Revisiting Double Ratio Estimation for Managing Risk in High Rigor Evaluation

Justin Spencer, Dan Greenberg, Terese Decker, Navigant Consulting, Boulder, CO

ABSTRACT

Evaluation requirements for energy efficiency programs have gotten more rigorous in many jurisdictions, including forward capacity markets. PJM and ISO-NE markets both require a one-tailed 90% confidence with 10% precision, and metering of coincidence factors (PJM 2010). At the same time, the payments available in these markets incentivize utilities and evaluators to rigorously evaluate programs.

In ratio estimation, the evaluator measures the ratio between the measured values and some prior estimate of the values on a site-specific basis, rather than measuring the mean savings. Ratio estimation may be applied to many sampling problems where there is some prior estimate of savings available. The use of double ratio estimation requires that there is low cost data collection method available to create an initial estimate of savings. By nesting more rigorous onsite measurements inside a large sample of lower rigor data collection, such as billing data or phone survey results, a rigorous result can be achieved at lower cost. Evaluators facing highly rigorous evaluation requirements should incorporate double ratio estimation more often, in order to reduce risk of not meeting confidence and precision targets.

This paper discusses the application of this method in two cases. For the Maryland utilities' custom programs a double ratio estimation used a larger sample of phone-supported engineering reviews in the first stage, combined with onsite metering for a subsample. For Con Edison, the approach employed a large phone survey with a smaller nested sample with onsite metering to determine room air conditioner usage.

Introduction

The following paper explains how double ratio estimation can be used in evaluation of gross impacts. While this method is used by evaluators across the country, there are many more opportunities for using double ratio estimation and other double sampling methods. The paper first lays out the general methodology, including the equations used to derive savings and uncertainty estimates. After the general methodology, the paper steps through the methods, results, and a brief discussion of results for each of three studies where this method has been employed:

- EmPOWER Maryland Custom Evaluation
- Con Edison Residential HVAC Evaluation

In ratio estimation, instead of measuring the mean of an uncertain variable, the evaluator measures the ratio between the measured values and some prior estimate of the values on a site-specific basis. Ratio estimation is used widely in evaluation of custom projects, but may be applied to many other sampling problems where there is some prior estimate of the quantity of interest available. Ratio estimation is a valuable tool whenever these prior estimates show covariance with the verified results, i.e. the sites with higher prior estimates have higher verified results. A ratio estimation technique calculates a ratio between the verified savings for a sample and the prior estimate for the same sample. Ratio estimation is a subset regression estimation. See Lohr 2010 for more information about ratio regression estimation. Double ratio estimation builds on ratio estimation by performing a ratio estimation in two stages.

The use of the double ratio estimation technique depends on there being a lower cost way of collecting data that is indicative of the final results, but not accurate enough to use on its own. (Wright et al 1994) By nesting more rigorous onsite measurements inside a large sample of lower rigor data collection, such as billing data or phone survey results, a rigorous result can be achieved at lower cost.

Double ratio estimation is especially effective when large outliers (sites with realization rates much higher or much lower than one) may be the primary drivers of the overall results, provided the first stage can effectively find these large outliers. The large sample in the first stage effectively measures the frequency of the large outliers, while the second stage acts to calibrate the results of the first stage to a more accurate set of results for a subsample, while using a much smaller sample than would be required if only the second stage data collection techniques were being used. Evaluators facing highly rigorous evaluation requirements should incorporate double ratio estimation more often, in order to maximize value and rigor and reduce risk of not meeting confidence and precision targets. For example, in custom programs, the ultimate confidence and precision are highly dependent on the quality of the ex ante estimates, which can vary widely from project to project and even year to year, as program participation changes. There is always a risk that the actual CV[1] (Coeffient of Variation) will be significantly higher than was assumed in sample design.

Both ratio estimation and double ratio estimation have problems with prior estimates that are near zero or may have the opposite sign of the verified savings. When a multiplier is applied to a zero or negative number, the result is generally no more accurate than the original estimate. In these cases, other statistical options should be explored. The simplest alternative is to do statistics on the mean of the population. Another option is to use a different regression estimator, either with a multiplier and a constant adder, or just a constant adder.

