Size-Based Probability Sampling with Constraints on Costs

Randall K. Powers and John L. Eltinge

Office of Survey Methods Research, U.S. Bureau of Labor Statistics

Powers.Randall@bls.gov

Abstract: This paper presents a simulation study of some properties of size-based probability sampling with unequal unit-level costs, subject to constraints on aggregate costs. Principal emphasis centers on the distribution of realized sample sizes; and on the distribution of estimation errors for a ratio estimator of per-unit population means.

Key words: anticipated variance; expected sample sizes; probability-proportional-to-size (pps) sampling; unequal-probability sampling; variable unit-level costs.

1. Introduction

In work with large-scale establishment surveys, the sizes of population units often display strong patterns of right skewness. Such patterns can have important effects on the efficiency of sample designs for two reasons. First, selection of units with probabilities proportional to size can produce marked reductions in the variances of standard point estimators for a given fixed sample size; see, e.g., Cochran (1977), Godambe (1982), Holmberg and Swensson (2001), Kott and Bailey (2000), Zangeneh and Little (2015) and references cited therein.

Second, in some cases the costs of data collection can vary substantially across sample establishments, with collection from larger or more complex establishments often incurring higher costs. Because survey field operations generally have fixed budgets, variable unit-level costs can present special challenges when one tries to optimize the balance between cost and estimator variance. For example, the combination of cost constraints and variable unit-level costs may lead to variability in realized sample sizes, which in turn can complicate efforts to control variances.

To explore this variable-cost issue, this paper presents the results of a simulation study based on two populations of establishments. Section 2 describes the populations, with special emphasis on six cost functions. Section 3 outlines the design of the simulation study and presents results for realized sample sizes and for the properties of a simple ratio-based point estimator. Section 4 reviews the concepts and results considered in this paper and suggests some potential areas of future work.

2. Finite Populations, Illustrative Cost Functions and Related Characteristics

This study used two relatively large finite populations (industries), labeled B and C, that were considered previously by Powers and Eltinge (2013, 2014); see these references for detailed descriptions of the populations. For the current work, three features are of special interest. First, the populations consist of establishments for which we have information for several consecutive quarters; we will focus primary attention on estimation of population means for the variables

y_{2i} = total wage payments by unit i in quarter 2; and

y_{4i} = total wage payments by unit i in quarter 4

Second, for each unit i we considered six distinct illustrative cost functions, defined as:

Cost0: $c_{0i} = 1$

Cost1: $c_{1i} = 1 + ln(e_{1i})$

Cost2: $c_{2i} = 1 + (e_{1i})^{1/2}$

Cost3: $c_{3i} = 1 + e_{1i}$

Cost4: $c_{4i} = 1 + (e_{1i})^2$

Cost5: $c_{5i} = 1 + exp(e_{1i})$

where e_{1i} is the establishment-level employment count in the first (reference) quarter for unit i. Note that the function Cost0 is constant for all units, and thus will lead to results that are equivalent to those obtained through standard probability-proportional-to-size designs with prespecified sample sizes. The functions Cost1 and Cost2 display relatively slow growth as the value of e_{1i} increases; and Cost3 and Cost4 display more pronounced rates of growth. In addition, Cost5 is intended to explore the effects of relatively extreme (exponential) rates of growth for costs. Tables 1 and 2 present summary statistics for industries B and C, respectively, including the population mean, standard deviation, and skewness coefficient, as well as the 0.10, 0.25, 0.50, 0.75 and 0.90 quantiles for each of the abovementioned cost functions, y_{2i} and y_{4i}. Note especially the severe skewness pattern for the function Cost5.
To explore in additional depth the population distributions of unit-level relative costs, for each unit i we computed the ratios

$$r_{1i} = (Unit\ i\ cost)/(Mean\ cost)$$

$$r_{2i} = (Unit\ i\ cost)/(Median\ cost)$$

separately for each of the cost functions Cost1 through Cost4. Figures 1 and 2 display boxplots of the resulting distributions of r_{1i} and r_{2i} for industries B and C, respectively. The functions Cost1 and Cost 2 have similar distributions for r_{1i} and r_{2i}, while the corresponding distributions for Cost3 and cost 4 display stronger patterns of dispersion. The function Cost5 has a very pronounced pattern of dispersion, as one would expect for an exponential function, and thus is omitted from Figures 1 and 2.

3. Simulation Results

Separately for industries B and C, we carried out a series of simulation exercises, each based on 10,000 independent replications. Each case was based on probability-proportional-to-size sampling based on size measure "c" as defined in Powers and Eltinge (2014).

