NAVIGANT

One Market Street
Spear Tower, Suite 1200
San Francisco, CA 94105
415.356.7100 Phone

Memorandum

To: **Aaron Lu, CPUC**

From: **Surya Swamy/Amul Sathe/Greg Wikler, Navigant; Brad Rogers, Tierra Resource Consultants**

Date: **May 1, 2015**

Re: **2015 and Beyond Potential and Goals Model Calibration**

This memo describes the approach to calibrating the 2015 and Beyond Potential and Goals (PG) model and discusses related calibration issues including the necessity of calibration, the data basis of calibration, which parameters are adjusted, effects of calibration, and interpreting calibration. In addition, this brief is intended to highlight the irrelevance of an "uncalibrated" forecast in response to stakeholder requests, as well as the need to separate scenario development from calibration activities.

Overview

Forecasting is the inherently uncertain process of estimating future outcomes by applying a model to historic and current observations. As with all forecasts, the PG model results cannot be empirically validated a priori, as there is no future basis against which one can compare simulated versus actual results. Despite that all future estimates are untestable at the time they are made, forecasts can still warrant confidence when historic observations can be shown to reliably correspond with generally accepted theory and models.

Calibration provides both the forecaster and stakeholders with a degree of confidence that simulated results are reasonable and reliable. Calibration is intended to achieve three main purposes:

- Ground the model in actual market conditions and ensure the model reproduces historic program achievements
- Ensure a realistic starting point from which future projects are made
- Account for varying levels of market barriers across different types of technologies and end uses

The PG model is calibrated by reviewing portfolio data from 2006 up through 2012 to assess how the market has reacted to program offerings in the past. The Navigant team used ex-post EM&V data from 2006-2012 as the calibration data and also compared results to the 2013-2014 compliance filing data.

The calibration data are used to inform the appropriate values for the customer willingness and awareness parameters that drive measure adoption during the model time horizon. These parameters are then considered to account for the range of factors—technological, economic, market, and program factors— that contribute to historic program achievements. This includes consumers' awareness of programs and their willingness to participate in them.

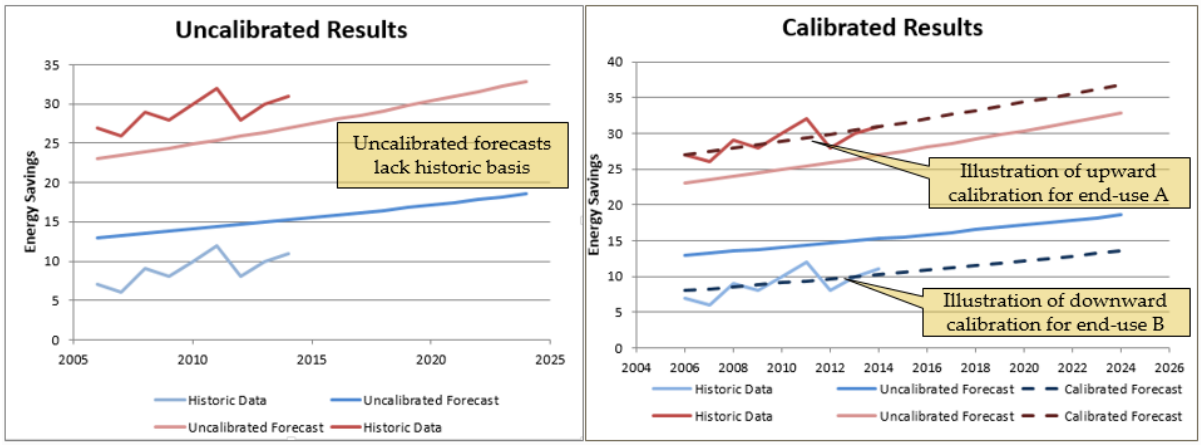
This calibration method (a) tracks what measures have been installed or planned for installation over an historic six-year period and (b) forecasts how remaining stocks of equipment will be upgraded, including the influence of various factors such as new codes and standards, emerging technologies, or new delivery mechanisms. The calibration approach is not applied to emerging technologies, as there is insufficient historical basis to adjust future adoption for these technologies.

Necessity of Calibration

Calibration refers to the standard process of adjusting model parameters such that model results align with observed data. In evaluative statistical models, calibration is called *regression*, and goodness of fit is typically the main focus since the models are usually simple. In situations of complex dynamics and non-linearity (as in this study), model sophistication and adequacy can become the main focus. But grounding the model in observation remains equally necessary. The ability of a forecast to reasonably simulate observed data affords credibility and confidence to forecast estimates.

