

## Double Ratio Estimation: Friend or Foe?

*Jenna Bagnall-Reilly, West Hill Energy and Computing, Brattleboro, VT  
Kathryn Parlin, West Hill Energy and Computing, Brattleboro, VT*

### ABSTRACT

Double ratio estimation was introduced as a sampling strategy for evaluating energy efficiency programs in the 1990's, and is often used as a way to reduce evaluation costs while still producing reliable results. The two-stage double ratio approach involves using a less expensive method on a larger sample, followed by a more expensive and rigorous method on a smaller nested sample of the first stage projects. The underlying assumption is that the first stage approach (e.g., desk reviews) improves the estimate and the second stage (e.g., rigorous on site measurement) further refines the results.

This double ratio approach, with desk reviews and rigorous M&V, was recently incorporated into a draft evaluation framework as the only sampling method to be used for evaluating custom commercial and industrial (C&I) programs. The evaluation structure in a jurisdiction in the Northeast provided a unique opportunity to test this approach, as C&I custom projects are evaluated through both savings verification and rigorous M&V. Assessing the effectiveness of this approach may avoid issues with future evaluations.

The results suggest that the double ratio approach using desk reviews as the first stage is not an effective use of the method, as it could introduce bias and reduce the relative precision of the results. However, double ratio estimation is a tool that could be useful for other applications. The results of our analysis assist us with identifying situations that are not appropriate for double ratio estimation, and contribute to the discussion of how to properly apply this approach.

### Introduction

Double ratio estimation was introduced as a sampling strategy for evaluating energy efficiency programs in the 1990's as a way to reduce evaluation costs while still producing reliable results. This method uses a two-stage approach to develop a realization rate of energy savings. The first stage involves using a less expensive evaluation method on a larger sample, followed by a more expensive and more rigorous evaluation method on a smaller, nested sample of the first stage projects.

A recently developed, draft evaluation framework for a jurisdiction in the Northeast requires double ratio estimation for custom C&I programs. This document further specifies desk reviews as the first stage of the approach, followed by more rigorous M&V on a smaller nested sample. While evaluation methods for other sectors and programs were left open for the evaluator to propose, double ratio nested sampling is the only available strategy specified for C&I custom programs.

Impact evaluation of C&I custom projects is often the most expensive part of evaluating any portfolio. The wide variation in the size and complexity of projects requires a range of techniques, and expensive, on site measurement is commonly a component of the evaluation plan for many projects. Double ratio estimation has the potential to reduce evaluation costs and produce results with similar or better precision by using a relatively inexpensive method (e.g., desk reviews) on a large sample of projects and limiting more rigorous methods (e.g., on site metering) to a much smaller sample.

The question is whether the double ratio, nested sampling approach can consistently produce reliable results when evaluating custom C&I programs in comparison to conducting rigorous M&V on a larger sample. Previous research by others suggests that results are mixed (Spencer, Greenberg, and Decker 2013). Our team was in a unique position to test the actual impacts of using the double ratio

approach for custom C&I programs. Over the past eight years, we have been conducting separate evaluations for the same programs:

1. Savings verification, conducted within four months and using only desk reviews
2. Impact evaluation for the ISO-New England Forward Capacity Market (ISO-NE FCM), requiring rigorous M&V and completed over an 18-month period

Consequently, we had historical data to test the double ratio approach and compare the results to the impact evaluation that was based on a larger sample.

The effectiveness of the double ratio approach is based on a key underlying assumption: the first stage method improves the estimation of the item of interest, e.g., that desk reviews provide a more accurate estimation of energy savings. Achieving the benefits of the double ratio approach are dependent on the validity of this assumption.

As part of our review of the evaluation framework, we conducted an analysis using historical data to answer four main questions:

1. Do desk reviews improve the *ex ante* savings estimates?
2. Does the smaller sample size of projects with rigorous M&V introduce potential for bias?
3. How does the precision of this approach compare to using the more rigorous method on a larger sample?
4. What are the potential cost savings for the double ratio approach?

This paper covers the background of the use of the double ratio estimation method, a description of the evaluation activities, the scope of the analysis, and results and outcomes. It ends with a discussion about how this analysis provides support to previous conclusions and lends to a broader conversation on when double ratio estimation is most appropriate.