3.1. Realized Sample Sizes Under Cost-Based Caps

For each function Cost0 through Cost5, respectively, we computed the mean and median cost that would have been incurred in unequal-probability sampling with a fixed sample size $n = 10$. We then defined these mean and median values to be "cost caps," and then selected sample units sequentially with per-unit selection probabilities proportional to size until the specified cost cap was reached. The resulting mean sample sizes are displayed in the first two rows of Table 3. Note that for the functions Cost1 through Cost4, the mean realized sample sizes were relatively close to the nominal sample size of 10. Thus, for these cases the mean of the realized sample sizes is relatively insensitive to the choice of the cost function. In contrast with this, under the Cost5 function, the mean sample sizes are much larger than 10. We repeated this process for nominal sample sizes of 15, 20, 25, 30 and 50, with the resulting mean realized sample sizes presented in the remaining rows of Table 3; the sensitivity results are qualitatively similar to those noted for the nominal sample size of 10. Table 4 presents corresponding results for the median realized sample sizes. Note especially that when the median-based cost cap is used, the median realized sample sizes were substantially less than the nominal sample sizes for the Cost3 and Cost4 functions; and that the median realized sample sizes were exceptionally small under the Cost 5 function. Tables 5 and 6 present parallel results for industry C.

3.2. Properties of Ratio Estimators of Mean Wages Per Unit

Under the sample design with cost caps described in subsection 3.1, we also computed standard combined-ratio estimators of the per-establishment mean wages for quarters 2 and 4, respectively. These ratio estimators used the first-quarter employment count e_{1i} the auxiliary variable, and weights were adjusted to account for the fact that the realized sample sizes were random, due to the use of the mean- or median-based cost caps.

Table 7 displays the properties of the estimators for the population means of y_{2i} and y_{4i} under a mean-based cost cap and a nominal sample size of 10. The third through fifth columns display the simulation-based bias, standard error and root mean squared error of the ratio estimator, and the sixth column presents the ratio of the square of the bias, divided by the mean squared error. The final column reports the ratio

$$scale_{n,\text{cost}} = \frac{rootMSE(n,\text{cost})}{rootMSE\left(n = 50, \text{cost} = C_0\right)}$$

Note that use of a divisor based on the nominal sample size of 50 and the constant cost function Cost0 provides a basis for comparison of mean squared error results across cases with different nominal sample sizes and different cost functions. Note especially that the "scale" ratios are relatively constant across cases, except for being substantially smaller for the Cost5 case. Table 8 presents corresponding results for the case of a nominal sample size of 50, and Tables 9 and 10 report results for the same cases, but with the use of a median-based cap on costs. In addition, Tables 11 through 14 present parallel results for simulations based on sampling from industry C. For all of these cases, the contribution of the bias to mean squared error was relatively small, so it is appropriate to focus primary attention on trade-offs between cost and variance.

Finally, to explore the distribution of estimation errors in additional detail, Figure 3 presents side-by-side boxplots for the errors in the ratio estimator for the mean of y_{2i} in industry B based on a nominal sample size of 10. Separate boxplots are provided for each of the functions Cost0 through Cost4, and for mean- and median-based cost caps. Note that the boxplots are relatively similar, except for a notably skewed distribution of errors for the median-based cost cap using the function Cost4. Figure 4 presents the corresponding set of boxplots for simulations with a nominal sample size of 50; in this case, the error distribution for the median-based cap and Cost4 is somewhat less skewed than in Figure 3. Figures 5 and 6 present parallel results for sampling from industry C.

4. Discussion

This paper has presented a simulation-based evaluation of some properties of unequal-probability sampling subject to cost caps with unequal unit-level costs. For the cases considered, efficiency results were relatively insensitive to moderate variability in cost functions (as reflected in Cost1, Cost2 and Cost3), but displayed a substantial amount of sensitivity to more severe variability in costs (as reflected in Cost4 and Cost5).

One could consider a number of extensions for the current work. For example, one could consider more complex cost functions through extensions of cost structures considered previously for other survey settings, e.g., Groves (1989), Karr and Last (2006), and LaFlamme (2008). In addition, it would be of interest to study more adaptive forms of cost management, beyond the use of fixed unit-level costs considered here. One example of such adaptive work would be the responsive design approach considered by Groves and Heeringa (2006). For example, one could consider expansion of the cost model to account dynamically for paradata like initial signals of cooperation from a selected sample unit.

5. Acknowledgements and Disclaimer

The authors thank Mike Buso, Robert Clark, Rick Clayton, Rachel Harter, Dave Hiles, Anders Holmberg, Phil Kott, Chester Ponikowski, Ken Robertson, Terrance Savitsky, Avi Singh, Michael Sverchkov and Daniell Toth for helpful discussions of size measures, cost functions and estimation methods in unequal-probability sampling. The views expressed here are those of the authors and do not necessarily reflect the policies of the United States Bureau of Labor Statistics.

6. References

Cochran, W.G. (1977). *Sampling Techniques, Third Edition.* New York: Wiley.

Godambe, V.P. (1982). Estimation in Survey Sampling: Robustness and Optimality. *Journal of the American Statistical Association* **77**, 393-403.

