportions of the four categories are written as a vector of length four, denoted by $\boldsymbol{\alpha}$, with its elements summing to one. For example, the true proportion vector of a MEP i with 15 percent of its total volume V_i being FCSP-NM, 10 percent FCSP-ME, and 1 percent SSUSP, is $\boldsymbol{\alpha}_i = (0.15\,, 0.10, 0.01, 0.74\,)$. Proportion vector $\boldsymbol{\alpha}_i$ and volume V_i completely define the mail mixture of MEP i for $i = 1, \dots, 13000$. $\boldsymbol{\alpha}$ is estimated from historical data for each MEP. For each of 522 MEPs with no historical data, $\boldsymbol{\alpha}$ is randomly selected from MEPs with data and small random perturbations are added.

The second level of variation is the day variation for a given MEP i, i.e, within-MEP variation. This is incorporated as a day specific proportion vector $\boldsymbol{\theta}_{ij}$ which is a sample from its parent distribution, a Dirichlet distribution with parameters λ_i and $\boldsymbol{\alpha}_i$ for all i. In other words, $\boldsymbol{\theta}_{ij}$ varies about its true proportion vector with a propensity of deviation characterized by a parameter λ_i for all $i = 1, \dots, 13000$ and $j = 1, \dots, J$, where J is the number of delivery days in a period. The larger the value of λ, the closer $\boldsymbol{\theta}$ tends to be to $\boldsymbol{\alpha}$. Given λ_i ,$\boldsymbol{\alpha}_i$, and V_i, within-MEP variation is modeled by a sample from its parent distribution:

$$\boldsymbol{\theta}_{ij} \sim Dirichlet(\lambda_i \boldsymbol{\alpha}_i)$$

for all i and j. λ_i is an unknown parameter which needs to be estimated. A Bayesian hierarchical model is applied, and this model is discussed later in this section. Under the two-stage sampling scheme (sample MEP days in the first stage, and then sample mailpieces from the selected MEP days in the second stage), the first-stage variance, among-test variation, in the variance formula provided in the proposal captures the sum of first- and second-level variations.

The third level of variation is due to subsampling of mailpieces for a given test day, within-test variation. We typically select 175 mailpieces, but the subsampling size is not fixed as we use systematic sampling with a prescribed skip interval for each MEP. We use an average daily volume of recent quarters and compute a skip interval so that 175 mailpieces are subsampled on average. For a given day j of MEP i, the true proportion vector is $\boldsymbol{\theta}_{ij}$ and we select n_{ij} mailpieces and observe y_{ijk} for $k = 1, \dots, n_{ij}$. n_{ij} is derived from day volume v_{ij} and the skip interval for MEP i. Since v_{ij} vary across j , so does n_{ij}. These variations, due to subsampling and random sample size, are modeled by

$$\mathbf{y}_{ij} \sim Multinomial(n_{ij}, \boldsymbol{\theta}_{ij})$$

for all i and j. $\mathbf{y}_{ij}$ is a vector of length 4 with each element representing the number of mailpieces falling into its category. Alternatively, we can consider it as multinomial with size 1 applied to individual mailpiece,

$$y_{ijk} \sim Multinomial(1, \boldsymbol{\theta}_{ij})$$

with $k = 1, \dots, n_{ij}$.

The hierarchical framework of simulations is depicted in the Figure 1. A randomly selected MEP, say MEP 7, is characterized by the proportion vector $\boldsymbol{\alpha}_7$ and known volume for the entire period, $V_7 = \sum_{j=1}^{J} v_{7j}$ where v_{7j} is the volume of MEP 7 on day j for $j = 1, \ldots, J$. The total number of mailpieces by product category for MEP 7 during the period is $\boldsymbol{\alpha}_7 V_7$. Then a randomly selected MEP day 5 is considered as a random sample from its parent distribution,

$$\boldsymbol{\theta}_{75} \sim Dirichlet(\lambda_7 \boldsymbol{\alpha}_7)$$

and n_{75} mailpieces are sampled from v_{75} mailpieces,

$$y_{75k} \sim Multinomial(1, \boldsymbol{\theta}_{75})$$

for $k = 1, \ldots, n_{75}$.