Double Ratio Estimation Methodology

A useful metaphor for double ratio estimation is the process used for extracting gold from river sediments. Prospectors are trying to separate gold from a bunch of gravel and have multiple methods available to them. Some methods are cheaper but offer lower accuracy – there will be other objects of similar density extracted with the gold. Panning is the most accurate method, in that the gold can be extracted in a pure form, but the process is labor intensive. What modern prospectors do is to combine a first stage of sluicing with a second stage of panning. In the first stage, huge volumes of river sediment are pumped through a sluice, which separates everything that has a similar density to gold from the other contents. This process is very efficient at sorting through high volumes, but the results are not pure gold. In the second stage, the extracted high density materials are panned to separate the gold from the other high density material. The result is pure gold. In double ratio estimation for evaluation, the first stage of information extraction is to perform a set of file reviews or phone verifications, which can be performed inexpensively on a large sample. This is equivalent to sluicing the river sediments to get high density material. In the second stage, a more accurate method (like onsite metering) is performed on a nested subsample of sites. This is equivalent to the panning method, where the results are pure gold.

Calculating Realization Rates with Double Ratio Estimation

In a double ratio estimation, there are three sets of numbers being compared:

- x_{jh} is defined as the tracking data estimate for a given sample point *j* in stratum *h*
- y_{jh} is defined as the first stage (phone or file review) estimate of savings for a given sample point *j* in stratum *h*
- z_{jh} is defined as the second stage (on-site metering or verification) estimate of savings for a given sample point *j* in stratum *h*

[1] The CV measures the "spread" of a data set. It is defined as the ratio between the standard deviation and the mean of a data set.

A double ratio estimation calculates two ratios, between the first stage and tracking and between the second stage and first stage. In cases where the prior estimates contain zeros or negatives, it may be preferable to calculate statistics on the mean in the first stage, rather than use a ratio.[2] The same general double sampling method applies, except for the use of standard statistics on the first stage. The first stage realization rate for the sample point, measuring the realization rate between the tracking and phone/file review estimate, RR_{jh1} is then calculated:

$$RR_{jh1} = \frac{y_{jh}}{x_{jh}}$$

The second stage realization rate for the sample point, measuring the realization rate between the more rigorous M&V method and the phone/file review estimate, RR_{ih2} is then calculated:

$$RR_{jh2} = \frac{z_{jh}}{y_{jh}}$$

The overall sample point realization rate RR_{jh} is then calculated as the product of the two stages:

$$RR_{jh} = RR_{jh1} \times RR_{jh2}$$

The stratum first-stage sample realization rate of stratum h is the sum of all phone/file-verified ex post savings in the sample of stratum h divided by the sum of all tracked ex ante savings in the sample (n = *j*) of stratum h, given by:

$$RR_{h1} = \frac{\sum_1^j y_{jh}}{\sum_1^j x_{jh}}$$

In the second stage, only a subsample of the sites in the first stage sample are used.[3] The stratum second-stage sample realization rate of stratum h is the sum of all the second stage ex post savings in the onsite subsample of stratum h divided by the sum of all the first stage ex post savings in the onsite subsample (n = *i*) of stratum h, given by:

$$RR_{h2} = \frac{\sum_1^i z_{ih}}{\sum_1^i y_{ih}}$$

The overall stratum realization rate, RR_h, is then calculated as the product of first and second stage realization rates:

$$RR_h = RR_{h1} \times RR_{h2}$$

The verified total savings estimate for stratum h is the sum of all tracked ex ante estimates in stratum h multiplied by the stratum realization rate, given by:

$$TS_h = RR_h \times \sum x_{kh}$$

The verified total savings for the program is the sum of the total savings in the individual strata:

[2] In some cases, a different regression estimator of the form mx + b may be more appropriate. See Lohr 2010 for more information on regression estimators.

[3] There are *k* members of the population, *j* members out of *k* in the first stage phone/file review sample, and i members out of *j* in the second stage onsite sample.

$$TS_p = \sum TS_h$$

The overall realization rate for the program is then calculated by dividing the total verified savings by the total tracked savings:

$$RR_p = \frac{TS_p}{TS'_p}$$

Calculating Confidence and Precision with Double Ratio Estimation

In ratio estimation, an estimate for each member of stratum h can be made by multiplying the sample stratum realization rate by the prior estimate. A residual error can then be calculated for each sample point in stratum h by taking the difference between the ratio estimate and verified ex post savings for the point. In double ratio estimation, the first stage error at each sample point is calculated by taking the difference between the first stage verified savings and the first stage realization rate times the tracked value:

$$e_{1jh} = y_{jh} - RR_{h1} \times x_{jh}$$

The sample variance of the first stage verified total savings in stratum h is derived from the stratum first stage residuals:

$$V_{h1} = \frac{1}{n_{h1} - 1} \sum_{1}^{j} e_{1jh}^2$$

The first stage finite population correction factor for stratum h, FPC_{h1}, is calculated using N_h, the stratum population and n_{h1}, the first stage sample size:

$$FPC_{h1} = \sqrt{\frac{N_h - n_{h1}}{N_h - 1}}$$

The first stage standard error for stratum h, SE_{h1}, is calculated using:

$$SE_{h1} = FPC_{h1} \times \frac{\sqrt{V_{h1}}}{\sqrt{n_{h1}}} \times N_h$$

The first stage relative precision for stratum h, RP_{h1}, is then calculated using the first stage total savings, TS_{h1}, standard error, SE_{h1}, and t-value, t_1, based on the first stage sample size, n_{h1}:

$$RP_{h1} = t_1 \times \frac{SE_{h1}}{TS_h} \times 100\%$$

In the case where the first stage estimates a mean value, rather than a ratio, the statistics calculation for the first stage is exactly the same as above, except that the individual error terms are calculated using:

$$e_{1ih} = y_{ih} - \frac{\sum_i^{n_{h1}} y_{ih}}{n_{h1}}$$

The second stage error at each sample point is calculated by taking the difference between the second stage verified savings and first stage verified savings:

$$e_{2ih} = z_{ih} - RR_{h2} \times y_{ih}$$

The sample variance of the second stage verified total savings in stratum h is derived from the stratum second stage residuals:

$$V_{h2} = \frac{1}{n_{h2} - 1} \sum_{1}^{i} e_{2ih}^2$$

The second stage finite population correction factor for stratum h, FPC_{h2}, is calculated using N_h, the stratum population and n_{h1}, the first stage sample size:

$$FPC_{h2} = \sqrt{\frac{N_h - n_{h2}}{N_h - 1}}$$

The second stage standard error for stratum h, SE_{h2}, is calculated using:

$$SE_{h2} = FPC_{h2} \times \frac{\sqrt{V_{h2}}}{\sqrt{n_{h2}}} \times N_h$$

The second stage relative precision for stratum h, RP_{h2}, is then calculated using the second stage total savings, TS_{h2}, standard error, SE_{h2}, and t-value, t_2, based on the second stage sample size, n_{h2}:

$$RP_{h2} = t_2 \times \frac{SE_{h2}}{TS_h} \times 100\%$$

The overall relative precision for stratum h, $RP_{ht,}$ is then calculated as the square root of the sum of the squares of the relative precisions for the two stages:

$$RP_{ht} = \sqrt{{RP_{h1}}^2 + {RP_{h2}}^2}$$

The total standard error for stratum h, SE_{ht}, is then calculated using the first stage t-value, t_1, and the stratum total savings, TS_h:

$$SE_{ht} = \frac{RP_{ht} \times TS_h}{t_1}$$

The standard error on the total program, SE_p is given by:

$$SE_p = \sqrt{\sum_{h} {SE_{ht}}^2}$$

The relative precision on the total program, RP_t, is calculated using the program total standard error, savings, and t-value, based on the total sample size across all strata:

$$RP_t = t \times \frac{SE_p}{TS_p} \times 100\%$$

EmPOWER-Maryland Custom Evaluation

Five Maryland utilities offer their C&I customers incentives to implement custom energy efficiency measures that don't lend themselves to treatment through prescriptive rebate programs. The measures implemented through these programs include a wide range of technologies and project-specific savings estimates that span several orders of magnitude. Accurately estimating savings for these projects often involves highly complex analysis and extended data collection periods. Typically the expense of such an effort is excessive for the customer or contractor implementing the project, so simplified methods are often employed to develop ex-ante savings estimates, with widely varying accuracy.

The resulting wide range in the accuracy of ex-ante estimates has historically resulted in unusually high coefficients of variation (exceeding 1.1 in some cases) of the ratio between ex-post and ex-ante savings. High CVs, in turn, necessitate large sample sizes to achieve a desired confidence interval. Because two of the EmPOWER utilities intended to bid the demand reduction from their custom programs into the PJM market, an evaluation methodology including on-site M&V was required.[4] In order to resolve the conflict inherent between the very high cost of the large on-site M&V samples required by a conventional evaluation approach and the relative magnitude of savings provided by these programs (each program was expected to contribute less than 5 percent of each utility's portfolio-wide demand savings), Navigant decided to pursue a double-ratio estimation approach.

