What are the benefits of weatherization?

ECON3391.01, Boston College

Intro

- Buildings are a major source of emissions
- Many programs aimed at reducing building energy use
- One of the largest is the WAP
- Expanded greatly under the stimulus
  - DOE says it was a huge success
  - Some economists question these results
- Debate is a good intro to importance of rigorous evaluation

The Weatherization Assistance Program

- Largest residential energy efficiency program in the US
  - Benefited over 7 million homes since 1976
- Provides free home improvements to low income households
- The 2009 ARRA (stimulus) dramatically increased WAP funding
  - $450 million in 2009
  - $5 billion 2011-2012
  - All owner occupied houses below 200% of poverty line eligible

The Weatherization Assistance Program (cont)

- WAP distributes funds to states based on climate and need
- States distributes the money to local implementers
- These sub-grantees identify and recruit eligible households
- Participants first receive a free audit to identify needs
    - insulation, new windows, furnace upgrades, etc.
- Auditor predicts the savings from each measure and makes recommendations.
- This audit data and information on local weather, etc are fed into a model, the National Energy Audit Tool (NEAT).
    - estimates energy savings
- Investments are provided for free

Why should you care about the net benefits of WAP?

Why should you care about the net benefits of WAP?

- Assessing the impact of weatherization is important.
- It is the federal government's largest effort to improve residential energy efficiency, a critical component of its strategy to combat climate change.
- It's important that the government does not waste money and gets the most bang for our (environmental) bucks.
- Results in this case are stark:
  - DOE finds benefits 4x the costs
  - Economists finds they are half the costs

DOE's evaluation found the WAP to be a resounding success

According to the , DOE press release:

- For every dollar invested in weatherization, $4.50 was generated in energy and non-energy benefits.
- Approximately 8,500 jobs were created or retained.
- Single-family homes saved an average of $283 annually on energy costs.
- Carbon reduction of 2.2 million metric tons.

Conclusion: "The results demonstrate that weatherization provides cost-effective energy savings and health and safety benefits to American families."

Here is the full report and summary of results.

How did the DOE arrive at these findings?

Fowlie, Greenstone and Worlfram critique

- Authors looked under the hood of the 4,500 page (!) DOE study
  - Our judgment is that many of the DOE's conclusions are based on dubious assumptions, invalid extrapolations, the invention of a new formula to measure benefits that does not produce meaningful results, and no effort to evaluate statistical significance
- Found numerous mistakes/ inconsistencies
- DOE had no control group for health and well being measures
- Treated population before and after, and a control group after only.
  - They averaged the differences between these groups.
  - "To the best of our knowledge, this approach has never been used in any textbook or research paper previously.....The approach to estimating nonenergy benefits is unrecognizable, and we believe the resulting estimates have no meaningful interpretation."

In DOE study, WAP investment costs still exceeded energy savings

- Start by checking if there is an energy efficiency gap
- DOE collected energy expenditures before and after weatherization for 16,000 households
    - Differencing gives an average annual savings of $223
- FGW argue that weatherization is a "purposeful" decision.
    - What does this mean? What are some stories here?

In DOE study, WAP investment costs still exceeded energy savings

- Start by checking if there is an energy efficiency gap
- DOE collected energy expenditures before and after weatherization for 16,000 households
    - Differencing gives an average annual savings of $223
- FGW argue that weatherization is a "purposeful" decision.
    - What does this mean? What are some stories here?
- DOE implicitly assumes no selection

FGW convert DOE's own estimates into NPV

Net Present Value of Energy Savings from DOE Study of ARRA Period

Time Horizon	Discount Rate		
	3 Percent	6 Percent	10 Percent
10 years	$2,003	$1,641	$1,370
16 years	$3,028	$2,254	$1,745
20 years	$3,646	$2,558	$1,899

Note: The table reports the net present value of energy savings (in dollars) implied by the DOE's household-level energy savings estimates ($223) using a range of discount rates and assumed time horizons. Reductions in energy bills associated with the estimates are assumed to accrue over the life of the measure.