Simulation steps are as follows:

1. Obtain $\boldsymbol{\alpha}_i$ from historical data and randomly select λ_i from a uniform distribution for all i.
2. Fix $\boldsymbol{\theta}_{ij}$ by sampling $\boldsymbol{\theta}$ from their respective parent distribution for all i and j. With known v_{ij}, volumes by product are fixed for all i and j. National totals are the sum of $\boldsymbol{\theta}_{ij} v_{ij}$. These make up the pseudo-population.
3. Fix a sample design and sample size, and estimator.
4. Randomly select MEP days from the subpopulation according to the sample design in step 3.
5. Expand the random sample in step 4 according to the estimator in step 3.
6. Repeat steps 4. and 5. to obtain 2,000 sets of estimates and compute CVs.

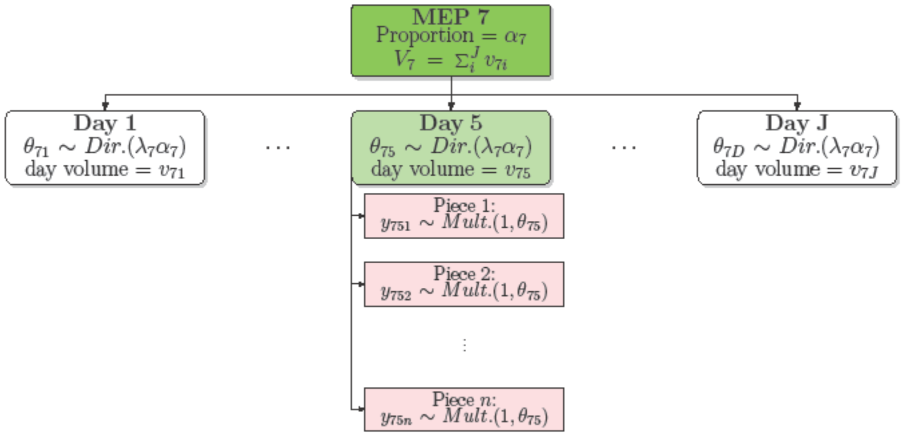

Figure 1: Hierarchical Structure of Digital Population Simulation

<u>Bayesian Hierarchical Model for λ:</u>

Because λ's are unknown, they are initially estimated for fifty-two randomly selected MEPs by a Bayesian hierarchical model. Fifty-two digital MEPs with thirty or more tests conducted in last five years are randomly selected to identify the variation of a hyper-parameter λ.[3] The hierarchical model is used strictly to identify a range of reasonable λ from fifty-two MEPs so that we can attach a value of λ to each MEP. Although a randomly assigned value is not reflective of the true value of λ, we are expecting an averaging-out effect to take place with 13,000 MEPs in the population. Ultimately, it is an overall similarity in variation of mail mixture of a pseudo-population

[3] The model requires extensive computing time due to its use of a Markov Chain Monte Carlo (MCMC) algorithm in obtaining posterior distributions. It was not feasible to apply the hierarchical model to all 13,000 MEPs.

that is critical in assessing effectiveness and impact of design and sample size. As such, simulation is not intended to exactly match all characters of individual MEPs.