The low-rigor first phase of this approach consisted of reviewing project files and conducting telephone surveys for a relatively large sample. Utilizing CVs determined for the most recent prior evaluation year, Navigant developed stratified samples for this phase designed to achieve 80 percent confidence and 20 percent precision. Based on the assumption that the first stage of the evaluation would result in significantly improved estimates of actual savings and therefore greater uniformity in the ratio between second stage results and first stage results, Navigant assumed a CV of 0.6 in designing the on-site M&V sample for the second stage of the evaluation. The second stage of the evaluation consisted of developing and implementing customized monitoring plans for each sampled project, followed by analysis and interpretation of data collected over at least a six-week period during the utility and PJM peak periods.

EmPOWER-Maryland Custom Results

Table 1 below summarizes the demand savings results and uncertainties for utility A.

Table 1: EmPOWER Maryland Custom Program Demand Savings Results: Utility A

	Stratum 1	Stratum 2	Stratum 3	Overall
Tracked Savings (Population) [A]	461	468	477	1,407
Tracked Savings (First stage sample) [B]	461	468	37	966
1st stage savings (1st stage sample) [C]	437	653	39	1,130
1st stage realization rate [D = C/B]	0.95	1.40	1.08	1.17
1st stage savings (2nd stage sample) [E]	334	246	22	602
2nd stage savings [F]	195	280	64	540

[4] See PJM Manual 18b (PJM 2010).

	Stratum 1	Stratum 2	Stratum 3	Overall
2nd stage realization rate [G = F/E]	0.58	1.14	2.98	0.90
Overall Realization Rate [H = D X G]	0.55	1.59	3.22	1.05
Verified Savings [I = A X H]	255	742	1,537	2,535
First stage sample sizes,	3	12	7	22
Second stage sample sizes	2	3	4	9
Population	3	12	82	97
First stage coefficients of variation	0	0	0.13	
Second stage coefficients of variation	0.52	0.33	0.09	
First stage relative precision	0%	0%	6.6%	
Second stage relative precision	32.4%	28.4%	7.4%	
Overall relative precision	32.4%	28.4%	9.9%	10.2%
Overall Standard Error Squared	2,561	24,114	11,547	38,222

- Stage 1 identified and corrected gross methodology, calculation or data entry errors, resulting in far greater agreement between the stage 1 and stage 2 results than between stage 1 results and tracking system values.
- For Utility A, there was far greater agreement between stage 1 and stage 2 results, resulting in much smaller coefficients of variation for the ratios from each stratum than was the case for the first stage of the evaluation. This brought relative precision at the program level to within the desired precision target with only 8 site visits.
- For Utility A, the stage 1 evaluation (file review supported by telephone verification) was unable to identify significant methodological, calculation or data entry errors, so the stratum-specific realization rates were relatively close to 1.0 and CVs were small. The more rigorous data collection and analysis of stage 2 (on-site M&V) resulted in substantially greater coefficients of variation than had stage 1. Nonetheless, the methodology resulted in a final overall relative precision well within the target even with the small sample of only 9 projects that got on-site M&V. This allowed Navigant to achieve its evaluation goals within the evaluation budget available for this program.

Con Edison Residential HVAC Evaluation

Navigant recently evaluated Con Edison's Room Air Conditioner and Residential HVAC programs. The objective of the study was to determine an accurate estimate of runtime hours, energy savings, and peak demand reduction for air conditioning in New York City and Westchester County. In this evaluation, the evaluation team used the double ratio estimation method with an engineering-based phone survey paired with billing data disaggregation as the first stage and onsite metering as the second stage. Navigant used the phone survey to ask program participants about their air conditioning usage on certain types of days and correlated to varying outdoor temperatures. Because air conditioning runtime hours are particularly difficult to estimate or predict, in the second stage, the evaluation team metered air conditioning energy consumption for a sample of the phone surveyed participants. Participants were divided into two strata: high population density (Manhattan, Brooklyn, Bronx) and medium population density (Queens, Staten Island, Westchester). The evaluation team sorted the sampled sites in ascending order based on the phone-predicted runtimes for a typical meteorological year. The high population density stratum was divided into three substrata for low, medium, and high phone-predicted runtime. The medium population density stratum was divided into two substrata for low and high phone-predicted runtime. Ratio estimation works poorly on numbers that are zero or nearly zero. Therefore, a regression estimation adder – rather than the typical regression ratio estimation — was used for the phone-predicted low runtime substratum in each population density strata. The method of using an adder creates an adjusted estimate that accounts for consumers who claim not to use their room air conditioner at all, but in fact have a low but non-trivial runtime as indicated by the metered data.