Groves, R.M. (1989). *Survey Errors and Survey Costs.* New York: Wiley.

Groves, R. M. and Heeringa, S. G. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society, Series A*, **169**, 439-457.

Holmberg, A. and B. Swensson (2001). On Pareto πps Sampling: Reflections on Unequal Probability Sampling Strategies. *Theory of Stochastic Processes* **7**, 142-155.

Karr, A. F. and Last, M. (2006). Survey Costs: Workshop Report and White Paper. National Institute of Statistical Sciences: Survey Cost Workshop in Washington, DC. http://nisla05.niss.org/a_liates/surveycost200604/whitepaper200609-_nal.pdf

Kott, P.S. and J.T. Bailey (2000). The Theory and Practice of Maximal Brewer Selection with Poisson PRN Sampling. *Proceedings of the Second International Conference on Establishment Surveys,* 269-278.

LaFlamme, F. (2008). Understanding Survey Data Collection through the Analysis of Paradata at Statistics Canada. *Proceedings of the Survey Research Methods Section, American Statistical Association,* 4217-4224.

Powers, R.K. and J.L. Eltinge (2013). Properties of Some Sample Designs Based on Imperfect Frame Information. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Powers, R.K. and J.L. Eltinge (2014). Effects of Imperfect Unit Size Information on Complex Sample Designs and Estimators. *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Zangeneh, S.Z. and R.J.A. Little (2015). Bayesian Inference for the Finite Population Total from a Heteroscedastic Probability Proportional to Size Sample. *Journal of Survey Statistics and Methodology* **3**, 162-192.

Table 1: Population-Level Descriptive Statistics for Industry B

Variable	Mean	Std Dev	Skewness	p10	p25	Median	p75	p90
Cost0	1.00	0.00	.	1.00	1.00	1.00	1.00	1.00
Cost1	3.44	1.30	-0.1886	1.69	2.39	3.64	4.33	5.09
Cost2	5.12	2.60	1.1932	2.41	3.00	4.74	6.29	8.75
Cost3	24.74	30.84	2.4570	3.00	5.00	15.00	29.00	61.00
Cost4	1515.19	4061.19	4.1598	5.00	17.00	197.00	785.00	3601.00
Cost5	7.3956E75	2.8515E77	38.5681	8.39	55.60	1202605.28	1.4463E12	1.142E26
Quarter 2 Wages	125653.67	192247.29	2.8780	6393.00	21000.00	57285.50	131659.00	340778.00
Quarter 4 Wages	120505.91	187320.35	2.9644	5964.00	18600.00	54873.00	126071.00	321534.00

Table 2: Population-Level Descriptive Statistics for Industry C

Variable	Mean	Std Dev	Skewness	p10	p25	Median	P75	P90
Cost0	1.00	0.00	.	1.00	1.00	1.00	1.00	1.00
Cost1	3.28	1.10	-0.3262	1.69	2.39	3.40	4.14	4.66
Cost2	4.59	1.81	0.5500	2.41	3.00	4.32	5.80	7.24
Cost3	17.16	15.23	1.3514	3.00	5.00	12.00	24.00	40.00
Cost4	494.12	855.02	2.7170	5.00	17.00	122.00	530.00	1522.00
Cost5	4.4348E28	1.1795E30	37.2446	8.39	55.60	59875.14	9744803447.25	8.6593E16
Quarter 2 Wages	58939.01	60265.60	1.6003	6076.00	14470.00	38126.00	82311.00	147395.00
Quarter 4 Wages	60099.10	64452.56	2.1995	6000.00	14065.00	38535.00	82874.00	148296.00