## Background

Ratio estimation takes advantage of an auxiliary variable, with a known value, that is correlated to a variable of interest to improve the precision of estimators of the mean or total population. Ratio estimation has been applied in the energy efficiency industry to improve the estimation of energy savings achieved by energy efficiency programs. Analyses produce *ex post* results that are compared to *ex ante* reported savings to establish a ratio (i.e., realization rate) that can be applied to the total population to improve the overall estimate of the program energy savings.

Double ratio estimation is a two-stage sampling approach with ratio estimation conducted at each level. In energy efficiency evaluation, this approach generally takes the following form:

1. Comparison of *ex post* to *ex ante* results using a large sample with a less expensive method of estimating *ex post* savings
2. More rigorous analysis on a smaller, nested sample of the first stage sample to develop an adjustment factor to the ratio estimated in the first stage

The ratio estimation approach works well for estimating the realization rate (RR), which is, by definition, the ratio of the *ex post* to *ex ante* savings.

Double ratio estimation has the potential to reduce evaluation costs and produce results with the same or better precision by using relatively inexpensive methods (e.g., desk reviews) on a large sample of projects and limiting more expensive methods (e.g., on site metering) to a much smaller sample. A number of assumptions must be met for this method to successfully cut evaluation costs while improving or maintaining the accuracy and precision of savings estimates. Previous studies

(Wright *et al.* 1994; Spencer, Greenberg, and Decker 2013) point towards the importance of the following assumptions:

- Relatively accurate, low cost method available for the first stage
- Large enough differentiation between per unit costs of the first stage and second stage methods (needed to produce cost-savings)
- Strong correlation between the results of the two different stages

Double ratio estimation is a two stage process. The RR for the desk reviews is calculated first, and then adjusted by the RR from the rigorous, on site M&V, as shown in equation (1):

$$RR = RR_1 \times RR_2 \quad (1)$$

Where  $RR_1$  is the realization rate from the first stage and  $RR_2$  is from the second stage.

The second stage RR is calculated using the evaluated savings from the first stage as the denominator with the evaluated savings from the second stage as the numerator. The relative precision captures the variations from both stages, as shown in equation (2).

$$RP = \sqrt{RP_1^2 + RP_2^2} \quad (2)$$

The overall RR is calculated as the product of the two stages (1). A more detailed review of the calculations is presented in “*Revisiting Double Ratio Estimation for Managing Risk in High Rigor Evaluation*,” Spencer, Greenberg, and Decker 2013.

The double ratio approach was initially proposed for energy efficiency program evaluation by Townsley and Wright in 1990. A brief summary of previous uses of the double ratio method is presented in Table 1.

Table 1: Previous Applications of Double Ratio Estimation to Energy Efficiency Program Evaluation

Application	Stage 1	Stage 2	Comments	Source
Initial proposal of method	Calibrated engineering models	End use metering	N/A	Townsley and Wright, 1990 (white paper)
PG&E Commercial, Industrial and Agricultural Rebate Program	Calibrated engineering models	End use metering	100% increase in the relative precision Possible reasons suggested: <ul style="list-style-type: none"> <li>• hard to obtain pre/post data</li> <li>• variability of in-project sampling</li> <li>• variability in estimation from monitored data</li> </ul>	Wright, <i>et. al.</i> , 1994
EmPOWER-Maryland custom C&I Projects	Desk reviews	Rigorous M&V	Designed for PJM FCM standards First stage results highly variable 2 <sup>nd</sup> stage sample sizes small (2 to 9) Overall precision 10.2% (met FCM)	Spencer, Greenberg, and Decker 2013
Consolidated Edison Room Air Conditioner	Phone survey/billing disaggregation	Onsite Measurement	Small improvement in relative precision Evaluators concluded no real advantage to double ratio	Spencer, Greenberg, and Decker 2013

## Description of Evaluations

The Vermont Department of Public Service (VDPS) is responsible for the evaluation of Efficiency Vermont's (EVT's) energy efficiency portfolio, as directed by the Vermont Public Service Board (PSB). Beginning with EVT's inception in 2000, the VDPS has conducted annual savings verification to determine whether EVT has met the goals as determined by the VDPS, PSB and EVT. In 2009, EVT bid its portfolio into the ISO-NE FCM, which requires a higher standard of energy savings verification. To meet the ISO-NE FCM standards, the VDPS undertakes a second round of rigorous impact evaluation over a 18-month period. The two evaluations are compared in Table 2.

