Sample Update Document
For
The Advance Retail Trade Survey (MARTS)

Abstract

This document contains specifications for selecting the 2003 sample for the MARTS. As such, it is a comprehensive guide to selecting a sub sample to enable reliable early estimates.

This document was provided to the STES timeliness framework by Tom Zabelsky:
thomas.e.zabelsky@census.gov

Changes from the last sample selection.

1. We used the six most recent monthly sales data collected from the Monthly Retail Trade Survey (MRTS) to compute the size measure for stratification. Last time, the sales data from the most recent Annual Retail Trade Survey (ARTS) were used.

2. We kept only good reporters, i.e. units reported to MRTS in at least four out of the last six months, in the MARTS sampling frame.

3. We collapsed some strata so that a collapsed stratum contains at least two units.

4. We assigned BSR2k big effect cases to appropriate strata according to their measure of size based on reported sales in the MRTS.

5. Bookstores (451211) were classified into three new sampling recodes, general bookstores (45121A), specialty bookstores (45121B) and college bookstores (45121C).

6. We carried the random groups from MRTS noncertainty units to MARTS sample units and randomly assign random groups to other MARTS noncertainty units. We did not reassign random groups to MRTS units as in the last sample selection.

7. We selected at least two units in each sampling stratum with the exception of bookstores. This will give us better chance to have at least one unit per stratum for MARTS estimation.

I. Frame Creation

The Advance Monthly Retail Trade Survey (MARTS) sampling frame is extracted from the Monthly Retail Trade Survey (MRTS) registry. We kept only MRTS units that are active as indicated on the registry. To improve the efficiency of our sample, we used MRTS units that responded in four of the six most recent months. We also used the MRTS sales in the last six months to compute the measure of size.

1. RISPN data

Location: EPBA22::RISPN_2K_DRESS_DB:[RISPN2K]RISPN_BSR2KDESS_DB.RDB
RISPN table: R_REPORTING_UNIT_TABLE
Select records that satisfy all of the following (valid units):

a) BSR_Panel_MRTS is 0, 1, or 5;
b) Report_code_MRTS is not 0, 7, or 8;
c) Pending_report_code is not B or C;
d) BSR_activity is not 0;
e) BMF_Activity is not 0;

Create a variable ID by dropping the first two-digits from ALPHA_EI_AREA.
Rename PART_COD to PART.

Keep the following variables:

Alpha/Employer ID	ID
Part Code	PART
Predecessor's alpha/EID	PRED_AEA
Predecessor's part code	PRED_PCD
Firm Name	FIRM_NAM
MRTS Weight	WGT_MRTS
MARTS Weight	MARTS_WEIGHT
MRTS Stratum	STRATUM_MRTS
Census 97 Equivalent Sales	CENSUS_EQUIV_SALES
MRTS Random Group	RG_MRTS

2. MRTS Detail Files

Location:
EPBA22::BSR2K_RETISPN:[ISPN.PROD.DAT]
File name:
RDETmmyy.SAS7BDAT
(where mm is the 2-digit month and yy is the 2-digit year). Select records that are in segment A.

For the MRTS detail file of the latest month (month t), keep the following variables:

Employer ID	ID
Part Code	PART
6 digit NAICS code	NAICS6
Current Month Preliminary Sales	CM
Past Month Final Sales	PM

Current Month Weighted Preliminary Sales	CMWGT
Past Month Weighted Final Sales	PMWGT
Current Month RI Code	SCMRI
Past Month RI Code	SPMRI

For month t-1, t-2, t-3, t-4 keep only four variables: ID, PART, PLANT, SPMRI and PMWGT. Merge these five data by ID, PART and PLANT and keep records that are in the most recent month (month t).
A unit is reported in a month in MRTS if the RI code for that month (SCMRI or SPMRI) has a value 1, 4, 5, 6, or 7.
If ID and PART in a detail file for earlier months have been changed and do not match to those in the current RISPN data, we updated the ID and PART using the link provided by predecessor's id/part code in the RISPN file.

3. Compute the annualized sales variable.

Select one record for each combination of ID/PART from the data in I.2. Then merge this data with the data in I.1. Keep records that are in both data sets in I.1 and I.2.