Although there are data supporting all underlying parameters in the PG model, much of the data are at an aggregate level that can be inadequate to forecast differences across the various classes of technologies and end uses. The customer willingness-to-adopt factor is a good example of this effect. Customers may exhibit certain average purchase tendencies in adopting measures based on their financial characteristics. However there may be features of certain end use technologies that cause customer behavior to vary from the average. Residential building envelope is an end use where adoption of measures like insulation is consistently lower than would be predicted compared with other end uses. Residential lighting adoption, on the other hand, performs better than the average predicted customer purchase tendencies, even after adjusting for differences in financial attractiveness. We often think of these differences as the influence of non-financial product attributes or of market barriers.

The charts below illustrate the concept of calibration. The chart on the left below shows how certain end uses may over predict (blue) or under predict (red) adoption compared to observations of program participation. By adjusting the customer willingness factors, as illustrated in the right chart below, the modeled results in past years become aligned with reported historical program achievements.



Note that model parameters and results may be increased *or* decreased depending on the end use. We do not "calibrate down" on aggregate, but rather just "calibrate" the end uses both up and down as appropriate based on the data, as shown in the chart on the right above.

Calibration is not an optional exercise in modeling. One might suggest that the average customer data should be sufficient to make a reliable aggregated forecast. However there are two important non-linearities that compel us toward a more granular parameterization:

- Program portfolios are not evenly composed across end-uses. This leads to an uneven weighting issue whereby average customer willingness may not lead to the correct calculation of total savings.
- The dynamics in the model regarding the timing of adoption can become incompatible with the remaining potential indicated by program achievements. For example, if the forecast results were not calibrated for CFL lighting in the residential sector, the saturation may remain inaccurately low in early years and indicate a larger remaining potential in future years. Thus calibrating a willingness parameter upward may increase its potential in the early years but decrease its potential in later years. This implies that in the absence of IOU program intervention, residential CFLs would have historically had much lower adoption. Calibration therefore allows us to capture these program influences to more accurately reflect remaining potential.

This discussion is intended to highlight the necessity of calibration and the effective irrelevance of uncalibrated parameters. It may be tempting to "relax" the calibrated parameters back toward the average to measure the effect of what could be possible. But the uncalibrated results can be difficult to interpret and almost certainly would not produce feasible results for certain end uses. Thus they provide no basis for a reasonable forecast. Instead, we treat the calibrated results as the most basic set of interpretable results from which alternate scenarios are developed. Changes to calibrated parameters are not returned to the uncalibrated averages, but are rather explicitly developed based on the feasibility of values that parameters might take over time and how quickly the change might occur. This is discussed more in the last section of this brief.

Interpreting Calibration

Calibration can constrain market potential for certain end uses when aligning model results with past IOU energy efficiency portfolio accomplishments. Although calibration provides a reasonable historic basis for estimating future market potential, past program achievements may not capture the potential due to structural changes in future programs or changes in consumer values. Calibration can be viewed as holding constant certain factors that might otherwise change future program potential, such as:

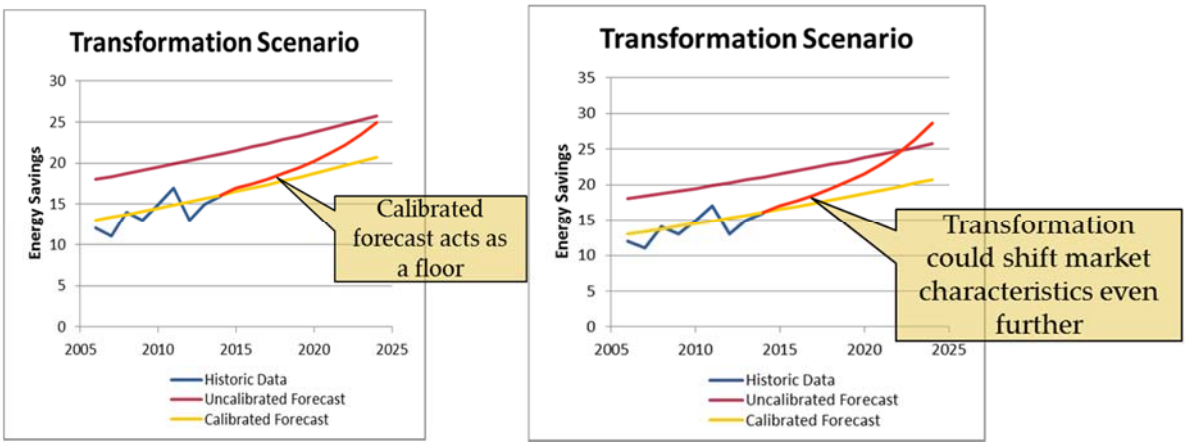
- Consumer values and attitudes toward energy efficient measures
- Market barriers associated with different end uses
- Program efficacy in delivering measures
- Program spending constraints and priorities

Changing values and shifting program characteristics would likely cause deviations from market potential estimates calibrated to past program achievements.