Figure 1: Distribution of r1 and r2, Industry B

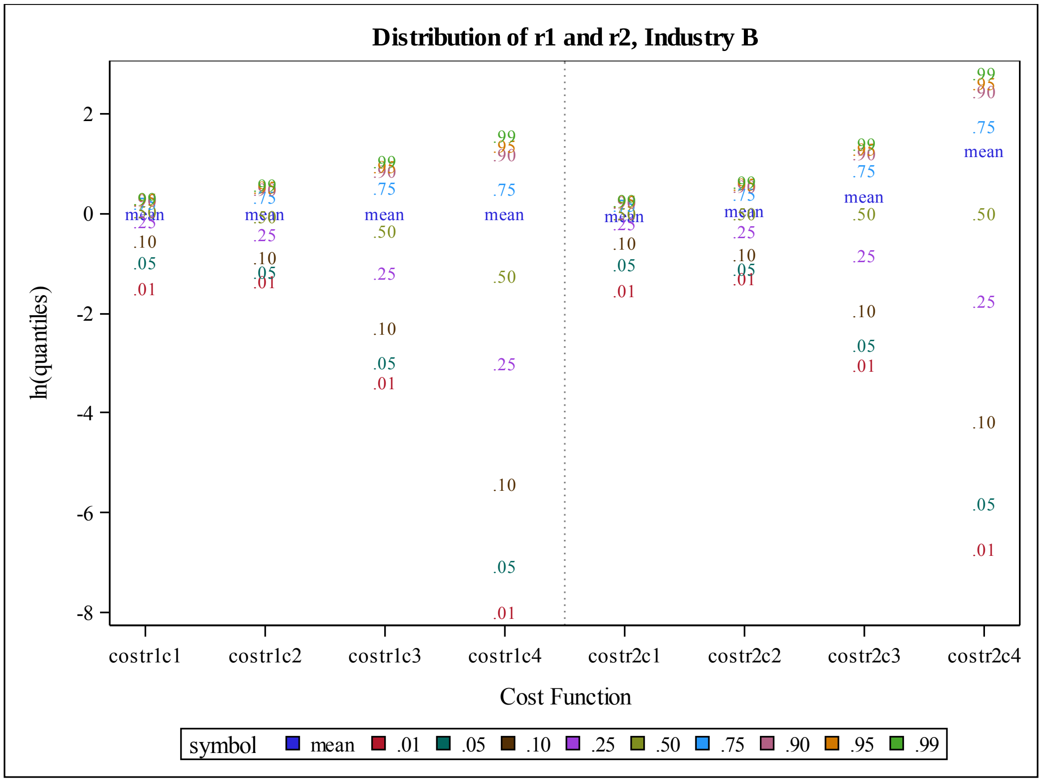

Figure 2: Distribution of r1 and r2, Industry C

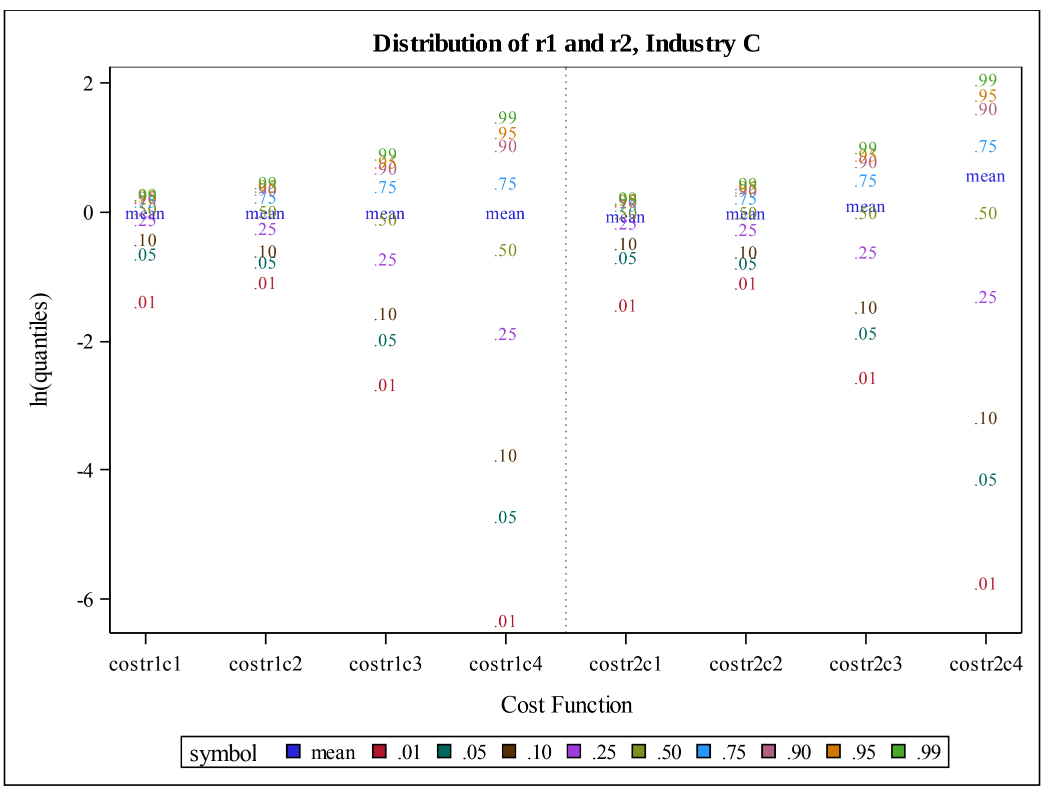

Table 3: Mean Sample Sizes for Industry B

Nominal sample size	Cost cap	Cost0	Cost1	Cost2	Cost3	Cost4	Cost5
10	mean	10	9.5805	9.7307	10.3377	11.6408	133.848
10	median	10	10.0051	9.1386	7.2235	3.7155	1.6120
15	mean	15	14.5942	14.7933	15.5723	17.2130	133.053
15	median	15	15.2420	13.9157	10.9230	5.2144	1.6155
20	mean	20	19.6111	19.8519	20.8281	22.7749	132.538
20	median	20	20.4628	18.6969	14.6525	6.7919	1.6179
25	mean	25	24.6286	24.9060	26.0536	28.3300	132.317
25	median	25	25.6994	23.4586	18.3271	8.3864	1.6619
30	mean	30	29.6642	30.0011	31.3289	33.9925	132.139
30	median	30	30.9521	28.2631	22.0380	10.0581	1.6594
50	mean	50	49.7532	50.2912	52.3953	56.4498	133.408
50	median	50	51.8972	47.3943	36.8003	16.4853	1.6561