- Cost of the program: $6,812 / $5,926 / $3,745 (unclear where the lower number comes from)
- DOE number higher: assumed (real) energy prices increase and investments last longer.
- Still only come to $3,190 in benefits.

Source: Fowlie (2015)

What about non-energy bill benefits?

- So there is no energy efficiency gap here.
- But the program still be justified by factoring other benefits.
- What might these be?

Problem: non energy benefits weren't measured

- Most estimates come from survey responses
- The DOE then takes several steps to translate these into $
- One major benefit is reduced hospitalization:
  - "In the past 12 months, has anyone in the household needed medical attention because your home was too cold (hot)?"
  - What do people think about this question?

Problem: non energy benefits weren't measured

- Most estimates come from survey responses
- The DOE then takes several steps to translate these into $
- One major benefit is reduced hospitalization:
  - "In the past 12 months, has anyone in the household needed medical attention because your home was too cold (hot)?"
  - What do people think about this question?
  - Using dubious formula, estimate medical attention declinded by 1.4 percent

Problem: non energy benefits weren't measured (cont)

- Want an estimate of reduced mortality.
  - However death statistics not collected.
- Take estimate and combine with many modeling assumptions:
  - assume what fraction of medical attention lead to hospitalization, ER visits or doctor's visits.
  - assume the fraction of these that lead to death
  - Combine with value of statistical life
  - "responses to a survey question about "medical attention" turn into $5,000 worth of benefits per weatherized household".
- Similar approach applied throughout the study
  - Responses to questions about sleepless nights are converted into productivity benefits that exceed $3,000 per household.

"Selective accounting?"

- Many assumptions also appear inconsistent with available evidence
  - "In the only report where indoor temperature is directly measured, DOE finds that average indoor temperature increased by 0.3 F in weatherized homes as compared to a control group. This very small change calls into question the basis for the claimed health effects due to reductions in thermal stress."
- The DOE also included statistically insignificant results in instances where benefits were measured:
  - This study found no significant changes in carbon monoxide following weatherization. Yet, the claimed benefits include reductions in CO poisoning.
- In contrast, weatherization was found to increase radon and formaldehyde levels. In the accounting of costs and benefits, however, no effort was made to quantify the potential health costs from the increase in these potential health risks.
- FGW conclude: "This selective accounting is a cause for concern."

ORNL's response to the FGW critique

ORNL's response to the FGW critique

- On the mortality benefits, ORNL presented results by "tiers" of precision
- Strongly rejected mortality assumptions were speculative
  - I think this is a great research topic
  - Unfortunately ORNL doesn't actually have the data to do it (email)
- Still does not not appreciate the flaws in his averaging approach
- What to do absent statistical significance?
- Basic rebuttal: "Ultimately, the evaluation team analyzed the benefits from multiple angles and determined that ancillary benefits not related to energy savings were not being fully recognized."

Fowlie, Greenstone and Wolfram (2015)

- 30,000 WAP-eligible households in Michigan
- Randomized encouragement design:
  - 25% of hh's put in treatment group
  - received extensive outreach and assistance signing up for WAP
  - but control households still eligible for WAP under own volition

Result 1: People do not seem to want to weatherize

- Average WAP household received $5,150 worth of home improvements **totally free**
- Yet in the control group, only 1% opt into the program
- In the treatment group, only 6% opt in despite extensive encouragement
- What do people make of this?

Result 1: People do not seem to want to weatherize

- Average WAP household received $5,150 worth of home improvements **totally free**
- Yet in the control group, only 1% opt into the program
- In the treatment group, only 6% opt in despite extensive encouragement
- What do people make of this?
- Takeway: there must be large hassle costs associated with this program.
    - What might those be?

