Summary of Revision

Revision for TSG-00707-2013
Paper title: "When Bias Matters: The Economics of Demand Response Baselines for Residential Customers"

December 23, 2013

Response to editor comments (Recommendation: Revise and Resubmit)

Comment E.1: Thank you for your submission. Four reviewers have analyzed the paper and they coincide that the paper has a value. However, while this paper is well researched and written, its results are not surprising even without a model. It is intuitive that positively biased baselines over-reward customers and are less attractive to utilities. The part of their conclusion of interest was that a positive bias can attract participation even with a low share – could be given more emphasis within the overall structure, among other topics. Please find below a number of significant comments and suggestions raised by each reviewer that should be carefully addressed by the authors.
Response E.1: Thank you for giving us the chance to make a major revision and resubmit. After carefully considering the reviewers' comments we have made substantial revisions. In addition, in Section I-C, where we described our contributions, we have emphasized the conclusion that more a positively biased baseline attracts customer participation. We hope that these revisions will satisfy the reviewers' requirements and comments.

Response to Reviewer 1

We would like to thank you for your comments; they were very useful for improving the quality of the paper. We have thoroughly revised the manuscript. In the following, we will explain in detail how we addressed all of your comments.

Comment 1.1: The four different methods proposed for the baseline calculations are simple methods in terms of accuracy of the calculations. Although the authors concluded from the results that accuracy of the baseline calculation is not effective as bias, why more sophisticated prediction algorithms are not considered (such as artificial intelligence methods) for higher accuracy in the baseline calculations?

Response 1.1: We explain our motivation on why we do not consider more sophisticated prediction algorithms in Section II (Demand Response Baselines).

DR baseline should be simple enough for all stakeholders to understand, calculate, and implement, including end-use customers [7], [8]. Thus, even though more sophisticated machine learning methods could deliver higher prediction accuracy, we do not consider them in this work. However, we refer interested readers to our previous work [17] on the performance analysis of more complex prediction methods, such as multilayer perceptron and support vector machine.

Comment 1.2: On page 3, column 1, first paragraph: in the common execution scenario proposed by the authors, it is mentioned that "the realized load will be calculated by reading the customers' smart meter data". I guess the authors agree with me that it brings on a huge amount of data and needs powerful and fast communication infrastructure. In what condition, the proposed scenario is practical? How often the data should be read? Is it possible to read millions of smart meters in real-time? What is the impact of communication delay?

Response 1.2: It is true that the amount of data communicated by smart meters is huge, and indeed, this is one of the many challenges that utility companies face nowadays. However, our framework does not need smart meter data in real-time. Assuming that the utility company does not need smart meter data for purposes other than DR, we consider two cases:

- if we would like to enable immediate feedback on customers' usage or reduction during a DR event, then, at the least, data needs to be communicated after each DR event, and assuming DR events last for one, two, or three hours, then communicating smart meter data every 60 minutes is fine;
- if immediate feedback is not needed, then to reduce communication cost (and frequency), smart meter data could be communicated once a day (or two/three times a day).

Comment 1.3: I feel that another table is required at the end of Section III to summarize the different baseline calculation methods. This table might contain the name of each method with a short explanation for each method.

Response 1.3: We have revised the paper according to your comment. At the end of Section II (Demand Response Baselines), we added a table which summarizes the different baseline methods:

Table 1: Summary of the baseline methods.

Baseline	Short description
HighXofY	average of the highest X of Y days
LowXofY	average of the lowest X of Y days
MidXofY	average of the middle X of Y days
Exp. moving avg.	weighted average of customer's consumption
Regression	linear regression of customer's consumption

note: all historical data considering only non-DR days preceding the DR event

Comment 1.4: On page 4, column 1, line 24: what does "generating load" mean? Did you mean the total cost of meeting load demand? Or cost of DR for the utility companies?

Response 1.4: We agree that this sentence was not clear. It is the total cost of meeting load demand. To avoid further confusion, we revised the sentence as follows: "We denote $c(L)$ as the total cost of meeting load demand L."

Comment 1.5: In Eq. 20, since the equation calculates the "true incentive", shouldn't be $L(\delta)$ instead of $B(\delta)$ in the equation?

Response 1.5: We added more information in the paragraph below Eq. 20 (now Eq. 18) and at the end of Section III-B (DR Event).