Our choice of prior distributions for the hyper-parameters λ and $\boldsymbol{\alpha}$ are dispersed and intend to be non-informative:

$$\lambda_i \sim Exponential(0.001)$$
$$\boldsymbol{\alpha}_i \sim Dirichlet(\mathbf{0})$$

where $\mathbf{0}$ represents a vector of length four with each element equal to 0.0000001. Then the rest of hierarchical model structure is as already described, i.e.,

$$\boldsymbol{\theta}_{ij} \sim Dirichlet(\lambda_i \boldsymbol{\alpha}_i)$$

and

$$\mathbf{y}_{ij} \sim Multinomial\left(n_{ij}, \boldsymbol{\theta}_{ij}\right)$$

A statistical software package called ‘Stan’ with a Hamiltonian Monte Carlo algorithm is used to obtain posterior distributions. Each chain uses the first 2,000 samples as warm-up and keeps 500 samples after the warm-up. The posterior inference is based on total of 2,000 samples from four chains. Diagnostics show no issues of convergence except for MEPs with extremely low proportions of the third category, SSUSPS. As mentioned, this category (certified, return receipt, and USPS mails combined) is extremely rare in the digital population.

The model is fit individually to fifty-two MEPs and posterior distributions of the hyper-parameters are obtained. Weighted averages from historical data are like the mean of posterior distribution of $\boldsymbol{\alpha}$. Therefore, the use of weighted average in lieu of posterior $\boldsymbol{\alpha}$ is corroborated. The posterior means of λ for fifty-two MEPs vary more or less uniformly within the interval (40, 200). MEPs with higher proportions of FCSP seem to have a smaller variation, uniformly within an interval (20,100). We, therefore, randomly assign λ_i to MEP i, for $i = 1, \ldots, 13000$, from uniform distribution with an interval (20,150) for MEP with the proportion of FCSP less than 0.4 and from uniform (20,60) for MEP with higher than 0.4 FCSP proportion. We use slightly smaller upper-bounds for uniform distributions so as to generate λ that exhibits a larger variance.

Figure 2 compares observed and simulated numbers of FCSP-NM and FCSP-ME for ZIP Codes XXXXX and YYYYY. There were forty and thirty-seven tests conducted for those two ZIP Codes respectively. We simulated the number of FCSP-NM and FCSP-ME based on the posterior estimates from the hierarchical model for forty and thirty-seven tests and histograms are shown side by side. The hierarchical model appears to generate similar numbers of products with a slightly larger variation. The other fifty MEPs showed similar results. CVs from the simulation are likely to be more conservative.