FGW encouragement activities

Encouraged group (households)	8,648
Initial home visits	6,694
Robo-calls	23,500
Personal calls	9,171
Follow up appointments	2,720
Average cost/encouraged hh	$55.00

Note: The table summarizes efforts to encourage a group of Michigan households to take up weatherization assistance. These households were selected randomly from a sub-population of households who were located in the service territory of our partner utility and presumptively eligible based on ex ante available income information.

The authors estimate two models of energy savings

$$ln(y_{imt}) = \beta \mathbf{1}\{WAP\}_{imt} + \alpha_{im} + \alpha_{mt} + \epsilon_{imt}$$

- Quasi-experimental approach: Difference in differences
  - Estimates average treatment effect on treated (ATET)
- Experimental approach: Instrumental variables
  - Typically we think of an RCT as randomly assigning $\mathbf{1}\{WAP\}$
  - But in this case it just increased the probability of assignment
  - Can use IV to predict probability of $\mathbf{1}\{WAP\}$, ie use treatment group assignment as an instrument
  - If treatment effects differ across households, this will recover the local average treatment effect (LATE)
    - Or the treatment effect of "compliers"

Review: Estimating the Causal Effect of an EE Investment

Each individual (i) has two potential outcomes Y (energy use)

- one where they have weatherization (Y_{i1})
- and one where they don't (Y_{i0})

Review: Estimating the Causal Effect of an EE Investment

Each individual (i) has two potential outcomes Y (energy use)

- one where they have weatherization (Y_{i1})
- and one where they don't (Y_{i0})

For any individual, the weatherization "treatment effect" is the amount of energy they save from weatherization: $Y_{i1} - Y_{i0}$

From a policy perspective, we may be interested in the average treatement effect (ATE): $\beta^{ATE} = E[Y_{i1} - Y_{i0}]$

Review: Estimating the Causal Effect of an EE Investment

Each individual (i) has two potential outcomes Y (energy use)

- one where they have weatherization (Y_{i1})
- and one where they don't (Y_{i0})

For any individual, the weatherization "treatment effect" is the amount of energy they save from weatherization: $Y_{i1} - Y_{i0}$

From a policy perspective, we may be interested in the average treatement effect (ATE): $\beta^{ATE} = E[Y_{i1} - Y_{i0}]$

Fundamental Problem of Causal Inference:
We only every observe one state of the world for each individual.

To make progress, we are forced to compare households that got the treatment to those that didn't. How bad is this?

Selection Bias

Want to learn about the true population average treatment effect (average savings from weatherization)

$$\beta^{ATE} = E[Y_{i1} - Y_{i0}]$$

What if we compare WAP and no-WAP homes:

$$\begin{aligned}
\hat{\beta} &= E[Y_{i1}|WAP_i = 1] - E[Y_{i0}|WAP_i = 0] \\
&= E[Y_{i1}|WAP_i = 1] - E[Y_{i0}|WAP_i = 1] \quad \text{(Causal Effect)} \\
&\quad + E[Y_{i0}|WAP_i = 1] - E[Y_{i0}|WAP_i = 0] \quad \text{(Selection bias)}
\end{aligned}$$

Our estimate of the energy saved is biased by "selection" – the difference in the untreated state of the world, across households who do and don't get WAP.

Who selects into weatherization?

Simple model of energy efficiency investment

- Individuals have expectations $\tilde{E}$ about how much energy they'd use with and without WAP, $\tilde{E}[Y_{i1}]$ and $\tilde{E}[Y_{i0}]$
  - Note that these expectations could be biased, in which case $\tilde{E}[Y_i]$ won't equal the true empirical average $E[Y_i]$
- Compare expected savings $\tilde{E}[Y_{i1} - Y_{i0}]$ against the "cost" of investment C_i.
  - Here WAP is "free", however you still need to take off work, clean your attic, etc

Who selects into weatherization?

