Who Wins the Pay-for-Performance Game? Measuring and Controlling Performance Risk

Hassan Shaban and Matt Golden, Open Energy Efficiency
David Tine, HSB

ABSTRACT

Recently, a combination of regulatory decisions and increased access to smart meter data that facilitates timely, cost-effective measurement of energy savings has created a surge of interest in pay-for-performance programs for energy efficiency. Pay-for-performance programs encompass a variety of program designs where financial incentives or payments, from a procurer or utility to an aggregator or customer, are tied to the quantity of delivered savings. More aggressive goals for energy and carbon savings coupled with declining returns from easy-to-implement efficiency measures, make pay-for-performance programs very attractive as a vehicle for shifting program performance risk from utilities and ratepayers to aggregators and even to participants (with certain program designs).

This paper presents an analysis of the sources of performance risk in pay-for-performance energy efficiency, how this risk is allocated among the different market participants and how these market actors can win the pay-for-performance game by measuring and controlling their risk and including it in their pricing models.

Introduction

Program administrators of pay-for-performance energy efficiency programs offer incentives or other monetary payments either directly to customers or to aggregators (entities that implement or procure energy efficiency in portfolios of buildings) in proportion to delivered energy savings (e.g. electricity, electric demand or gas). These savings are typically measured during a performance period that starts when the energy efficiency intervention is completed and may extend for several months or even several years following this date. This model is in contrast to deemed or custom energy efficiency programs, where incentives are calculated based on projected savings and are usually paid upfront, close to the date when the intervention is completed.

Pay-for-performance is not a new concept in energy efficiency. In fact, some versions of pay-for-performance have been implemented for the past 25 years; a comprehensive review of past and current pay-for-performance programs was presented in a recent NRDC report (NRDC 2017). However, there has been a recent surge of interest in the pay-for-performance approach arising from a confluence of a number of factors including market trends, regulatory changes, and technology, which include:

- Tightening codes and standards, and diminishing low-hanging opportunities for energy savings, coupled with increasingly aggressive energy savings goals have introduced opportunities for (and necessitated) innovative business models that remove barriers to participation in energy efficiency. By encouraging private investment in energy efficiency assets that come with stable cash flow returns, pay-for-performance offers a first step toward reducing or eliminating a primary barrier to implementing energy efficiency: upfront and transactive costs.

- Regulatory and policy changes have reinforced the need to value energy efficiency based on its ability to change consumption (California State Legislature 2015, State of New York Public Service Commission 2016). By extension, these policies have also called out performance and outcome-based incentives as a preferred construct for program and market design (California Public Utilities Commission 2018, Office of the Governor of the State of Oregon 2017).

- Technological advances have made distributed computing and data storage ubiquitous. Consequently, performance measurement (M&V) can be automated and repeated faster and cheaper than ever before, facilitating pay-for-performance for huge energy efficiency portfolios. Moreover, a huge influx of new "smart" technologies with high but untested potential for providing savings requires program models that allow for their inclusion without waiting for drawn out pilot evaluations. Pay-for-performance is one such model, since it reduces ratepayer and utility risk (as explained in later sections), allowing these technologies faster access to market (if they actually deliver energy savings).

Even with all of its promises of better savings outcomes, higher quality interventions, fairer allocation of incentives, and more satisfied customers and ratepayers, pay-for-performance is still viewed suspiciously, as it introduces new risks and re-allocates other risks among the different stakeholders. In short, pay-for-performance is a game-changer and the players are understandably nervous. But, why? As long as risk can be quantified, it can be priced. And perhaps more importantly, pay-for-performance places the risks squarely with the parties that are in the best position to measure and control them. This paper seeks to drill in to the components of performance risk in pay-for-performance energy efficiency and dissect how the various market players can handle them.

What is performance risk?