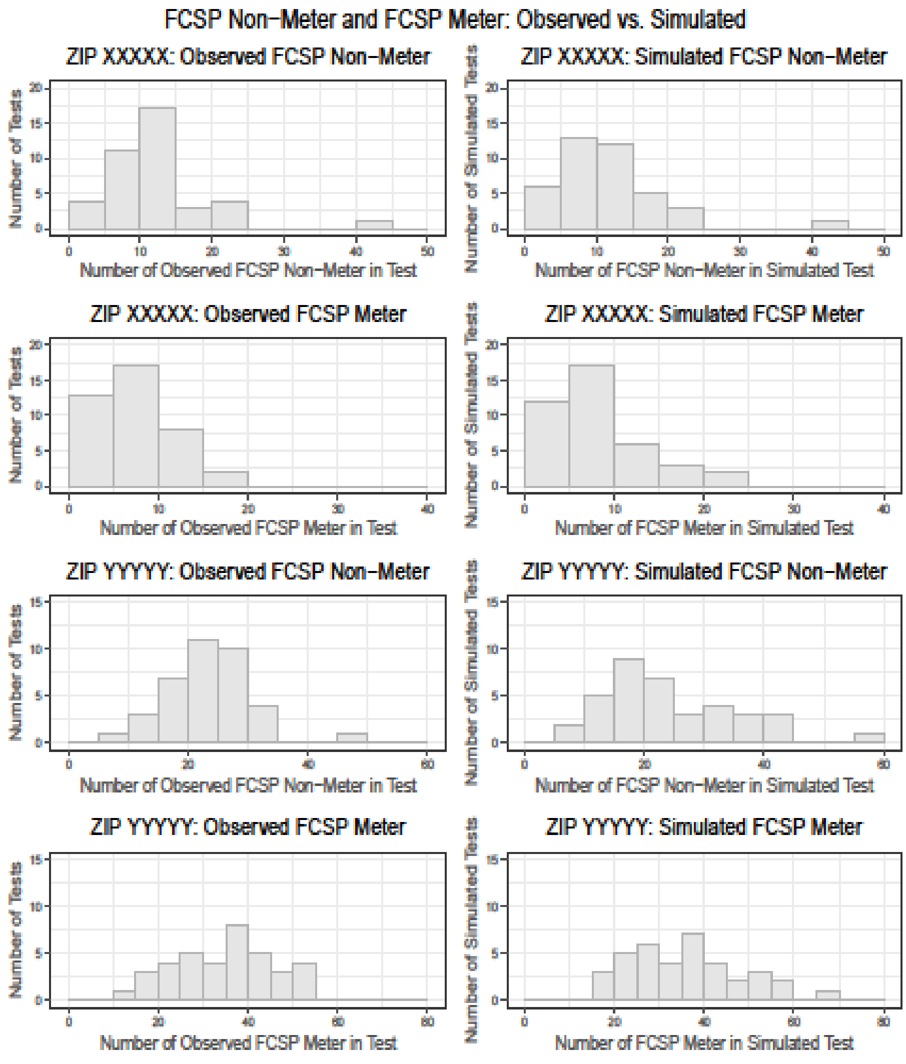

Figure 2: Comparisons of Observed and Simulated FCSP Non-Meter and FCSP Meter for ZIP 20003 and 60603.

<u>Non-Digital Simulation:</u>

There is no frame to build a pseudo-population as was done for the digital simulations, because many of the approximate 55,000 MEPs remain unsampled in a given year. Instead, we used the completed Quarter 2, FY2020 ODIS-RPW tests, and a bootstrapping simulations (see below). Products considered in this non-digital population of Quarter 2 FY2020 tests are FCSP stamped, FCSP meters, FCSP cards, First-Class (FC) Flats, Media and Library Mail, Certified Mail, Return Receipt, FC Package Service – Commercial, FC Package Service – Retail, Domestic Priority Mail – Commercial, Domestic Priority Mail – Retail, and USPS Mail.

A standard bootstrap understates the true variation because the method accounts for among-test variation (the first-stage variance), but within-test variation (the second-stage variance) is not addressed. Within-test variation is expected to be much smaller than among-test variation. For example, the ratio of the second-stage variance to the first-stage variances is 0.08 to 0.13 based on the digital simulation, and the relationship is expected to hold for the non-digital population as the second-stage variances for both populations are due to the use of a systematic sampling which is characterized by a multinomial distribution.

Bootstrap using with-replacement sample is likely to contain the same tests sampled multiple times. Especially under the current design with over 2,200 strata, there are strata with a small number of tests. For example, a stratum consists of 100 MEPs may have three tests in a quarter. Then repeated with-replacement samplings of size three are not likely to simulate the potential variation that is exhibited by the 100 MEPs in the stratum. When a test in selected multiple times, we add random perturbation to the estimates for duplicate tests. If a test is selected twice, for example, a number u from a uniform distribution with an interval (0,1) and a random sign s (+ or -), then estimates of one of duplicates are multiplied by $(1 + su)$ to have new estimates. The addition of random perturbation inflates CVs by 5-8 percent compared to CVs from bootstraps without random perturbation. CVs estimated from the bootstrap with random perturbation are further inflated by 10 percent to compensate for the lack of within-test variation.

Simulation steps are as follows:

1. Fix a sample design and sample size, and estimator.
2. Obtain a bootstrap sample under 1.
3. If a test is selected multiple times, then add random perturbations to duplicate(s).
4. Expand the test-level estimates according to 1 to obtain national level estimates for the products.
5. Repeat steps 2 through 4 to obtain 2,000 sets of estimates, and compute CVs.
6. CVs in 5 are multiplied by 1.1 to make up for the second-stage variance.