With reference to our old notation: In the real world, while $L(\delta)$ is not known, $B(\delta)$ and $R(\delta)$ are known. We assume that the utility company publishes $B(\delta)$ and $R(\delta)$. This does not violate customer's privacy (because both $B(\delta)$ and $R(\delta)$ are aggregate information over all customers). In addition, publishing $B(\delta)$ and $R(\delta)$ provides feedback to customers on how they perform in response to DR event δ as a community. This feedback on performance, and the company's transparency could potentially increase customer participation. Due to the fact that $L(\delta)$ remains unknown to the customers, we use $B(\delta)$ as the approximation of $L(\delta)$.

We revised our notation to make the paper clearer, as was also suggested by another reviewer. We distinguished whether the load actually occurred or not:

- we now denote the load that occurs in reality as "actual load" ℓ (previously original load ℓ in the absence of a DR event, and realized load r in the presence of a DR event),
- we now call the original load ℓ in the presence of a DR event, i.e., the energy that a customer would have consumed, as the "true baseline" b^*,
- the baseline load b, i.e., the estimate of the true baseline, is now emphasized as "predicted baseline" (with the same letter b).

We summarized the changes on the table below. We also have ensured that the changes are consistent throughout the paper. We hope this revision made the paper clearer.

Table 2: Notation revision

DR event	old term		revised term	
absence	ℓ, L	original load	ℓ, L	actual load
presence	r, R	realized load	ℓ, L	actual load
	b, B	baseline load	b, B	predicted baseline
	ℓ, L	original load	b^*, B^*	true baseline

Comment 1.6: On page 4, column 2, lines 39-41: why the proposed function is called "utility" function (which is confusing with the utility as the company whose responsibility is generating and distributing electricity to different customers and billing them each month)? Isn't it better to call it "profit" function?

Response 1.6: We have addressed this comment by calling it "profit" function (instead of "utility" function) consistently, throughout the paper. Thus, we hope that there will be no more confusion with the "utility" company.

Comment 1.7: On page 6, column 1, line 23: the "In term of bias" should be "In terms of bias" which is the common expression.

Response 1.7: We have revised the paper according to this comment. We have replaced "In term of bias" with "In terms of bias".

Comment 1.8: On page 7, column 1, lines 34-41 and in Figs. 3(a) and 3(a): when customers behave as a free riders, they still get incentive. Why the utility companies should pay incentive to the customers who do not participate in the DR by reducing their loads? How can you justify this?

Response 1.8: We have revised the paper by taking into account this comment. We have clarified this matter in Section V-B, at the end of paragraph "Additional profit of naïve and rational customer model".

As for the free riders ($\gamma = 0$) their true incentive is 0, because they do not carry out any reduction in consumption. However, there are some customers whose loads are overestimated by the baseline methods, and others whose loads are not. While the non-overestimated customers do not receive any incentive, the overestimated customers do receive some incentives (free riders). This is why the sum of the received incentives (and additional profit) over all customers when $\gamma = 0$ is positive.

Response to Reviewer 2

We would like to thank you for your comments; they were very useful for improving the quality of the paper. We have thoroughly revised the manuscript. In the following, we will explain in detail how we addressed all of your comments.

Comment 2.1: You introduce the MidXofY baseline on p2 and it seems very reasonable. Why wasn't it mentioned again later or included in the results section?

Response 2.1: We have revised the paper considering this comment. We have included MidXofY in the result section. That is, we added Mid4of6 to our net-benefit analysis (Section V).

Comment 2.2: Clarify the difference between baseline and original consumption and their use in the study earlier if possible, the reader keeps wondering how the original can ever be known until p4 where you explain the dataset which provides original and you model realized.

Response 2.2: We have clarified the difference between baseline and original consumption at the beginning of Section II (Demand Response Baselines). We also have explained that the original consumption is provided by our dataset, and that we model customer response to get the realized load (now, actual load) in Section III-B (DR Event), i.e., when we introduced these notations for DR events for the first time.
We revised our notation to make the paper clearer, as was also suggested by another reviewer. We distinguished whether the load actually occurred or not:

- we now denote the load that occurs in reality as "actual load" ℓ (previously original load ℓ in the absence of a DR event, and realized load r in the presence of a DR event),

- we now call the original load ℓ in the presence of a DR event, i.e., the energy that a customer would have consumed, as the "true baseline" b^*,
- the baseline load b, i.e., the estimate of the true baseline, is now emphasized as "predicted baseline" (with the same letter b).