Energy efficiency projects are financed, incentivized and/or executed with the expectation of a certain level of savings – typically based on engineering estimates or deemed values (e.g. from past EM&V studies). Performance risk is meant to capture the probability of a given portfolio of energy efficiency projects achieving these projected savings, i.e. the probability of having a realization rate different from 100%. In pay-for-performance programs, performance risk may dissuade implementers or aggregators from bidding on programs and procurements, while for energy efficiency investors, performance risk may introduce unacceptable uncertainty on their expected returns. However, if these sources of risk can be quantified, then the *unit savings price* (e.g. $/kWh or $/therm) can be adjusted to account for the risk of procuring savings when pricing or bidding for pay-for-performance programs. If you can measure it, you can try to control it. And if you cannot control it, you can adjust your price to account for it.

Risk re-allocation with pay-for-performance

In non-pay-for-performance EE programs, participants receive an upfront incentive from program implementers. These incentives come from the program administrators (e.g. utilities) or ultimately from the ratepayers in ratepayer-funded programs. Even if an intervention does not yield the expected level of savings, participants keep their incentives and program implementers generally still receive their contractual program administration fees. The program administrator may not achieve their mandated savings goals, so they shoulder some of the performance risk, especially if there are penalties involved. Moreover, in some cases, they may be forced to procure other sources of energy if their conservation targets are not met. Many of the negative outcomes of unrealized savings also fall on the shoulders of the ratepayers, who collectively invested in energy efficiency interventions with the intention of deferring the higher costs of grid infrastructure, but ended up getting less than what they paid for.

With pay-for-performance, the risk equation is dramatically reconfigured. Ratepayers and program administrators assume almost none of the risk, since they are only paying for realized savings. Program participants may or may not be affected, depending on the program design and incentive structure. Customer-facing pay-for-performance usually comes with other problems, as any perceived risk to customer incentives may deter participation. Program implementers and aggregators assume almost all of the performance risk, especially if participants still receive fixed upfront incentives.

While neither of these options can completely eliminate the risks inherent in energy efficiency interventions, pay-for-performance does help align incentives with the stakeholders who hold the most power to influence outcomes. If aggregators are held accountable for the savings that they deliver, they will be incentivized to perform high-quality interventions, target customers with the largest potential to save energy, and adjust their business models based on measured risk. If these risks are quantified, then this uncertainty can be taken into account when setting the unit savings price. If risks are underestimated, then aggregators are incentivized to improve their analytical processes to stay in business. If risks are significantly overestimated by an aggregator, then their savings price will be correspondingly high, making them less competitive in open bidding for a procurement. The procurer is also responsible for making sure that any winning procurements satisfy key return on investment or cost-effectiveness metrics, in order to protect rate-payers. One problem with shifting most of the risk to aggregators is that they may already be operating on very thin profit margins and taking on additional risk may be unsustainable. This model may also initially favor large aggregators with the financial liquidity and analytical infrastructure to jump in head first. However, since almost all risks associated with energy efficiency performance can be quantified in one way or another, portfolio insurance and financing will help mitigate these risks for aggregators, large and small, and help level the playing field. A similar alignment of incentives with the right stakeholders holds for customer-facing pay-for-performance (performance-based incentives), especially for managed facilities (large commercial and industrial). Pay-for-performance incentives encourage operational practices that maximize energy efficiency savings, and by extension, performance payments.

Source of performance risk

In general, pay-for-performance risks can be categorized into *implementation risks* and *evaluation risks.* Implementation risks are related to the selection, implementation and ongoing support and maintenance of the energy efficiency interventions. Also part of this risk is accurately predicting the savings potential of the selected energy efficiency interventions. Evaluation risks are related to the measurement and verification of energy savings stemming from the interventions.

Table 1. Sources of performance risk

Implementation risks	Evaluation risks
Ex ante savings prediction uncertainty	Implementation variance in savings evaluation
Quality of interventions	Non-routine events
Customer participation	Data quality
	Weather

A core consideration for any energy efficiency program is the eligibility requirements, a set of guidelines that establish which buildings or customers are allowed to participate in the program. In non-pay-for-performance programs, the fact that an intervention was performed is usually sufficient to claim the associated savings. The number of participants is usually a big determinant of total program savings, so these eligibility requirements can be relatively lax. However, with pay-for-performance, screening potential participants for energy savings potential becomes much more critical to running a successful program, as is the quality of interventions. Implementation risks will always exist. However, focusing on quality can mitigate them, with the additional benefit of creating satisfied customers.