Con Edison HVAC Results

Table 2 below summarizes the energy savings results and uncertainties for the Con Edison room air conditioner study.

Table 2: Con Edison Residential Room Air Conditioner Program Results

	High 1	High 2	High 3	Medium 1	Medium 2	Overall
Tracked Savings (Population) [A]		1,248			1,130	2,378
Tracked Savings (First stage sample) [B]	N/A	N/A	N/A	N/A	N/A	N/A
First Stage Runtime (1st stage sample) [C]	288	618	1368	371	1028	
1st Stage Realization Rate [D = C/B]	N/A	N/A	N/A	N/A	N/A	N/A
1st Stage Runtime (2nd stage sample) [E]	151	312	641	201	509	
2nd Stage Runtime [F]	240	465	288	192	321	
2nd Stage Realization Rate [G = F/E]	N/A	1.49	0.45	N/A	0.63	1.08
Overall Realization Rate [H = D X G]	N/A	N/A	N/A	N/A	N/A	N/A
Verified Average Runtime [I = G X C]	377	922	614	362	648	
Population		31,526			27,667	
First Stage Sample	32	30	33	49	41	185
Second Stage Sample	10	10	9	11	14	54
First Stage CV	0.65	0.23	0.34	0.75	0.38	
Second Stage CV	0.71	0.46	0.36	0.87	0.47	
First stage relative precision	20%	7%	10%	18%	10%	
Second stage relative precision	43%	28%	23%	50%	23%	
Overall relative precision	48%	29%	26%	53%	25%	15%

The results shown in Table 2 lead to the following conclusions about the use of double ratio estimation in this evaluation:

- The second stage CV was not nearly as low as had been assumed. This means that the first stage phone/billing result was not very accurate for room air conditioners. The benefits of using the double ratio method were small – 15% overall relative precision vs. 17-18% if only the onsite logged data was used.

- In the future, there are significant improvements that could be made for rigorous evaluation of room air conditioners. One option is to analyze first stage results in time to oversample higher usage strata. Another option is to survey people after the completion of the summer cooling season, instead of before, which would allow them to provide more accurate data about their usage for the room air conditioner in question. Yet another option is to drop the double ratio estimation method in favor of a larger traditional metering study.

Discussion and Lessons Learned

The two studies discussed in this paper tell two different stories about the use of double ratio estimation. Two-stage sampling methods do not always work – they depend on having a relatively accurate low-cost method available for the first stage. The first stage of the Con Edison Room AC study utilized phone surveys of participants, which turned out to be rather inaccurate. In contrast, the EmPOWER Maryland custom evaluation used engineering estimates combined with phone interviews to get surprisingly accurate results. Double ratio methods can also be used to validate the use of other lower-cost methods to assess program savings. The results of these two studies generally showed that the lower cost methods have questionable accuracy, meaning that they have large methodological error *when used alone*. As more double ratio estimations are performed, data may support further quantification of typical methodological error for certain low cost evaluation techniques.

In general, double ratio estimation reduces the risk of falling grossly short of confidence and precision targets when evaluating programs with uncertain CVs. The method is particularly applicable to evaluation of programs with custom calculation methods that have highly variable assumptions associated with them. Evaluators should be on the lookout for new low-cost data collection methods that might benefit from inclusion in a two-stage sampling approach. For example, increased availability of interval metering data should offer new opportunities for inexpensive analysis and adjustment of savings, but these methods will need to be calibrated by comparison to high rigor approaches. As an example, Navigant is in the midst of a double ratio metering study of furnaces in Illinois. The results of this study were not available in time to go to press, but the evaluators used customer billing disaggregation in the first stage, combined with furnace metering in the second stage of the double ratio estimation.

References

KEMA 2009. *Design 2000plus Lighting Hours of Use and Load Shapes Measurement Study.*

Lohr S 2010. *Sampling: Design and Analysis, 2nd Ed.*

PJM 2010. *Manual 18b: Measurement and Verification of Energy Efficiency Resources.*

Wright, R. L., Herowitz, M., Obtsfeld, I., and Butler, S. 1994. “Double Ratio Analysis: A New Tool for Cost-Effective Monitoring.” *In Proceedings of the ACEEE 1994 Summer Study on Energy Efficiency in Buildings,* Asilomar, Calif. American Council on Energy Efficiency.