We use 97 census pseudo sales as the measure of size. We first compute the annualized sales, AnnSales, for each ID and sampling recode combination. Let *i* be a ID and sampling recode combination. Let $S_t(i), S_{t-1}(i), S_{t-2}(i), S_{t-3}(I,) S_{t-4}(i), S_{t-5}(i)$ be the (unweighted sales) for the last six months for *i*. (preliminary sales for month t and final sales for other months). If combination *i* had positive sales in at least one of the last six months, then

$$Annsales(i) = \frac{S_t(i) + S_{t-1}(i) + \cdots + S_{t-5}(i) \cdot 6}{n(i)}$$

where,

$S_t(i), S_{t-1}(i)$ etc. = sales for *i* in month *t, t-1* etc. and
$n(i)$=number of months (during the last six months) in which the unit *i* had positive sales.
If an ID and sampling recode combination *i* did not have any positive sales in the last six months, then
$Annsales(i) = 0$.

4. Create the ID level data

Assign a major KB (sampling recode) for an ID that is the sampling recode that has the largest Annsales within the ID. Compute the Annsales and Census1997 Equivalent Sales at the ID level by summing the Annsales and Census Equivalent Sales within the ID. The number of months reported for an ID is the maximum number of months reported for its parts.

5. Compute Stratification Measure-of-Size (PS97)

For each ID, we used the BSR2k parameter study stratum-boundaries and the ID's major KB recode to assign the ID to two sets of size strata. The first set (Censales strata) is assigned using

the Census Equivalent Sales. The second set (Annsales strata) is assigned using c·AnnSales, where c is a sampling recode level constant. For each sampling recode, we compute N_0, the number of IDs that Censales stratum=Annsales stratum and N_1, the number of IDs that Censales stratum and Annsales stratum differed by 1. Then the measure of Size variable *PS97* is defined as c_0·AnnSales, where the sampling recode level constant c_0 is the constant that maximizes $2 \cdot N_0 + N_1$.

6. Create the preliminary frame and the sampling frame

The MARTS preliminary sampling frame contains one record per valid ID. It is the ID level data in I.4 with the addition of the measure-of-size variable PS97. This preliminary frame contains both units that are good reporters and units that are not good reporters.

A good reporter is a unit who reported in at least four months in the last six months and has a positive measure of size.

The sampling frame is formed from the preliminary sampling frame by selecting units (IDs) that are good reporters. We also include new units that were in MRTS sample for less than six months provided that the unit was in MRTS for at least three months and reported for at least two months. For units in the take-all KBs, all units in the preframe are selected in the frame.

II. Certainty Determination and Stratum Assignment

Like in the MRTS, MARTS uses two level strata. The primary level is the sampling recode. The secondary level is the size.

There are a few criterions we use to determine certainty units. We select a number of MRTS certainty units that have the largest measure of size in each KB to be certainty units. We worked with the staff in the Retail and Wholesale Indicators Branch to determine the number of this kind of certainties for each sampling recode. There are some MRTS certainty units that are not selected as certainties using the previous criterion, but the analysts in the survey area want to include these units in the MARTS sample. These units are the must certainties. A list of must certainties are provided by the analysts. There is also a list of sampling KBs (452111, 452112, 452113, and 452910) in which the analysts want to include all valid units in the sample. Hence, a unit is chosen to be a MARTS certainty unit if it satisfies one of the following:

1. It is in a take-all KB.
2. It is a MRTS certainty unit and its measure of size is among the largest for its sampling recode.
3. It is a MUST case.

We applied the measure-of-size (PS97) and MRTS stratum boundaries to assign units to secondary strata. Units whose measure of size is above the MRTS certainty cutoff but not selected as MARTS certainty are assigned to an additional stratum.

Secondary strata that have less than two sampling units are collapsed with other strata to form new secondary strata that have at least two units (at least one units in bookstore strata, 45121A, 45121B and 45121C). Near certainty stratum will be collapsed with the certainty stratum if it does not contain any MRTS non-certainty unit and 1) it has less than two units or 2) its sample size (calculated in part III) equals its stratum size.

III. Sample Size Computation

The methodology of computing the sample size for this sample update is the same as that used in the previous sample.