Simple model of energy efficiency investment

- Individuals have expectations $\tilde{E}$ about how much energy they'd use with and without WAP, $\tilde{E}[Y_{i1}]$ and $\tilde{E}[Y_{i0}]$
  - Note that these expectations could be biased, in which case $\tilde{E}[Y_i]$ won't equal the true empirical average $E[Y_i]$
- Compare expected savings $\tilde{E}[Y_{i1} - Y_{i0}]$ against the "cost" of investment C_i.
  - Here WAP is "free", however you still need to take off work, clean your attic, etc

Who adopts weatherization?

1. Imagine C_i is the same for everyone, and everyone has perfectly rational expectations of savings. Then the people who adopt are the people with the highest savings. Averaging this, would clearly lead to an underestimate.
2. Imagine households vary in their adoption costs C_i or expectational errors $\tilde{E}[Y_{i1} - Y_{i0}] - E[Y_{i1} - Y_{i0}]$. Bias depends on the sign of this correlation.

What if we run an experiment?

- Ideally we would randomly assign households to weatherization.
- Since we assign WAP_i randomly, our selection bias term $E[Y_{i0}|WAP_i = 1] - E[Y_{i0}|WAP_i = 0]$ equals zero.

What if we run an experiment?

- Ideally we would randomly assign households to weatherization.
- Since we assign WAP_i randomly, our selection bias term $E[Y_{i0}|WAP_i = 1] - E[Y_{i0}|WAP_i = 0]$ equals zero.
- Unfortunately, in the real world we can't force people to participate in this program.
- We can however, *encourage* them.
    - this could be a monetary incentive, advertising, persuasion, etc
- If we encourage people randomly, we can use that as an instrument to estimate *a* casual relationship.

Causal estimation under random encouragement

Define treatment (Z_i) as whether or not you're encouraged. Taxonomy of household types (t)

- **Always takers ($t = A$):** take WAP without encouragement ($WAP_i = 1$ if $Z_i = 1$ or if $Z_i = 0$)
- **Compliers ($t = C$):** only take WAP if encouraged ($WAP_i = 1$ if $Z_i = 1$; $WAP_i = 0$ if $Z_i = 0$)
- **Never takers ($t = N$):** don't take WAP even if encouraged ($WAP_i = 0$ if $Z_i = 1$ or if $Z_i = 0$)

	Treatment	Control
WAP	Always takers / Compliers	Always takers
No WAP	Never takers	Never takers / Compliers

Intent to treat/ Randomized encouragment

Let $Pr(t)$ be the probability of each type compliance type in the population, and let Y^t_{WAP} be the average electricity use given WAP state.

$$E[Y|Control] = E[Y^N_{i0}]Pr(N) + E[Y^C_{i0}]Pr(C) + E[Y^A_{i1}]Pr(A)$$

$$E[Y|Treatment] = E[Y^N_{i0}]Pr(N) + E[Y^C_{i1}]Pr(C) + E[Y^A_{i1}]Pr(A)$$

From these two equations, its clear that they **only** group we learn about from this experiment is the compliers.

Intent to treat/ Randomized encouragment

Let $Pr(t)$ be the probability of each type compliance type in the population, and let Y^t_{WAP} be the average electricity use given WAP state.

$$E[Y|Control] = E[Y^N_{i0}]Pr(N) + E[Y^C_{i0}]Pr(C) + E[Y^A_{i1}]Pr(A)$$

$$E[Y|Treatment] = E[Y^N_{i0}]Pr(N) + E[Y^C_{i1}]Pr(C) + E[Y^A_{i1}]Pr(A)$$

From these two equations, its clear that they **only** group we learn about from this experiment is the compliers.

Define the average savings for this group as the Local Average Treatment Effect (LATE).

If we simply take the raw average of the treatment and control groups, we get the Intent to treat effect (ITT), which is the LATE times the probability of compliers in the population:

$$E[Y|Treatment] - E[Y|Control] = (E[Y^C_{i1}] - E[Y^C_{i0}])Pr(c)$$

Better LATE then never?

This tells us that even with an experiment (or a good instrumental variable), we can *never* learn about the true average savings for the entire population.

The best we can do is to learn the average (causal) effect for housholds that would get WAP if encouraged, but otherwise would not.