We summarized the changes on the table below. We also have ensured that the changes are consistent throughout the paper. We hope this revision made the paper clearer.

Table 3: Notation revision

DR event	old term		revised term	
absence	ℓ, L	original load	ℓ, L	actual load
presence	r, R	realized load	ℓ, L	actual load
	b, B	baseline load	b, B	predicted baseline
	ℓ, L	original load	b^*, B^*	true baseline

Comment 2.3: Minor edits:
- p1-40: bring out $\rightarrow$ demonstrate
- p2-6: baselines on THE stakeholder utility
- p4-28: thought OF as

Response 2.3: We corrected the typos. In addition, a professional proofreader revised the new version of the manuscript.

Response to Reviewer 3

We would like to thank you for your comments; they were very useful for improving the quality of the paper. We have thoroughly revised the manuscript. In the following, we will explain in detail how we addressed all of your comments.

Comment 3.1: There is probably a bit more introductory material than is strictly necessary – anyone who reads this paper in this journal is likely to already understand why DR is used and how it works – but even reducing the amount of introductory material wouldn't allow the paper to shrink by a full page, so they might as well leave that material in, now that it's there.

Response 3.1: Based on this comment and one from another reviewer, we have reduced the introductory material. Furthermore, we also integrated the original manuscript's Section II (Related Work) into Section I (Introduction). This allowed us to adhere to the 8-page limit and add some revision materials.

Comment 3.2: First, the Regression Baseline definition seems odd. If I understand the notation, the regression prediction for the load at a given day and time does not include a constant term or any explanatory variables, but is simply a weighted sum of the loads on previous days at that time. There's nothing wrong with such a model but I don't think it is one that is in common use. Regression models I'm familiar with in this area usually include an explanatory variable, usually outdoor temperature or a time trend. I question whether it's even worth including this regression model; perhaps predictably, it does not fit very well. (I'm slightly surprised the regression models don't have lower MAE, though, closer to the ISONE model: I would have thought there were enough adjustable regression parameters to essentially duplicate the weighting used for ISONE.) Given that this model includes no predictive variables, the authors should avid implying that the other models improve better than regression models in general, when all that has been shown is that they behave better than this particular model.

Response 3.2:
- We realized that our regression formula did not express precisely what we meant. We have now revised the description, also taking into consideration your comment. Because our dataset does not contain temperature or other measurements which could potentially be explanatory variables, we do not include them in our regression models. Our feature vector consists of historical consumption at the same hour of day. Furthermore, we implicitly incorporate calendar features, by differentiating the regression for weekday/weekend and for the different hours of the day. We set the length of the feature vector such

that it is more than enough to capture the weekly trend, i.e., 7 for weekday and 5 for weekend. Note that in order to reach the same day one week prior to any specific weekday (or weekend day), we need to go back at least 5 previous weekdays (or 2 previous weekend days).

- We have explicitly stated in Section IV-B (Analysis), right after the comparison between LowXofY and Reg1/Reg2, that one should not conclude from our experiment that LowXofY is better than regression models in general. Adding some additional explanatory variables to the current regression models could increase their predictive power, which leads to better accuracy.

Additionally, we also performed a little analysis on the historical consumption feature [17]. See Figure 1 below for an example. The highest correlation is given by the consumption of h-1, h-2, h-3, h-24, h-48, h-72, ... Thus, when we would like to have a prediction for the day-ahead, h-1, h-2, and h-3 become unavailable, and we are left with h-24, h-48, h-72, ..., which are essentially the consumption levels for the same hour on the previous days.

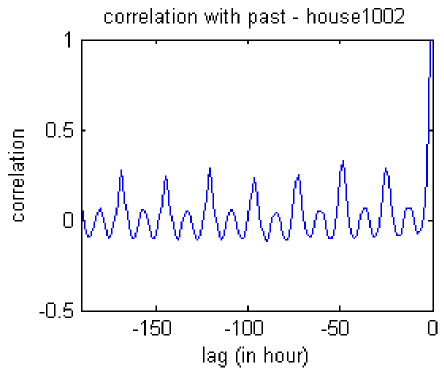

Figure 1: Auto-correlation of an individual household's consumption (id 1002).