Table 4: Median Sample Sizes for Industry B

Nominal sample size	Cost cap	Cost0	Cost1	Cost2	Cost3	Cost4	Cost5
10	mean	10	10	10	10	11	158
10	median	10	10	9	7	3	1
15	mean	15	15	15	16	17	155
15	median	15	15	14	11	4	1
20	mean	20	20	20	21	23	156
20	median	20	21	19	15	6	1
25	mean	25	25	25	27	28	156
25	median	25	26	24	19	8	1
30	mean	30	30	31	32	35	157
30	median	30	31	29	22	9	1
50	mean	50	50	52	55	59	158
50	median	50	53	49	38	16	1

Table 5: Mean Sample Sizes for Industry C

Nominal sample size	Cost cap	Cost0	Cost1	Cost2	Cost3	Cost4	Cost5
10	mean	10	9.5321	9.5570	9.7081	10.1688	77.819
10	median	10	10.1016	9.6944	8.7709	5.7447	1.6701
15	mean	15	14.5326	14.5592	14.6956	15.1613	80.258
15	mean	15	15.3854	14.7558	13.2702	8.5276	1.6976
20	mean	20	19.5107	19.5387	19.6952	20.1355	95.314
20	median	20	20.6511	19.8066	17.7837	11.3179	1.6828
25	mean	25	24.5280	24.5549	24.7020	25.1636	100.735
25	median	25	25.9414	24.8987	22.3081	14.1628	1.7353
30	mean	30	29.5236	29.5391	29.6918	30.1704	102.179
30	median	30	31.2106	29.9534	26.8101	17.0232	1.7678
50	mean	50	49.5196	49.5486	49.7420	50.2848	121.480
50	median	50	52.3525	50.2328	44.9051	28.1750	1.7593

Table 6: Median Sample Sizes for Industry C

Nominal sample size	cap	Cost0	Cost1	Cost2	Cost3	Cost4	Cost5
10	mean	10	9	10	10	10	62
10	median	10	10	10	9	5	1
15	mean	15	14	15	15	15	63
15	median	15	15	15	13	8	1
20	mean	20	19	19	20	20	84
20	median	20	21	20	18	11	1
25	mean	25	24	25	25	25	90
25	median	25	26	25	22	14	1
30	mean	30	29	29	30	30	92
30	median	30	31	30	27	17	1
50	mean	50	49	49	50	50	123.5
50	median	50	52	50	45	28	1

Table 7: Properties of Estimator for Mean Wages, Nominal n=10, Mean-Based Cap, Industry B

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	1679.66	19485.63	19557.89	.007375637	1.90193
cost1	2	1751.96	19840.97	19918.17	.007736568	1.93696
cost2	2	1770.52	19931.76	20010.24	.007828835	1.94592
cost3	2	1810.40	20198.91	20279.88	.007969306	1.97214
cost4	2	1923.30	20815.56	20904.23	.008464937	2.03285
cost5	2	609.33	14035.40	14048.62	.001881199	1.36617
cost0	4	1627.67	21486.18	21547.74	.005705937	2.06045
cost1	4	1644.52	21718.81	21780.98	.005700619	2.08276
cost2	4	1691.22	21859.02	21924.35	.005950435	2.09647
cost3	4	1670.96	21871.53	21935.26	.005802886	2.09751
cost4	4	1805.48	22533.98	22606.19	.006378666	2.16167
cost5	4	483.43	13495.20	13503.86	.001281585	1.29128

Table 8: Properties of Estimator for Mean Wages, Nominal n=50, Mean-Based Cap, Industry B

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	324.747	10278.07	10283.20	.000997317	1.00000
cost1	2	331.056	10328.08	10333.38	.001026403	1.00488
cost2	2	310.501	10321.49	10326.16	.000904169	1.00418
cost3	2	342.229	10415.86	10421.48	.001078389	1.01345
cost4	2	291.787	10419.29	10423.37	.000783639	1.01363
cost0	4	328.052	10452.62	10457.76	.000984026	1.00000
cost1	4	338.701	10459.89	10465.37	.001047426	1.00073
cost2	4	318.310	10457.15	10461.99	.000925706	1.00040
cost3	4	368.286	10518.87	10525.31	.001224334	1.00646
cost4	4	328.375	10485.86	10491.00	.000979731	1.00318
cost5	4	476.688	13577.09	13585.46	.001231176	1.29908