Table 2: Comparison of ISO-NE FCM and Saving Verification (SV) evaluations

	SV	ISO-NE FCM
Purpose	Assess progress toward meeting goals	Meet ISO-NE FCM requirements
Method	Desk reviews	Rigorous M&V
Types of Savings Covered	kWh savings, winter and summer kW, MMBtu savings, TRB	Winter and summer kW, also kWh if all inputs are collected
Sampling/RR Method	Stratified ratio estimation	
Stratification	Upper level by program type/ Lower level by project size	
Largest stratum	Census	
Sample	Every third year, use same sample for FCM & SV	

## Savings Verification

Savings verification includes a paper review intended to identify errors in calculations, assumptions and methodology made by EVT in their savings claim. EVT's entire portfolio is included in the review, which covers the energy savings, demand savings, other fuel savings or extra use and all other inputs into the total resource benefit (TRB) calculation. Desk reviews are conducted for a sample of the custom C&I projects, and prescriptive savings are compared to the Technical Reference User Manual (TRM) for the rest of EVT's portfolio.<sup>1</sup>

Other common verification methods, such as direct measurement, participant surveys and on-site verification, cannot be carried out within the available time frame. The desk reviews for the custom C&I projects are limited to review of the project-level documentation and an assessment of errors in engineering methods or inputs. On a case-by-case basis, and time permitting, participant billing data may be reviewed for large retrofit projects.

<sup>1</sup> Efficiency Vermont *Technical Reference User Manual: Measure Savings Algorithms and Cost Assumptions*, TRM User Manual No. 2014-87, Effective Date: March 16, 2015

## Forward Capacity Market Evaluation

In order to participate in FCM, providers of energy efficiency resources must verify efficiency savings in compliance with the ISO-NE standards.<sup>2</sup> The primary purpose of the FCM impact evaluation is to determine the RRs for the winter and summer peak kW reduction. The impact analysis is designed to determine the kW reduction from the regional electrical system. In most cases, the energy savings (kWh) are also calculated as all of the necessary data has been collected.

Custom projects require M&V through one of four paths that are largely consistent with the International Performance Measurement and Verification Protocol (IPMPV). On-site measurement and/or utility interval data are used for project level analyses. Evaluators conducted the on-site measurements for small and medium projects, and EVT conducted the on-site measurement for the large projects under the direction of the DPS's Evaluation Team.

### Scope of the Analysis

In PY2013, the same sample of custom C&I projects were evaluated under both savings verification and FCM. Some projects were dropped for the FCM evaluation due to lack of participant cooperation or other logistical issues. However, both savings verification and FCM rigorous analyses were conducted for most of the sampled sites. This overlap provides the historical data required to conduct a comparison of the two-stage double ratio approach and single-stage rigorous M&V to analyze the potential impacts of the double ratio method.

Thirty-six of the PY2013 custom C&I projects had verified energy savings (kWh) for both savings verification and FCM. While the methods used to develop verified energy savings are different, as described above, both evaluations used stratified ratio estimation in calculating the RR's. These thirty-six projects were used to simulate the impacts of the double ratio approach as compared to rigorous M&V on a larger sample. The analysis has four main components:

1. Comparison of savings verification and FCM project-by-project RRs
2. Investigation of reasons for adjustments to the *ex ante* savings at the project level for the two different evaluation methods
3. Monte Carlo simulations assuming a larger desk review of all 36 projects for stage one and rigorous M&V on a smaller sample of 18 projects, evenly distributed through the strata, for stage two
4. Cost comparison of the per project costs for the desk reviews and rigorous M&V and total evaluation costs for different scenarios

The four topics are listed in Table 3 with the analysis approach for each.