Comment 3.3: Moving on... The "Evolutionary-rational" model suggests one way in which customer behavior could respond to experience. It might even be a reasonable model. But it seems very arbitrary. Perhaps the authors can cut a paragraph or two of intro material after all, and add a couple of paragraphs to explain where the "evolutionary-rational" model comes from. (I also have to question how rational the evolutionary-rational model actually is: a really rational customer who understands how the incentive system works would attempt to determine the key parameter values that define the system and to optimize her behavior given those values.)

Response 3.3: We have rewritten the first paragraph in this part to explain the motivation/rationale behind this model. This "evolutionary-rational" model is more about adapting its decision (gradually) based on its past experiences rather than determining the key parameter values of the system. The formula is inspired by the exponential moving average, to gradually forget the past and gives more importance to more current information. Finally, to avoid any confusion due to the model's name, we have changed its name to the "adaptive" model.

Comment 3.4: The authors talk about "original load" and introduce "true consumption" as a synonym, but primarily use the term "original load." I suggest getting rid of the term "true consumption" altogether: it's very confusing, since "true consumption" sounds like it would mean the actual load during the event, but it doesn't. I don't even like the term "original load", since this load does not in fact occur at any point. In some papers the term "true baseline" is used the way "original load" is used here, and "predicted baseline" is used the way "baseline" is used here. I prefer the "true baseline"/"predicted baseline" terminology to the "original load"/"baseline" terminology, but if other reviewers are OK with the terminology used in the paper then I am too.

Response 3.4: We have revised the paper considering your comment. We have removed the term "true consumption" altogether. In addition, we have made some changes to the terminology to make the paper clearer to readers. We denoted a customer's actual load (in the absence/presence of a DR event) as ℓ (previously ℓ and r, in the absence and in the presence of a DR event, respectively). We denoted predicted baseline (or baseline) as b, and true baseline as b^* (previously ℓ). We summarized the changes below. We have also ensured that the changes are consistent throughout the paper. We hope that the revision made the paper clearer.

DR event	old term		revised term	
absence	ℓ, L	original load	ℓ, L	actual load
presence	r, R	realized load	ℓ, L	actual load
	b, B	baseline load	b, B	predicted baseline
	ℓ, L	original load	b^*, B^*	true baseline

Comment 3.5: The x-axis on Figure 2 intersects the Bias axis at 0.05 instead of 0.00. I assume this was done on purpose to avoid obscuring the points clustered along bias=0, but it's easy to assume it goes through 0, at least at first, which is needlessly confusing. There's no perfect solution, but I think a better one would be to move the x-axis all the way to the bottom of the plot (just as the bias axis is all the way at the left), and then to draw a thick gray horizontal line at bias=0.

Response 3.5: We have revised the paper accordingly. We have moved the x-axis all the way to the bottom of the plot to avoid confusion.

Response to Reviewer 4

We would like to thank you for your comments; they were very useful for improving the quality of the paper. We have thoroughly revised the manuscript. In the following, we will explain in detail how we addressed all of your comments.

Comment 4.1: The introduction can be improved. The authors should focus on presenting their own message instead of spending too much space motivating DR. After reading the introduction, it is unclear what's the main contribution and objective of this paper. The reviewer also suggests that authors integrate section II (Related Work) into section I (Introduction).

Response 4.1: We have revised the paper considering your comment. We have reduced the space for the motivation and integrated Section II (Related Work) into Section I (Introduction). In addition, we have made our contribution and objectives explicit by putting them all together in Section I-C (Overview of Contributions)

Comment 4.2: The authors often use absurd words where they should be more specific. For example, the authors often say "study the impact of DR baseline applied to residential customers". The authors should be more specific about "the impact". Does it mean the willingness to participate in the DR or something else? Another example is that the authors use "incentive" here and there, but for most cases it is not clear what the meaning of the "incentive". There are other cases as well.

Response 4.2: We have revised the paper according to your comment. We have made clear that the impact studied is the economic benefit and customers' willingness to participate in a DR event. We have also clarified that the incentives that we consider in this work are monetary incentives, such as bill rebates, redeemable vouchers or discounts. We added this clarification in Section I. Moreover, a professional proofreader revised the new version of the manuscript.

Comment 4.3: (1) and (2) defines $l_i(d)$ and $b_i(d)$, but I'm not sure where the authors use them.