Table 9: Properties of Estimator for Mean Wages, Nominal n=10, Median-Based Cap, Industry B

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	1679.66	19485.63	19557.89	0.007376	1.9019
cost1	2	1702.37	19416.37	19490.86	0.007629	1.8954
cost2	2	1941.44	20629.92	20721.07	0.008779	2.0150
cost3	2	2529.70	24459.55	24590.01	0.010583	2.3913
cost4	2	8685.77	62841.36	63438.78	0.018746	6.1692
cost5	2	18761.11	110874.41	112450.50	0.027835	10.9354
cost0	4	1627.67	21486.18	21547.74	0.005706	2.0605
cost1	4	1598.04	21243.18	21303.20	0.005627	2.0371
cost2	4	1816.70	22530.25	22603.38	0.006460	2.1614
cost3	4	2487.30	26796.70	26911.89	0.008542	2.5734
cost4	4	9015.73	89596.80	90049.26	0.010024	8.6108
cost5	4	18263.95	131528.80	132790.80	0.018917	12.6978

Table 10: Properties of Estimator for Mean Wages, Nominal n=50, Median-Based Cap, Industry B

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	324.75	10278.07	10283.20	0.000997	1.0000
cost1	2	323.25	10205.12	10210.24	0.001002	0.9929
cost2	2	337.74	10518.06	10523.49	0.001030	1.0234
cost3	2	411.55	11624.62	11631.91	0.001252	1.1312
cost4	2	1095.78	17108.32	17143.38	0.004086	1.6671
cost5	2	16595.32	106853.01	108134.04	0.023553	10.5156
cost0	4	328.05	10452.62	10457.76	0.000984	1.0000
cost1	4	320.88	10304.29	10309.29	0.000969	0.9858
cost2	4	359.46	10689.07	10695.11	0.001130	1.0227
cost3	4	444.79	11987.66	11995.90	0.001375	1.1471
cost4	4	963.35	17919.67	17945.55	0.002882	1.7160
cost5	4	14787.37	109055.43	110053.41	0.018054	10.5236

Table 11: Properties of Estimator for Mean Wages, Nominal n=10, Mean-Based Cap, Industry C

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	786.048	8453.51	8489.98	.008572071	2.25926
cost1	2	762.754	8597.97	8631.73	.007808595	2.29699
cost2	2	773.714	8587.81	8622.59	.008051657	2.29456
cost3	2	752.098	8561.03	8594.00	.007658755	2.28695
cost4	2	778.278	8694.06	8728.82	.007949843	2.32282
cost5	2	522.112	9439.25	9453.68	.003050180	2.51571
cost0	4	810.885	10198.12	10230.30	.006282631	2.38195
cost1	4	805.560	10509.57	10540.40	.005840919	2.45415
cost2	4	822.865	10580.70	10612.64	.006011873	2.47097
cost3	4	763.455	10481.64	10509.41	.005277284	2.44693
cost4	4	753.815	10549.17	10576.07	.005080207	2.46245
cost5	4	427.453	7723.65	7735.47	.003053541	1.80107

`

Table 12: Properties of Estimator for Mean Wages, Nominal n=50, Mean-Based Cap, Industry C

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	126.133	3755.73	3757.85	.001126623	1.00000
cost1	2	135.637	3769.09	3771.53	.001293371	1.00364
cost2	2	130.745	3762.96	3765.23	.001205782	1.00196
cost3	2	133.153	3758.46	3760.82	.001253540	1.00079
cost4	2	139.809	3762.07	3764.66	.001379178	1.00181
cost5	2	134.286	5326.61	5328.31	.000635161	1.41791
cost0	4	160.317	4291.94	4294.93	.001393306	1.00000
cost1	4	165.951	4311.41	4314.60	.001479387	1.00458
cost2	4	160.100	4301.34	4304.32	.001383478	1.00219
cost3	4	170.989	4306.01	4309.40	.001574353	1.00337
cost4	4	169.634	4308.54	4311.88	.001547721	1.00395
cost5	4	90.478	5123.11	5123.91	.000311807	1.19301

Table13:Properties of Estimator for Mean Wages, Nominal n=10, Median-Based Cap, Industry C

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	786.05	8453.51	8489.98	0.008572	2.2593
cost1	2	779.66	8344.45	8380.79	0.008654	2.2302
cost2	2	779.85	8537.25	8572.79	0.008275	2.2813
cost3	2	862.58	9040.94	9082.00	0.009021	2.4168
cost4	2	1552.81	12157.74	12256.50	0.016051	3.2616
cost5	2	6760.38	41075.08	41627.69	0.026374	11.0775
cost0	4	810.89	10198.12	10230.30	0.006283	2.3819
cost1	4	801.24	10187.42	10218.88	0.006148	2.3793
cost2	4	821.48	10469.98	10502.16	0.006118	2.4452
cost3	4	864.80	11041.84	11075.65	0.006097	2.5788
cost4	4	1634.33	15588.76	15674.19	0.010872	3.6495
cost5	4	6990.78	51635.69	52106.77	0.018000	12.1322