Table 3: Evaluation Topics and Analysis Approach

Topic to be Tested	Analysis Approach
Effectiveness of desk reviews	Comparison of savings verification & FCM RR's Investigation into reasons for adjustments
Potential for bias	Investigation into reasons for adjustments Monte Carlo simulation

---

<sup>2</sup>ISO New England Manual for Measurement and Verification of Demand Reduction Value from Demand Resources Manual M-MVDR, Revision: 6, Effective Date: June 1, 2014

Compare precision	Comparison of savings verification & FCM RR's Monte Carlo simulation
Assess cost savings	Cost comparisons

### Comparison of Savings Verification and FCM Projects

This analysis focused on the overlapping 36 C&I custom projects from PY2013, as these projects had verified energy savings (kWh) under both savings verification and FCM evaluations. As the two evaluations incorporate distinctly different methods of analyzing energy savings, three hypotheses were developed to explain why the RRs are different:

1. Evaluators may have missed errors or overlooked corrections to energy savings calculations during the savings verification desk reviews.
2. Baseline assumptions may be different under savings verification (which often utilizes the Vermont TRM) and FCM (which is based on ISO-NE FCM standards).
3. On-site measurement or interval meter data analysis may provide critical information about the actual use of energy efficient measures, which was unknown at the time of the desk review.

The evaluation team compared project-level results across the two evaluations to determine why the project-level RR differed. Projects were grouped into the following four categories:

1. No savings verification adjustment
2. Similar savings verification & FCM adjustments
3. FCM adjustment much higher than savings verification
4. FCM adjustment much lower than savings verification

The RR's for these four groups were compared to assess the validity of our hypotheses and their effect on resulting RR's.

### Investigation into Reasons for Adjustments

To assess the differences in RR, the evaluation team reviewed reasons for individual project-level changes to energy savings and identified the drivers of the overall differences in RR. Based on this assessment the 36 projects were grouped into four mutually-exclusive groups according to the magnitude and direction of the energy savings adjustment. The four groups are defined in Table 4 below.

Table 4: Definition of Savings Verification (SV) and FCM Project Categories

Group	Number of Projects	Description	Purpose
A	16	No SV adjustment/ large FCM adjustment	Were adjustments missed in SV?
B	6	SV & FCM adjustments similar	Why were results similar?
C	5	FCM adjustment much higher than SV	Why was FCM higher?
D	9	FCM adjustment much lower than SV	Why was FCM lower?

## Monte Carlo Simulations

The double ratio approach with nested sampling assumed desk reviews of the 36 projects and rigorous M&V for 18 projects. Due to the relatively small sample size, the two upper-level strata, retrofit projects and new construction/market opportunity projects (NC/MOP), were combined. A Monte Carlo simulation was conducted for the second stage sample as described below.

- Eighteen projects were randomly selected without replacement from the 36 overlapping projects, and the RR, standard error and relative precision at the 80 percent confidence level were calculated for each run.
- The sample sizes for the nested sample were set for each stratum at approximately half of the number of overlapping projects and the random selection was conducted by stratum.
- One thousand simulations were conducted and results compared.
- Savings verification verified kWh was used as the denominator and the FCM verified as the numerator for the second stage RR.

As the precision for the savings verification analysis was good (3 percent relative precision at the 80 percent confidence level) and thus would make little difference to the overall precision, we simplified the calculations by including only the second stage results in the calculation of relative precision.

## Cost Comparisons

For comparison purposes, actual per project costs from the most recent evaluation cycles were used to estimate the overall evaluation costs using the double ratio approach that incorporates savings verification as the first stage and a second stage nested sample of rigorous M&V. This approach is compared to evaluating the total sample using the more rigorous FCM evaluation (FCM Only). This comparison provides an estimate of the cost savings that could be realized through using the double ratio approach.

## Results and Outcomes

The following sections outline the results from the four components of our analysis: comparison of savings verification and FCM RR's, our investigation into the savings verification and FCM reasons for adjustment, double ratio nested sampling simulation and cost comparison.

### Comparison of Savings Verification and FCM RR's

Table 5 demonstrates that the RR's from the two evaluation methods for the C&I custom projects were substantially different. For savings verification, the kWh RR for the overlapping projects was 96 percent +/-3 percent based on only desk reviews as described above, at the 80 percent confidence level. For FCM, the RR for the same sample of projects was 74 percent +/-14 percent.

The double ratio approach is based on the assumption that the desk reviews improve the accuracy of the savings, (*i.e.*, the verified savings from the desk reviews are closer to the actual savings than the program reported savings). To assess the validity of this assumption, we calculated the Pearson's correlation coefficient between the savings verification and FCM RR's at the project level. The correlation was -0.15, indicating a weak negative correlation. This means there is essentially no relationship between the savings verification and FCM RR's at the project level. These results bring into question this underlying assumption supporting the use of the double ratio approach.