Another key question for pay-for-performance energy efficiency is how to evaluate and measure "performance", i.e. energy savings. Any number of measurement and verification (M&V) methods can be applied including metered savings, deemed savings, submetering, and building simulations. Metered savings methods such as IPMVP Option C (International Performance Measurement & Verification Protocol Committee, 2002), CalTRACK (CalTRACK Working Group 2017), and ASHRAE Guideline 14 (ASHRAE, 2014) are the generally preferred options, as they are widely accepted standards and provide savings results at the meter, making the energy savings of pay-for-performance programs more likely to be accepted as true grid resources; in other words, they paint a true picture of the impact of energy efficiency interventions. Additionally, when energy efficiency projects are bundled into portfolios, the uncertainty of metered savings can be dramatically reduced, especially if the embedded projects have been underwritten to a certain quality and performance standard. Other evaluation methods can be employed as fallbacks, when a building's energy data cannot be utilized to fit a valid baseline energy model. Different evaluation methods introduce different evaluation risks with pay-for-performance, some of which will be discussed in the following.

Several source of performance risk are discussed below in more detail:

Ex ante savings prediction uncertainty

Underlying most decisions to proceed with an energy efficiency intervention is an estimate of the potential return on investment (energy savings). These predicted savings can be obtained using a variety of methods (aggregating historical data, building simulations, engineering equations etc.) and every prediction method has a certain level of uncertainty stemming mainly from built-in assumptions.

This type of risk can be quantified by examining the historical performance of the savings prediction methods versus the performance evaluation method that will be employed in the program (usually referred to as backcasting). Backcasting is especially accessible for mature energy efficiency programs and measures with a long history in the market and sufficient collected data. However, it is based on the assumption that projects executed in the future can deliver similar performance as those executed in the past, which may not always hold true, if building operations or other factors that affect energy use evolve with time. For newer measures, available data may be scarce and the prediction uncertainty may be more difficult to estimate. High-uncertainty prediction methods should be improved or replaced, and if neither is possible, then this uncertainty must be built into the unit savings price (Mathew et al. 2004).

Quality of interventions

Different contractors have different levels of experience and workmanship, which can result in dramatically different savings outcomes, even when the same equipment is installed in very similar buildings. Buildings also have different underlying characteristics that make them good candidates for some interventions but not others.

Intervention quality can be dramatically improved by program implementers by putting in place robust QA/QC procedures that ensure that measures are installed correctly and in accordance with industry best practices. Additionally, energy audits and pre-intervention analytics can play an important role in selecting the most suitable interventions for each building or portfolio of buildings. Underwriting requirements from insurance or financial entities can support the quality control process to ensure a consistent level of performance across the portfolio.

Customer Participation

Participation levels and customer acquisition costs have a dramatic impact on aggregate program savings and getting them right is essential to accurate planning and program design. Aggregators are particularly vulnerable as they need to assume fixed overhead costs, so lower-than-expected participation can significantly affect their bottom line.

Historical program participation data can yield valuable insights about the potential interest in various types of energy efficiency interventions and can aid in the development of more realistic participation numbers, in advance of program launch. Real-time feedback on program performance can also be crucial in detecting any unforeseen trends and taking remedial action in a timely manner (e.g. shifting more budget dollars to marketing, if participation is lagging).

Implementation variance in savings evaluation

Methodologies for estimating project performance (savings) vary from one analyst to the next, or even from one year to the next, and present one of the largest sources of performance risk, and one of the easiest to eliminate. Based on prior experiences with CalTRACK methods developement process (CalTRACK working group 2017), seemingly minor differences in methodology could easily result in large differences in savings (and payments).

With any evaluation methods that form the basis for program payments, the entire analysis process from data cleaning through aggregating portfolio savings, including all methodological choices, must be fully documented and made transparent so that multiple parties with the same input data can be confident that they can independently reproduce the same savings results. Evaluation methods should be built into procurement contracts, such that aggregators and bidders know upfront how they will be evaluated. Any required methods updates can be applied at pre-defined timelines (e.g. every calendar year, similar to many Technical Resource Manual (TRM) updates for deemed savings).