1. CV constraints

We determined sample sizes for each sampling recode and measure of size stratum. We refer to each combination (sampling recode, measure of size stratum) as a "two-level stratum".
We designed sample sizes to meet the following the cv constraints in Attachment 4. They were obtained via the following MRTS CV constraints located at:

EPBA22::CBUS_BSR2K3:[PRDB.RETAIL.FRAME.MRTS].FINAL121
Subsetting criterion: Level not in ('00RFA', '72') and substr(Level,1,3) ne '721'
Selected variables: Level CV

In the previous MARTS sample, the MRTS cvs were multiplied by the factor 1.7/.895. To increase the sample sizes to meet the requirement, we reduced the CVs constraints for selected sampling recodes as requested from the Retail and Whole Sale Indicators Branch.

2. Calculating initial sample sizes

We calculate for each level L for which a CV constraint, cv_L , was supplied

$$Mult_L := (\sum_h N_h s_h) / (\sum_h N_h s_h^2 + (cv_L * (\sum_h Y_h))^2)$$

where h runs through the two-level strata in level L
N_h := the size of stratum h
s_h := the standard deviation of EstSales in stratum h
Y_h := the total of EstSales in stratum h

We calculate for each two-level stratum h an initial sample size
$n_h := \max\{Mult_L : L \text{ contains } h\} N_h s_h$

We determine the collection F of two-level strata for which $n_h \geq N_h$. We reset $n_h := N_h$ for these strata. (These are the strata where the initial sample size oversampled, and we are fixing ("freezing") their sample sizes to be the stratum sizes. These strata will be called "frozen".)

Three additional constraints are also applied. The first is the maximum weight constraint that assures the final weights will not exceed 500 (for recode 445110, 722110, 722211, and 722213). The second constraint is the minimum sample size constraint that requires a sample of at least

two units in a stratum. The third constraint is that the sample size can not exceed the number of units in MARTS sampling frame in a stratum.

3. Re-calculate the sample sizes for the non-frozen strata

We re-calculate for each level L for which a cv constraint, cv_L , was supplied

$$Mult_L := (\Sigma_h\ N_h\ s_h)\ /\ (\Sigma_h N_h\ s_h^2 + (cv_L * (\Sigma_h\ Y_h))^2 - Freeze_L)$$

where h now runs through the non-frozen strata h in level L

$$Freeze_L := \Sigma_i\ \{ (N_i\ s_i)^2\ /\ n_i\ - N_i\ s_i^2\ : \text{frozen strata i in L}\}$$

We re-calculate for each nonfrozen stratum h

$$n_h := \min (N_h , \max\{Mult_L : \text{L contains h}\} N_h\ s_h)$$

4. Rounding the sample sizes

We used a relatively simple method to round the sample size. This method rounds up sample sizes whose fractional parts are close to 1 and rounds down sample sizes whose fractional parts are close to zero while keeping the total amounts rounding up and the total amounts rounding down relatively balanced.

Sampling recode 443120 (computer stores) is a second stage take-all KB. All units in the sampling frame in this KB will be selected in the sample.

IV. Sample Selection

We applied a stratified systematic PPS selection available in a SAS procedure SURVEYSELECT to select the MARTS sample.

PPS certainties need to be identified in order to use this SAS procedure. A unit *i* in stratum *L* is a PPS certainty if

$$\frac{PPSMOS_i}{\sum_{l\in L} PPSMOS_l} \geq \frac{1}{n_L}$$

where n_L is the sample size in stratum *L*.

In selecting the 2003 MARTS sample, we experienced a significant drop in the number of PPS certainties from the previous MARTS sample selection. One explanation is that our procedure to that minimizes the number of stratum-jumpers when we compute pseudo-sales (PS97) may have an effect on reducing the MOS discrepancies within strata. Consequently, it reduces the number of PPS certainties.

The sampling frame was then sorted by sampling recode, size strata, MRTS random group and the pseudo sales. We use a stratified systematic PPS selection procedure to select the sample.