Table 5: Comparison of Savings Verification and FCM RR's

Scenario	Sample Size	RR	RR Standard Error	Relative Precision at 80% Confidence
FCM Only	36	74%	0.079	14%
Savings Verification Only	36	96%	0.026	3%
Pearson's Correlation Coefficient for Project-Level RR's	-0.15			

### Investigation into Savings Verification and FCM Reasons for Adjustments

Although there was no relationship between the savings verification and FCM RR's, there still remains a question about whether desk reviews could be improved to be more effective. To investigate this possibility, we conducted an extensive review of the savings verification and FCM findings by project to try to understand why the desk review and rigorous M&V results varied.

This process showed that savings verification and FCM adjustments were made for completely different reasons, with the exception of one project. The differences in RRs at the project level were overwhelmingly due to the difference in availability of information gathered through direct measurement, on-site visits and participant interviews, as is shown in Table 6 below.

Table 6: Comparison of Reasons for Savings Verification and FCM Adjustments

Hypothesis	Number of Projects	Percent of Projects	Comments
Evaluators missed errors in desk reviews	4	11%	Two of these projects had additional adjustments due to on-site data collection in FCM.
Savings verification and FCM baselines were different	2	6%	The different baselines applied to only one measure in each of these two larger projects.
On site information critical to estimate savings	29	80%	These adjustments were made based on information that was not available for savings verification.
Savings verification and FCM results very similar	1	3%	Program reported savings were supported by FCM results.

These results suggest that it is most likely not possible to improve the accuracy of desk reviews without incorporating direct assessment of the operating conditions of the efficient equipment. This



finding is consistent with other studies (Frischmann and Kroll, 2012; Spencer, Greenberg, and Decker 2013).

### Simulation of Double Ratio Nested Sampling

Monte Carlo simulation modeling was conducted to assess the effects of the double ratio approach on both precision and bias. While previous papers have focused primarily on precision, small sample sizes in the second stage could also introduce bias. The RRs from the 1,000 Monte Carlo simulations range from 55 percent to 85 percent, with a mean of 72 percent, slightly lower than the FCM RR of 74 percent. The relative precision ranges from 17 percent to 29 percent, with a mean of about 19 percent, which is higher than the FCM value of 14 percent. A summary of the results is shown in Table 7 below.

Table 7: Summary of Simulation Double Ratio RR's and Relative Precision

Simulation Runs	RR	Relative Precision
Mean	0.72	0.19
Median	0.72	0.19
Minimum	0.55	0.17
Maximum	0.85	0.29
FCM Reference	0.74	0.14

Figure 1 below shows how the RR from the simulation runs compared to the FCM RR. This frequency chart is bimodal, with less than 30 percent of runs resulting in a RR within +/-2 percent of the FCM RR. While over 30 percent of the simulation runs produced RR's that were *more than 5.0 percent lower* than the FCM RR, less than 10 percent produced RR's that were *more than 5.0 percent higher*. This highly unsymmetrical result suggests that smaller second stage sample sizes could result in introducing bias to the RR's.