Response 4.3: We have removed the definition of $b_i(d)$. We used $l_i(d)$ in Section II (Demand Response Baselines). We have now brought the definition closer to where we use it for the first time.
We revised our notation to make the paper clearer, as was also suggested by another reviewer. We distinguished whether the load actually occurred or not:

- we now denote the load that occurs in reality as "actual load" ℓ (previously original load ℓ in the absence of a DR event, and realized load r in the presence of a DR event),
- we now call the original load ℓ in the presence of a DR event, i.e., the energy that a customer would have consumed, as the "true baseline" b^*,
- the baseline load b, i.e., the estimate of the true baseline, is now emphasized as "predicted baseline" (with the same letter b).

We summarized the changes on the table below. We also have ensured that the changes are consistent throughout the paper. We hope this revision made the paper clearer.

Table 4: Notation revision

DR event	old term		revised term	
absence	ℓ, L	original load	ℓ, L	actual load
presence	r, R	realized load	ℓ, L	actual load
	b, B	baseline load	b, B	predicted baseline
	ℓ, L	original load	b^*, B^*	true baseline

Comment 4.4: (8) should be $b_i(d,t)$; and the description and notations for this part (Exponential moving average baseline) can be much improved.

Response 4.4: We have revised the paper according to your comment.

- We have fixed the typo $b_{i,d}(t)$ to $b_i(d,t)$.
- For the explanation of the exponential moving average baseline: instead of introducing new notation $\Phi(\sim, d)$, we have now re-used our existing definition $D(Y,d)$ to express the set of Y most recent non-DR days preceding the day d with the same day type as d, i.e., $D(\infty, d)$. We hope that the explanation for this part is now more concise and clearer.

Comment 4.5: How to determine the regression coefficient in (9)?

Response 4.5: We have added the explanation (in the paragraph after the regression equation, Eq. 7 in the revised manuscript) that we estimate the regression coefficient using ridge regression. However, we also mentioned that other estimation methods can also be used.

Comment 4.6: For (20), it is impossible for the customer to calculate its tv_i because it requires the global information of $B(\delta)$ and $R(\delta)$. Who should calculate tv_i? It is also impractical for the utility to calculate tv_i otherwise the utility can use the true information $l_i(\delta)$ to give incentives. The reviewer is confused with the incentive model and not sure how to use it practically.

Response 4.6: With reference to our old notation: we assume that customer i is able to calculate tv_i based on the assumption that the utility company publishes $B(\delta)$ and $R(\delta)$. We now motivate this more clearly in the end of Section III-B (DR Event). Publishing this $B(\delta)$ and $R(\delta)$ does not violate customer privacy (since both of them are aggregate information over all customers). In addition, DR performance feedback can also be useful to foster customer participation [18].
Note that, in the revised paper, $R(\cdot)$ is replaced by $L(\cdot)$. We described our new notation in Response 4.4.

Comment 4.7: The authors use the summation of loads over the whole DR event to calculate the cost, savings, and to distribute the incentives, which the reviewer disagree with. Notice that the authors assume a DR event lasts for couple hours. If a DR event last for a long time, then cost, savings, and incentives should be defined for each time slot itself instead of the whole DR event.

Response 4.7: We define a DR event as a tuple consisting of a start time and an end time, and we compute the cost, savings, and incentives (or DR performance) over the period of the event.

- In our experiments, a DR event lasts for several timeslots (or hours, since we have hourly timeslots). We then compute the DR performance (by summing the loads) over the period of the event. We did it this way in order to be as close as possible to the real-world scenario, which also computes the DR performance over the period of the event [8, page 5].
- However, our framework can be generalized to compute DR performance per timeslot. In fact, a DR event that lasts several timeslots can be divided into smaller events of one timeslot each, for which the DR performance is computed on a per timeslot basis.

Comment 4.8: The reviewer also has a concern about the interactions between the DR behavior and baseline calculation. In the future grid, if customers often participate in DR events, then the resulting behavior will affects the calculation of baseline because the baseline already take DR into accounts. Does the model require customers to report their intended (original) consumption? If it does, then why not just use the reported information to calculate everything rather than the estimated baseline?

Response 4.8: We have mentioned in the beginning of Section II (Demand Response Baselines) that the baseline calculation does not include the days where DR events happen. Thus, the baseline calculation is not affected by customers' DR behaviors. More specifically, baseline calculation for day d uses historical data from $D(Y, d)$, which is a set of Y most recent non-DR days preceding the day d having the same day type (weekday/weekend) as d.