Does calibrating to historic data constrain the future forecast? In a strictly numeric sense, yes. If a certain end use is calibrated downward or upward, then future adoption and its timing are affected. However this should not be interpreted as "calibration constrains the level of adoption that we think is possible." Rather calibration provides a more accurate estimate of the current state of customer willingness, market barriers, program characteristics and remaining adoption potential. One forecast scenario might assume that the underlying conditions remain the same—a sort of business as usual scenario. We might develop another scenario such that it represents a transforming market based on agreed-upon end state parameter values appropriate for the end use market. For insulation that may mean a slight improvement, for water heating a greater improvement, and for lighting perhaps little change is warranted if fewer market barriers exist today.

One interpretation is that the calibration process creates a floor for the remaining potential. Market barriers, customer attitudes, and program efficacy generally move in the direction of improvement. The extent to which a market or program can improve should not be compared to the uncalibrated results, but rather to the vision for what is reasonably possible for the paramters describing each end use. This may require little change, some change, or greater change in parameter values for different end-uses. But improvements to parameter values are based on their own merits and feasibility, and are independent of the uncalibrated paramter values and results.

The charts below show two illustrative end uses where there is a calibrated base scenario (yellow) and alternative high scenarios (red) that are independent of the uncalibrated numbers (dark red). The chart on the left below shows a high forecast that may increase but still not meet the uncalibrated forecast, while the chart on the right shows a high forecast that exceeds the adoption of the uncalibrated forecast. The relation to the uncalibrated forecast is effectively arbitrary.



Implementing Calibration

Calibration examines three types of parameters to best align results with past program achievements:

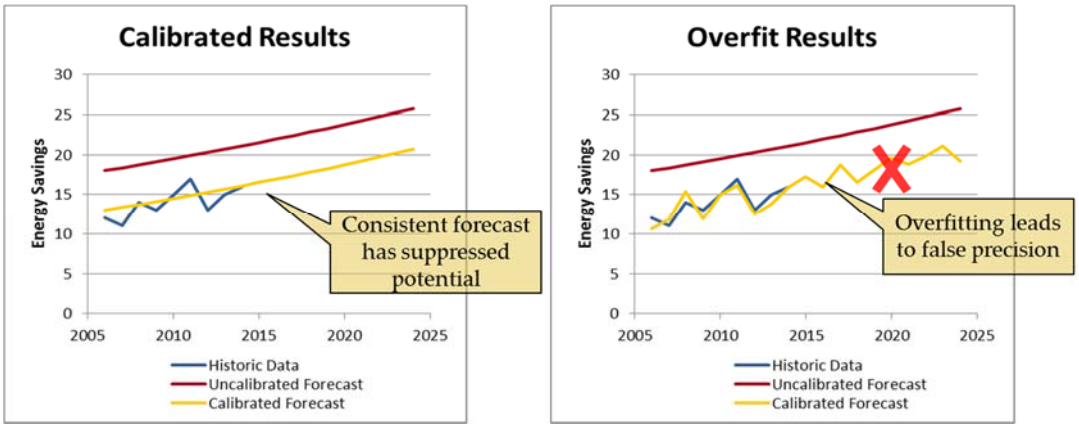
- *Willingness parameters*
  - Primary target of calibration,
  - *Implied Discount Rate* – the iDR is adjusted when perceived market barriers are higher or lower than typical measures, or when factors other than financial characteristics may play a larger role in purchase decisions,
  - *Sensitivity* – the consumer sensitivity to the differences in financial attractiveness is adjusted when markets are considered mature and customer primary focus is measure financial attractiveness.
- *Awareness parameters*
  - Sometimes used, but only after willingness,
  - Results are generally insensitive to awareness factors when measures are replace on burnout (RoB) with a measure life greater than 5 years because stock turnover dominates the timing,
  - *Word of mouth and marketing factors* - For retrofit and short-lived measures awareness can be adjusted to better fit the timing of market growth.
- *Initial awareness*
  - Less influential, but frequently used to align the curvature of the adoption with 2013 market saturation data.
  - Used to align the curvature of adoption timing with the estimated willingness and starting saturations.

Parameters are adjusted to fit historic observations during the calibration period. Then the parameters are applied to the forecast period, which begins in the year of most recent density data vintage. Calibrating parameters up and down can have different effects in a dynamic model depending on the initial saturation (i.e., density) data. For example, calibrating up can increase both historic and future adoption if the initial saturation is low. If initial saturation is high, then calibrating up can increase past adoption in the model leaving less for future years.

Once the consumer preference parameters are calibrated, the model forecast begins in 2013 by applying known market saturation data of that same vintage. Forecasts indicate the saturation of measures over time under the expected IOU future program influences.