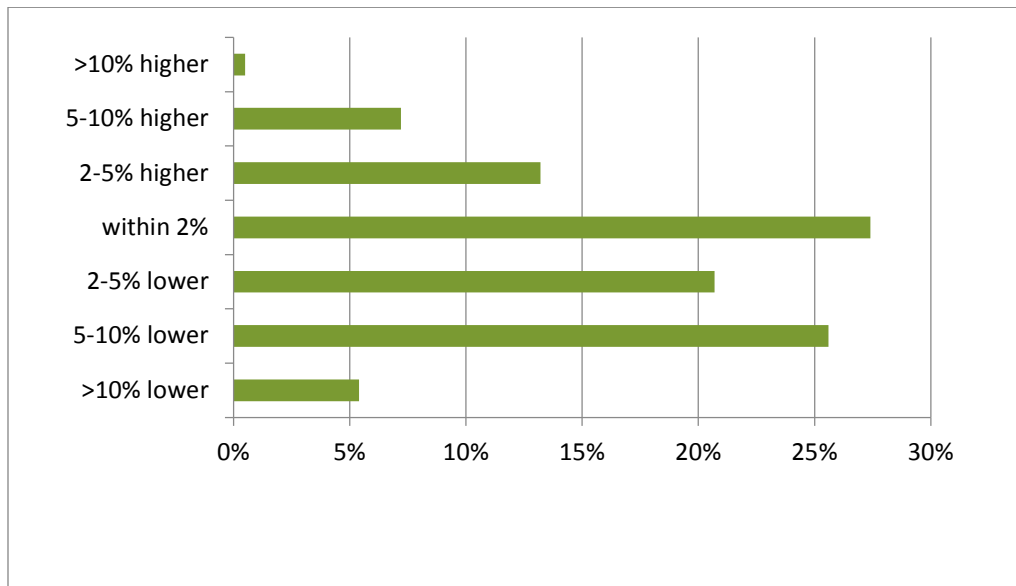


Figure 1: Results of Simulation of Double Ratio Nested Sample Approach

As this analysis suggests the double ratio method may not be an effective method for this application, we investigated further to try to understand why our analysis produced this result. It appears that the combination of the small sample sizes and high variability within the strata leads to this outcome. For example, one project within a stratum may have an RR at either extreme; in a larger sample, the impact of this project is mitigated. However, with a small second stage sample, the inclusion of this project could have a substantial impact on the results.

In the savings verification and FCM evaluations, a census of the largest projects are evaluated to reflect the high variability of these complex projects. Consequently, there is no sampling error for this stratum. Verifying a census of this stratum reduces the uncertainty of the evaluation as a whole. In the double ratio nested sample, half of these large projects were selected, which increases the sampling uncertainty and may also introduce bias if one or more of the largest projects have substantially higher or lower realized savings than the other projects.

This analysis indicates that the nested double ratio sampling approach can lead to both a loss in precision and introduction of bias to the results, at least for this application. This analysis also demonstrates that issues can arise with double ratio estimation when the methods used in the two stages do not capture adequate variability to improve precision. This is particularly problematic when desk reviews are used at the first stage method, as our results indicate this method does not the accuracy of savings estimates.

### Cost Comparison

Overall evaluation costs will vary according to the relative costs of the methods used for stage one and stage two, as well as the relative sample sizes. The sample size for the FCM evaluation was 68 projects. For the double ratio approach, one strategy would be to conduct desk reviews for 68 projects and rigorous M&V for 34 projects. Some recent studies using the double ratio approach have second stage sample sizes more in the range of a quarter to a third of the total number of desk reviews, however these smaller sample sizes also increase the potential for bias.

For comparison purposes the per project costs were used to estimate the overall evaluation costs using the double ratio approach that relies on desk reviews as the first stage and a second stage nested sample of rigorous M&V. This approach is compared to evaluating the total sample (68 projects) using the more rigorous M&V evaluation. The approximate costs of the two approaches are provided in Table 8 below. Average per-project costs for desk reviews and rigorous M&V reflect actual per-project costs during the most recent evaluation cycles.

Table 8: Results of Cost Analysis of Double Ratio Approach

Scenario	Sample Size	Per Project Cost	Total Component Cost	Total Evaluation Cost
<b>Rigorous M&amp;V (for comparison)</b>	68	\$7,000	\$476,000	<b>\$476,000</b>
<b>Double Ratio (with nested sample)</b>				<b>\$374,000</b>
First stage: desk reviews (large sample)	68	\$2,000	\$136,000	
Second stage: rigorous M&V (nested sample)	34	\$7,000	\$238,000	

This analysis suggests that, in this case, the double ratio approach would result in approximately 20 percent cost savings compared to the rigorous M&V, however this cost savings may potentially come with a loss of overall precision and the possibility of introducing bias. While it is certainly likely that the double ratio approach can reduce evaluation costs, the questions about the impacts on precision and bias remain.

## Discussion

The double ratio approach as proposed in the draft evaluation framework is based on the premise that the initial desk reviews improve the accuracy of the estimates of *ex-post* savings, which will then be further improved through rigorous M&V for a smaller sample. However, the results of our analysis of historical data do not support this underlying assumption, as the evidence suggests that desk reviews do not improve the accuracy of the estimates in comparison to using rigorous M&V.