What do people think of this?

Result 2: The energy savings are less than the upfront costs

Panel A: Dependent variable is monthly energy consumption (in logs)

	Total Energy		Gas	Electricity
	(1)	(2)	(3)	(4)
	OLS-FE	IV-FE	IV-FE	IV-FE
WAP	-0.10**	-0.20*	-0.21**	-0.10
	(0.01)	(0.08)	(0.08)	(0.10)
Imputed baseline consumption MMbtu/month	7.52		6.38	2.19
F-statistic	.	267.41**	260.10**	266.78**
Households	27,990	27,229	26,054	27,115
Observations	1662781	1653583	1528526	1638337

- Realized savings were 10-20%
- IV $<$ OLS. Suggests selection bias term $E[Y_{i0}|WAP_i = 1] - E[Y_{i0}|WAP_i = 0]$ is greater than zero.
- ie households selecting into WAP would have had higher than average non WAP usage

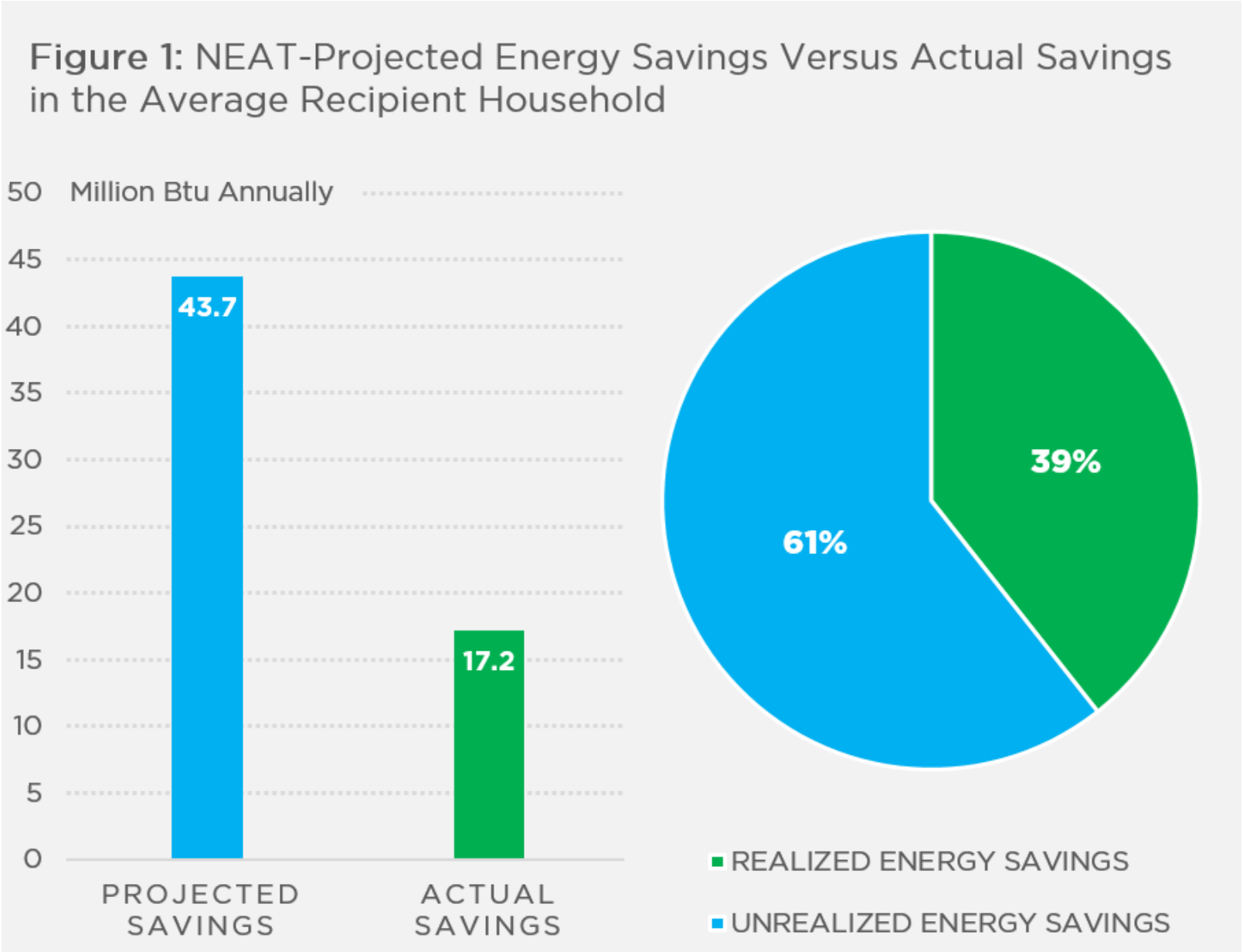
While these savings are substantial, they are significantly less than the up front costs