Table14:Properties of Estimator for Mean Wages, Nominal n=50, Median-Based Cap, Industry C

Cost	Quarter	bias	stderr	rootMSE	biasratio	scale
cost0	2	126.13	3755.73	3757.85	0.001127	1.0000
cost1	2	116.81	3659.56	3661.42	0.001018	0.9743
cost2	2	129.11	3741.26	3743.49	0.001190	0.9962
cost3	2	151.06	3930.20	3933.10	0.001475	1.0466
cost4	2	278.05	5036.77	5044.44	0.003038	1.3424
cost5	2	7459.34	45245.84	45856.60	0.026460	12.2029
cost0	4	160.32	4291.94	4294.93	0.001393	1.0000
cost1	4	148.09	4173.52	4176.15	0.001258	0.9723
cost2	4	160.56	4273.59	4276.60	0.001409	0.9957
cost3	4	173.20	4501.92	4505.25	0.001478	1.0490
cost4	4	306.01	5683.98	5692.21	0.002890	1.3253
cost5	4	6651.14	44085.46	44584.36	0.022255	10.3807

Figure 3: Boxplot of the Distribution of Estimation Error by Cost, Industry B, n=10

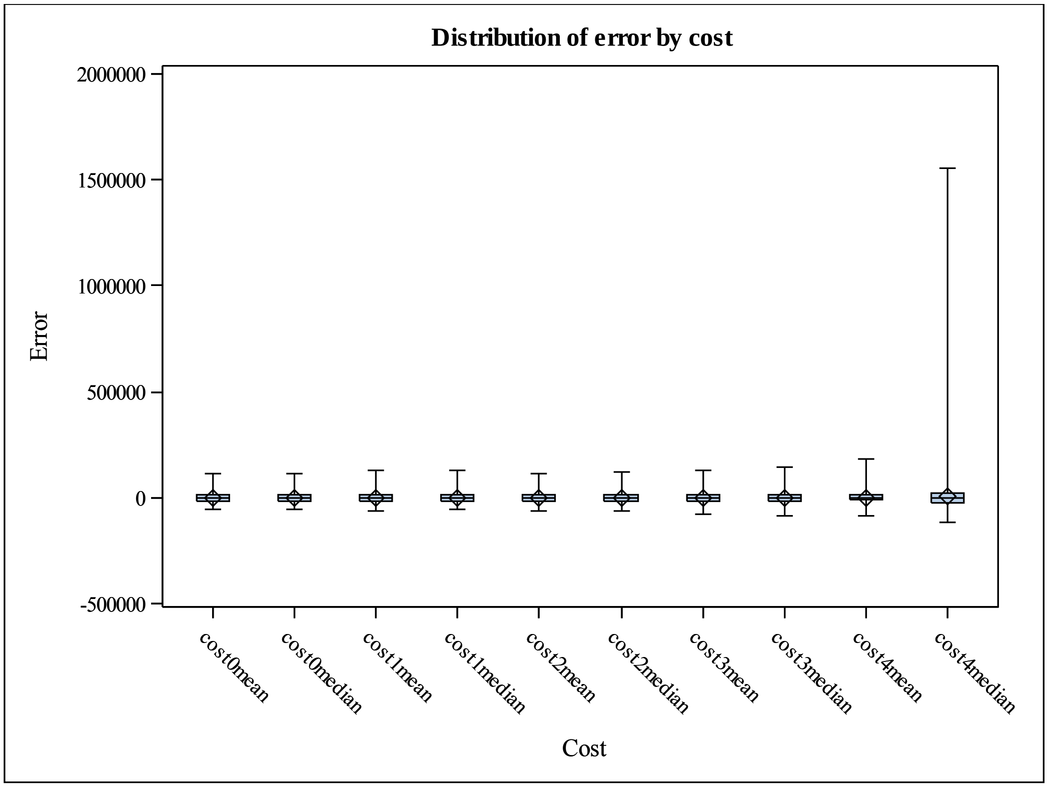

Figure 4: Boxplot of the Distribution of Estimation Error by Cost, Industry B,, n=50