Results shown in Table 1 suggest that there are two possible problems with using the nested sampling approach given the historical savings verification and FCM verifications:

- Loss of precision, with relative precision ranging from 17 percent to 29 percent, as compared to the 14 percent when the entire FCM sample is used.
- There appears to be a substantial likelihood of introducing bias, as 30 percent of the simulated double ratio runs produced RR's that were more than 5 percent *lower* than the RR from the whole M&V sample and less than 10 percent of the runs produce RR's that were more than 5 percent *higher* than the FCM RR.

Both the bias and precision issues indicate that this approach is not likely to meet the ISO-NE FCM standards.

The differences in RRs between savings verification and FCM at the project level were overwhelmingly due to the difference in availability of information gathered through direct measurement, utility interval data, on-site visits and participant interviews. These sources of information seem to be critical for understanding and addressing the key factors that affect the realization of energy savings for custom C&I projects. It does not appear that the results from desk reviews can be improved without direct contact with the site to determine how the efficient equipment is operated.

The double ratio approach is also based on the assumption that desk reviews are substantially less expensive than on site M&V. Our experience indicates that the detailed review of project files often requires back-and-forth with the program staff and understanding the program analysis involves a substantial time commitment, particularly for complicated projects. In contrast, rigorous M&V requires a solid understanding of the project but does not require a deep investigation of the methods and input used to estimate the savings, which saves some time and could potentially reduce the overall cost savings. Using actual costs for desk reviews and rigorous M&V, the potential cost savings may be in the range of 20 percent.<sup>3</sup>

## Conclusions

A recently cited advantage for use of double ratio approach is achieving reliable evaluation results at a lower cost through the use of less expensive desk reviews for a larger sample and more rigorous M&V for a smaller sample. Our test of these assumptions through the use of historical data from the savings verification and FCM evaluations suggests that the underlying rationale for the double ratio nested sampling needs to be re-visited. Some of the key issues arising from this analysis are as follows:

1. The assumption that desk reviews improve the accuracy of the savings estimates as compared to rigorous M&V is not supported by our analysis.

---

<sup>3</sup> Per-project cost estimates assume that similar staff would be utilized for both the desk reviews and rigorous M&V metering.

2. The small sample sizes used in the nested double ratio approach introduce the potential for bias, possibly of a substantial magnitude.
3. The potential cost savings may be smaller than anticipated and may not be worth the potential negative impacts on the evaluation results.

However, double ratio nested sampling may be a useful tool if we can construct a framework based on a less expensive evaluation method that actually improves the estimation of savings. The first stage verification approach needs to be equally rigorous; a double ratio approach relying on less rigorous desk reviews will not produce the intended outcome.

A less expensive but rigorous potential first stage verification method is billing and/or advance metering interface (AMI) data analysis. This approach allows all facilities or homes with sufficient billing records to be included in the first stage analysis. The second stage could incorporate on site measurement for a relatively small sample of sites. Using this approach, on site metering could be used to ensure the AMI data is correctly interpreted and improve the AMI analysis for sites that were not metered. Weather-dependent measures may be good candidates for this type of approach as the signal-to-noise ratio in the billing or AMI data is lower. We are currently working on this type of approach for a residential heat pump project.

## References

- ERS, Navigant Consulting Inc., Opinion Dynamics Corporation, and Apprise, May 2014. "Consolidated Edison Company of New York: Residential Electric HVAC Program Impact Evaluation Summary." Consolidated Edison Company of New York, 2014.
- Frischmann, Michael and R. Kroll, "On-site Measurement and Verification versus Project File Desk Review," International Energy Program Evaluation Conference, Rome, Italy, 2012.
- ISO New England Manual for Measurement and Verification of Demand Reduction Value from Demand Resources, Manual M-MVDR, Revision: 6 Effective Date: June 6, 2014.
- Lohr, Sharon L. *Sampling: Design and Analysis*. Duxbury Press, 1999.
- Spencer, J., Greenberg, D., Decker, T. "Revisiting Double Ratio Estimation for Managing Risk in High Rigor Evaluation," International Energy Program Evaluation Conference, Chicago, August, 2013.
- Townsley M., and R. L. Wright, "Measuring DSM Impacts: End-Use Metering and the Engineering Calibration Approach." *End-Use Information and its Role in DSM*, The Fleming Group, 1990.
- 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, CA. American Council on Energy Efficiency.