Panel B: Present value of (discounted) savings

Time Horizon	Discount rate		
	3 percent	6 percent	10 percent
10 years	$2,003	$1,728	$1,443
16 years	$2,949	$2,373	$1,837
20 years	$3,493	$2,693	$1,999

- Average WAP household received $5,150

Engineering model projected savings were 2.5 times larger



Source: Fowlie (2015)

Even accounting for environmental externalities, the program still has negative costs

Ex Ante (NEAT) vs Ex post (FGW) benefit estimates

Panel C: Social internal rate of return		
10 years	-1.0%	-20.0%
16 years	5.4%	-9.5%
20 years	7.0%	-6.1%
Panel D: CO_2 abatement cost - 3 percent discount ($/ton CO_2)		
10 years	$29	$552
16 years	-$19	$329
20 years	-$35	$255

Could low savings be due to a rebound effect?

- What is the rebound effect?
- Do you think it would be large here?

Could low savings be due to a rebound effect?

- What is the rebound effect?
- Do you think it would be large here?
- Why does it matter if there's a rebound effect?

We don't have much evidence on rebound effects

- Typically hard to measure
- FGW randomly contacted a subset of treated and non-treated households
- Went to 1,658 households on cold days
    - Asked what the thermostat was set to
- 899 people let them in the house (688 let them close the door)
    - Moved to the center of the room and waved two thermometers around

Result 3: No evidence of a rebound effect

	Thermometer		Thermostat	
	(1)	(2)	(3)	(4)
Base temperature	72.36**	72.17**	69.26**	68.91**
	(0.95)	(1.24)	(0.96)	(1.29)
Weatherized home	0.57	0.65	-0.57	-0.56
	(0.41)	(0.44)	(0.29)	(0.33)
Heating Degree Days	-0.16**	-0.15**	0.04	0.05
	(0.03)	(0.04)	(0.03)	(0.04)
Propensity Score Weights?	N	Y	N	Y
R-squared	0.02	0.02	0.01	0.01
Observations	1359	1359	899	899

Note: The table reports measured indoor temperature differentials across weatherized (WAP) and unweatherizedhouseholds. Columns (1) and (2) have the indoor thermometer temperature reading as a dependent variable while columns (3) and (4) use the survey thermostat readings. Columns (2) and (4) are weighted so that surveyed population better represents total quasi-experimental sample. Standard errors clustered at the household level.

What are some critiques of the FGW study?

What are some critiques of the FGW study?

- This is just one study.
  - External validity?
- WAP is about more than just energy savings
  - other benefits not measured?

Call in the auditors?

- What did people learn from this debate?

Call in the auditors?

- What did people learn from this debate?
- I personally think it's unfair lump ORNL together with polluters in India
- But selective attention affects all of us, and ORNLs response shows clear signs of this.
- Highlights the benefits of independent evaluators

Summary

- Energy efficiency
  - engineering models over estimate
  - people really can't be bothered
  - rebound effect small
- Broader
  - importance of transparent and independent evaluation
  - RCTs in energy able to deal with selection