Granularity of Calibration

The calibration process is undertaken at the sector and end use level for program activity in years 2006 to 2012.[1] The calibration accordingly accounts for the *cumulative* effect of market and program activity during these years. In our experience, this level is sufficient to capture the major differences in customer attitudes at the sector and end use level and to produce stable, reliable results over the forecast period. Overfitting the data can produce erratic model behavior that is beyond the precision of the forecast and the data that we use.



The data used for calibration are the ex post, gross evaluated program data. These data have units of energy savings such as MWh and Therms saved. By adjusting consumer preference parameters we can align the adoption and savings forecast over the calibration period with the actual evaluation data. This alignment is used by adjusting the consumer parameters for each sector, utility, and end use. The model is not calibrated at the building type or measure level for three reasons:

- The gain in precision of the results from calibrating at a lower level is expected to be negligible owing to the precision of the data sources for non-calibrated model inputs (e.g., density, building stocks, and calibration data)
- Calibrating at the lowest level of the model may give an appearance of rigor. But it is unlikely that customer preferences are represented by such sophisticated and highly dimensional reasoning. In other words, a highly granular model of consumer preferences would be at odds with the relative simplicity of the reasoning that consumers apply when making a purchase decision.
- Optimizing the non-linear model at the measure and building type level is a computationally intractable task that would require division into many batches—an enormously work- and time-intensive task due to the complexity of the model. It is not clear that such a path would lead to more accurate results and indeed might take away valuable resources from completing other aspects of the study scope,

[1] Evaluation ex-post gross data were used for 2006-2012 from the CA Standard Program Tracking Database

The end use/sector/multiyear level of calibration was chosen because:

- The model variance is mostly explained at the sector and end use level making this level adequate to account for the most influential non-linear effects,
- The precision of lower level calibration results is not significantly improved beyond the chosen level,
- It is unlikely that in deciding to adopt a measure, consumers show very different purchase behavior toward similar technologies,
- Individual year calibration data are too noisy and inconsistent to fit and may lead to unreliable predictions.
- The chosen level of calibration strikes the right balance of analytical benefit versus cost.

Calibration of the PG model is performed at the back end of the modeling process in that input willingness and awareness parameters are iteratively (and manually) adjusted in the back end of the model until alignment is reached with ex post, gross evaluated data program data over the calibration period. The manual nature of this iterative task results in a lengthy process that requires repeatedly running the model, one sector and IOU at a time, to calibrate at the end-use level.

Scenario Analyses

This section offers an auxiliary discussion about scenario analyses not directly related to the process of calibration but brought up by stakeholders in relation to discussions about calibration.

Explicit Scenarios

Calibrated parameters provide the starting point for interpretable quantitative results. Scenarios are developed as explicit modifications to key variables the calibrated forecast such that the results can be easily interpreted. Multiple key variables can be changed in the calibrated forecast to produce results under different scenarios. These key variable fall under two categories:

1. Exogenous variables (events and outcomes that cannot be influenced)
2. Endogenous variables (events and outcomes that can be influenced)

Disentanglement of parameter uncertainty from policy and program levers in scenarios

One factor that has obfuscated the interpretation of scenarios in the 2013 study is the combination of exogenous parameter uncertainty (e.g., retail rates, building stocks, technology curves, etc.) with the endogenous variables that may be influenced by policy and program implementation (e.g., measure inclusion criteria, codes and standards, variable incentive levels, or market transformation activities). This conflation of exogenous and controllable parameters within the scenarios made them difficult to interpret. Separation of exogenous parameter uncertainty from parameters that may be influenced or controlled will help disambiguate the meaning of the scenarios.

Navigant believes it is important to consider the effects of exogenous parameter estimates as a statement about the range of uncertainty stemming from several important factors that are beyond stakeholder's control--an effective uncertainty band. Then other parameters that represent the influence of policy and program decisions might be used to estimate credible increases in adoption, beyond the base calibrated results that might be achieved.

Maximum Achievable Potential

In previous discussions, some stakeholders have expressed a desire to use estimates of economic potential to convey the upper bound of what is possible. Although economic potential has a financial basis, it does not have a market basis. In particular, economic potential has no consideration of customer preferences nor does it account for the turnover of stock and the time scale of diffusion for different classes of technologies. For instance, future potential for ROB and long-lived measures generally are constrained by stock turnover rates which is not captured within economic potential. This leaves a disconnect and a gap between economic potential and the upper bound of what could maximally be achieved with market-based program activities under idealized market conditions. Furthermore, the *maximum achievable potential (MAP)* is not a result that would likely be achieved under current conditions, but rather provides a maximum benchmark against which future market and program potential can be interpreted. The idea of MAP is one that would not penalize future potential based on current conditions, but rather show that programs will include strategies that might remove barriers over time which could lead to higher market adoption rates. In essence, such a scenario would illustrate future shifts in programmatic priorities and consumer attitudes that would increase future savings. Navigant will develop details for the MAP scenario as part of Stage 2 work.