The MARTS weight has 3 components: MRTS sampling weight, a P-factor, and second phase selection weight.
The P-factor accounts for the discrepancies between the number of MRTS active units and the number of units in the MARTS sampling frame (good reporters).

$$\text{P-factor:} \quad = \frac{\text{Number of valid cases on RISPN in a stratum}}{\text{Number of valid cases on MARTS selection frame in a stratum}}$$

The numbers on the numerator and the denominator are weighted by MRTS weight. The stratum used here was created from the BSR2k MRTS stratum assignments, where each certainty stratum is split into two: one for MARTS certainties and one for the MRTS certainties that are not MARTS certainties.
PPS Measure of size = $MRTS\ wgt \times Pfactor$

The total selection probability has three components:

1) MRTS selection probability is $\frac{1}{MRTS\ wgt}$
2) Frame restriction is represented by $\frac{1}{pfactor}$
3) Second stage selection probability is

   $= n_{h\,MARTS} \times \frac{MRTS\ wgt \times pfactor}{\sum_h MRTS\ wgt \times pfactor}$, where n_{hMARTS} is the initial sample size for stratum *h.*

If we multiply all 3 of the selection probabilities, we see that the final selection probability is

$$MARTS(p) = \frac{1}{MRTS\ wgt} \times \frac{1}{pfactor} \times n_{h\,MARTS} \times \frac{MRTS\ wgt \times pfactor}{\sum_h MRTS\ wgt \times pfactor}$$

Simplifying the above equation we see that the full weight simplifies to be $\frac{N_{U\,h}}{n_{MARTS\,h}}$, where $N_{U\,h}$ is the stratum size in the universe, and $n_{MARTS\,h}$ is the initial MARTS sample size in stratum *h.*

V. Random Group Assignment

MARTS and MRTS both compute their variances using the random group method. MARTS sample is a sub-sample of the MRTS sample. We did not reassign random groups to the noncertainty units in MRTS as in the last MARTS sample selection.

We assign random group to the units selected in the MARTS sample as follows:

MARTS sample units	MARTS random group
MRTS random group=0 and MARTS weight>1	Random assignment
Others	MRTS random group

There are several reasons a sample unit (who must be a MRTS certainty) could be assigned to MARTS random group 0. 1) It is a certainty unit during certainty determination. 2) It is in a near certainty stratum that was later collapsed with the certainty stratum. 3) This MRTS certainty unit is in a stratum in which all units were selected and its P-factor is one.

Note.
We excluded active MRTS units whose MRTS report code (PCD_MRTS) =0,7,or 8 (confirmed out-of-business, confirmed out-of-scope and confirmed refusal) and MRTS units whose pending report code is B or, C (pending refusal and pending out-of-business) from the preframe in the 2003 MARTS sample. This same procedure was used in the BSR2K Marts sample. This resulted in the total weighted count of the MARTS preframe being substantially below the total count of the MRTS frame or the total weighted number of units in current MRTS reporting unit files. If we would include this units in the preframe, the p-factors (see IV) would be as large as 7.9 and the MARTS weights would be as big as 2,500 for some units. We believe that we need to conduct more research to address this issue the next time we draw the MARTS sample.

Attachment 1. Definitions of selected RISPN and ARTS variables

RISPN variables

RISPN variable name	Definition, and possible partial list of values
Alpha_EI_Area	alpha for alpha sample units EIN for EI sample units
SSEL_SIC	RISPN KB
BSR_Panel_MRTS	Panel
Firm_Name	Name
Census_Equiv_Sales	Census equivalent sales
Weight_MRTS	MRTS sample weight
Stratum_MRTS	MRTS (measure of size) stratum
Report_Code_MRTS	Report code 0: confirmed out of business 7: confirmed out of scope 8: confirmed refusal
Pending_Report_Code	Pending report code B: pending refusal C: pending out of business
BSR_Activity	BSR activity code 0: inactive 1: active
BMF_Activity	BMF activity code 0: inactive 1: active
RG_MRTS	MRTS random group

Attachment 2. MARTS sampling recodes

Attachments 2 and 3 use the following aggregate levels.