Non-routine events

Unexpected changes in occupancy, equipment or other factors that influence a building's energy use may cause metered savings to deviate significantly from the projected savings for energy efficiency projects. Some of these changes may cause positive biases and other may cause negative biases, but it is expected that they will average out to zero for portfolios with a very large number of projects, unless there are other systemic effects in play (e.g. economic recession).

This type of risk can be quantified using statistical simulations to determine the portfolio size required to deliver a certain level of savings within an acceptable degree of uncertainty. This approach operates under the assumption that past projects are representative of future projects, and thus may be more suitable for more mature programs and measures. In general, increasing the portfolio size is a potential solution that can mitigate this risk. Alternatively, adding contractual clauses that define non-routine events and how they should be handled (e.g. alternative modeling approaches) can be used, especially for smaller portfolios.

Data Quality

A large fraction of energy use data is inherently noisy and models calibrated to this data typically suffer from poor predictive power. If project savings are low, they may be easily lost in the noise.

The noise in the data is usually expressed through model-fit metrics; for example, the Coefficient of Variation of Root Mean Squared Error (CVRMSE) is a commonly used metric for statistical models of baseline energy consumption. One way to mitigate this risk is by pre-screening candidate buildings on baseline model fit, which may limit the pool of potential participants but reduces risk for the portfolio as a whole. This method is suitable for applications where precision for an individual building is important (e.g. customer-facing pay-for-performance). An alternative method, for portfolio-based programs, is to increase the portfolio size, which enhances statistical power and enables the detection of smaller intervention effects (savings) in the data.

Weather

Metered savings for weather-sensitive measures depend on the temperature in the performance period. For example, savings from a cooling system measure may be significantly lower if the summer temperatures are relatively mild in the performance period. Moreover, if weather conditions experienced in the performance period are significantly different from those experienced in the baseline period, the baseline model may not accurately capture the counterfactual.

Both types of weather risk can be quantified based on historical weather observations, which are easily accessible and can inform of the possibility of anomalous weather conditions, either in the baseline or performance periods. There are also weather related financial products (e.g. derivatives) that could be utilized to mitigate the weather risk associated with the portfolio of projects.

Summary

Performance risk with energy efficiency is inescapable. Pay-for-performance does not create major new risks; instead, it reallocates the performance risk mainly to market players like program implementers, aggregators, insurance companies, and financial entities. This paper has identified the major sources of performance risk for pay-for-performance programs and metered savings and illustrates that all of these sources can be measured and accounted for when developing pricing models. Transparency, analytics and quantification of risks can help reduce total risk and ensure that the pay-for-performance game is a win-win-win for all parties involved.

References

NRDC. 2017. *Putting Your Money Where Your Meter Is*, NRDC Report R-16-09-A, January 2017.

California State Legislature. 2015. *California Assembly Bill 802 Energy Efficiency.* https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160AB802

State of New York Public Service Commission. 2016. *Case 14-M-0101 – Proceeding on Motion of the Commission in Regard to Reforming the Energy Vision*, Issued and Effective May 19, 2016.

California Public Utilities Commission. 2018. *Decision 18-01-004.* http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M205/K560/205560586.PDF

Office of the Governor of the State of Oregon. 2017. *Oregon Executive Order 17-20.* http://www.oregon.gov/gov/Documents/executive_orders/eo_17-20.pdf

Mathew, P., Kromer, S., Sezgen, O. and Meyers, S. 2004. *Actuarial pricing of energy efficiency projects: lessons foul and fair*. Energy Policy 33 (2005) pp. 1319-1328.

CalTRACK working group. 2017. CalTRACK Methods Development Issues. https://github.com/CalTRACK-2/caltrack/issues

ASHRAE. 2014. *Guideline 14-2014 - Measurement of Energy, Demand, and Water Savings.*

International Performance Measurement & Verification Protocol Committee. 2002. *International Performance Measurement & Verification Protocol - Concepts and Options for Determining Energy and Water Savings*.