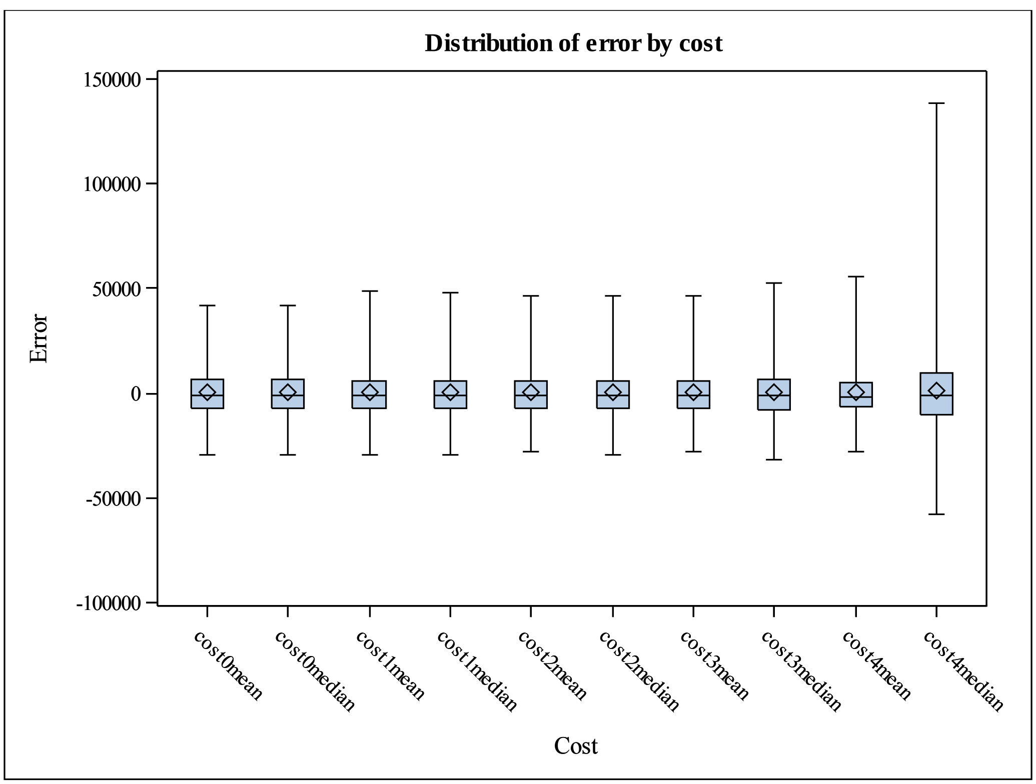

Figure 5: Boxplot of the Distribution of Estimation Error by Cost, Industry C, n=10

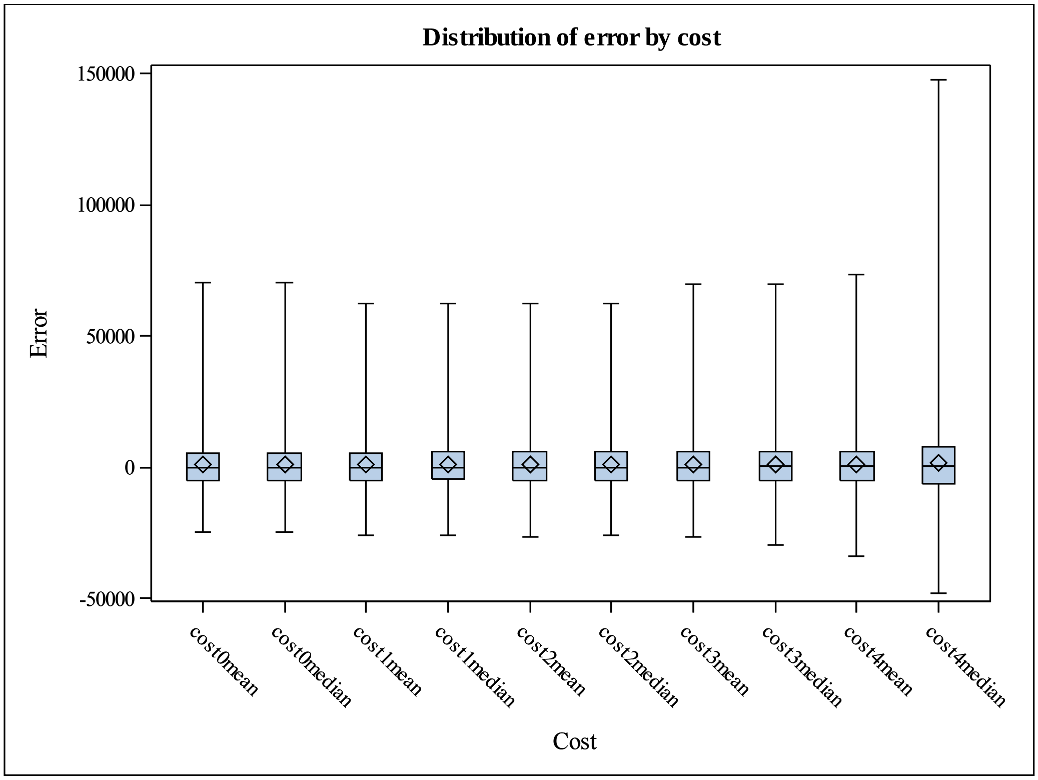

Figure 6: Boxplot of the Distribution of Estimation Error by Cost, Industry C, n=50