Definitions of selected aggregate levels

Aggregate level	Definition
04445	440000∪ 450000
04445x1	(440000 \ 441000) ∪ 450000
0CONTROL	(440000 \ (441100∪ 441300∪ 443120∪444000)) ∪ (450000 \ (453210∪ 453930∪ 454310))
0gaf	442000∪ 443000∪ 448000 ∪ 451000∪ 452000
4461x0	446100 \ (446110 ∪ 446130)

MARTS SAMPLING RECODES

452111	441110	448130	442100	45121A	4461X0
452112	441120	448140	444200	45121B	
452113	441210	448150	445300	45121C	
441221	441310	448190	448200		
441222	441320	448310	453100		
441229	442210	448320	453300		
442291	443120	451110	454100		
442299	443130	451120	454200		
443111	444110	451130	722100		
443112	444120	451140	722400		
445291	444130	451220			
445292	444190	452910			
445299	445110	452990			
451212	445120	453210			
453991	445210	453220			
453998	445220	453910			
454311	445230	453920			
454312	446110	453930			
454319	446130	454390			
722211	447110	722310			
722212	447190	722320			
722213	448110	722330			
	448120				

Attachment 3. Desired numbers of certainties by sampling recode

These are the number of largest units that are designated as certainties.

RECODE	NUMCERTS
441110	39
441120	13
441210	5
441221	3
441222	5
441229	4
441310	9
441320	4
442100	21
442210	11
442291	3
442299	6
443111	12
443112	16
443120	All MRTS certainties
443130	2
444110	7
444120	2
444130	5
444190	14
444200	14
445110	62
445120	4
445210	2
445220	5
445230	2
445291	3
445292	5
445299	2
445300	8
446110	11
446130	4
4461X0	3
447110	25
447190	10
448110	13
448120	20
448130	3
448140	11
448150	2
448190	5
448200	9
448310	9
448320	4
451110	9
451120	3
451130	2
451140	2
451211	4
451212	3
451220	5
452111	TAKE ALL
452112	TAKE ALL
452113	TAKE ALL
452910	TAKE ALL
452990	35
453100	3
453210	5
453220	4
453300	3
453910	2
453920	1
453930	4
453991	1
453998	3
454100	22
454200	2
454311	14
454312	7
454319	1
454390	4
722100	51
722211	44

722212	4
722213	13
722310	6
722320	2
722330	2
722400	2

Attachment 5. MARTS Sample Size and CV by Sampling Recode

Sampling recode	# of Certainties	# of Noncertainties	Sample Size	# in Frame	BSR2K Size	Design CV	Achieved CV	BSR2K Design CV
441110	39	353	392	868	364	1.57%	1.57%	1.70%
441120	13	97	110	192	101	7.76%	7.76%	8.50%
441210	5	15	20	33	22	17.00%	13.52%	17.00%
441221	3	15	18	33	21	17.00%	13.74%	17.00%
441222	5	24	29	52	32	11.56%	11.52%	11.56%
441229	4	20	24	61	20	17.00%	14.20%	17.00%
441310	9	62	71	115	82	8.50%	5.03%	8.50%
441320	4	32	36	61	33	8.50%	7.19%	8.50%
442100	22	103	125	235	101	3.80%	3.78%	5.10%
442210	11	46	57	106	56	10.20%	8.15%	10.20%
442291	3	13	16	28	16	17.00%	16.52%	17.00%
442299	18	38	56	81	46	5.80%	5.81%	17.00%
443111	13	104	117	226	123	4.68%	4.67%	4.68%
443112	16	58	74	138	79	3.40%	3.39%	3.40%
443120	52	135	187	187	352	5.10%	4.72%	5.10%
443130	3	15	18	27	15	9.50%	9.59%	9.50%
444110	6	8	14	23	14	5.10%	2.13%	5.10%
444120	2	12	14	22	14	17.00%	12.26%	17.00%
444130	5	41	46	105	51	6.80%	6.69%	6.80%
444190	14	114	128	238	128	11.90%	4.30%	11.90%
444200	15	42	57	106	52	10.88%	9.01%	10.88%
445110	62	204	266	417	275	1.02%	1.04%	1.02%
445120	4	37	41	54	42	15.30%	10.40%	15.30%
445210	2	24	26	61	27	13.60%	13.44%	13.60%
445220	5	22	27	54	32	13.60%	12.85%	13.60%
445230	2	18	20	34	17	17.00%	17.41%	17.00%
445291	3	14	17	32	22	13.60%	12.12%	13.60%
445292	5	9	14	23	20	17.00%	15.56%	17.00%
445299	6	15	21	45	19	17.00%	15.64%	17.00%
445300	20	89	109	244	139	5.95%	5.95%	5.95%
446110	11	56	67	106	73	3.40%	3.33%	3.40%
446130	4	21	25	37	28	8.50%	8.34%	8.50%
4461X0	3	27	30	48	32	13.60%	12.11%	13.60%
447110	26	150	176	315	185	3.40%	3.24%	3.40%
447190	10	132	142	167	115	6.00%	5.98%	6.80%
448110	13	40	53	87	58	5.10%	5.43%	5.10%
448120	21	75	96	139	99	2.55%	2.54%	2.55%
448130	5	14	19	22	15	17.00%	17.39%	17.00%
448140	12	39	51	76	47	3.40%	3.26%	3.40%
448150	2	14	16	24	14	16.66%	14.99%	16.66%
448190	5	14	19	33	22	16.66%	14.33%	16.66%
448200	12	33	45	64	42	4.25%	4.00%	4.25%
448310	9	52	61	119	60	6.80%	6.76%	6.80%
448320	4	9	13	17	16	13.60%	12.09%	13.60%
451110	9	52	61	125	71	6.63%	6.62%	6.63%
451120	6	17	23	29	31	5.10%	4.73%	5.10%
451130	2	14	16	24	15	17.00%	11.00%	17.00%
451140	4	12	16	22	16	17.00%	16.84%	17.00%

Sampling recode	# of Certainties	# of Noncertainties	Sample Size	# in Frame	BSR2K Size	Design CV	Achieved CV	BSR2K Design CV
451211	.	.	.	.	41	5.10%	4.51%	5.10%
45121A	2	19	21	32	.	.	4.94%	.
45121B	1	14	15	26	.	.	14.57%	.
45121C	1	9	10	13	.	.	12.50%	.
451212	3	16	19	22	18	17.00%	17.07%	17.00%
451220	5	15	20	36	22	17.00%	8.60%	17.00%
452111	19	.	19	19	31	0.00%	0.00%	0.00%
452112	8	.	8	8	15	0.00%	0.00%	0.00%
452113	1	.	1	1	3	0.00%	0.00%	0.00%
452910	7	.	7	7	7	0.00%	0.00%	0.00%
452990	35	22	57	78	63	6.80%	5.74%	6.80%
453100	3	35	38	55	46	10.00%	9.49%	13.60%
453210	5	24	29	68	29	4.25%	3.91%	4.25%
453220	4	47	51	91	56	10.00%	9.07%	17.00%
453300	3	45	48	59	32	10.00%	10.45%	17.00%
453910	3	25	28	39	21	10.00%	10.58%	17.00%
453920	1	34	35	52	21	12.00%	12.64%	17.00%
453930	4	32	36	65	30	16.49%	8.79%	16.49%
453991	1	48	49	69	40	10.00%	10.11%	11.90%
453998	3	42	45	69	38	10.00%	10.27%	17.00%
454100	32	148	180	223	166	2.60%	3.33%	3.40%
454200	2	13	15	23	17	17.00%	17.31%	17.00%
454311	24	30	54	119	57	8.50%	8.31%	8.50%
454312	7	27	34	57	33	8.50%	7.74%	8.50%
454319	8	7	15	24	26	17.00%	16.09%	17.00%
454390	4	35	39	57	46	17.00%	11.63%	17.00%
722100	52	182	234	330	210	2.30%	2.33%	3.40%
722211	44	138	182	251	156	2.39%	2.47%	5.10%
722212	4	25	29	52	32	5.10%	5.07%	5.10%
722213	13	46	59	123	59	5.10%	5.04%	5.10%
722310	9	9	18	25	17	10.20%	6.82%	10.20%
722320	2	16	18	32	19	17.00%	14.05%	17.00%
722330	3	17	20	23	18	17.00%	16.70%	17.00%
722400	2	87	89	196	102	6.80%	6.69%	6.80%
	844	3757	4601	7699	4725			

Attachment 6. Record layout for register creation load file

We issued an ASCII file containing information for load into the MARTS register about the selected units. Following are the variables and the range of character fields they occupy in the issued file.

Variable	Lacation	Format
1. Primary identification field (alpha or EIN)	1-10	$10.
2. Sampling recode	11-16	$6.
3. MRTS weight	17-25	9.4
4. MARTS weight	26-34	9.4
5. MARTS random group assignment	35-36	$2.
6. Payroll-based sales	37-51	15.2.
7. Stratification measure of size	52-63	12.
8. MARTS stratum	64-65	2.