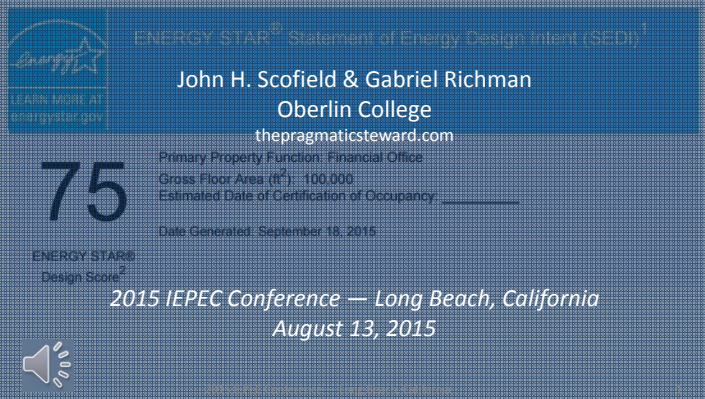
# Results of Validation Tests Applied to Seven ENERGY STAR Building Models

John H. Scofield & Gabriel Richman
Oberlin College
thepragmaticsteward.com

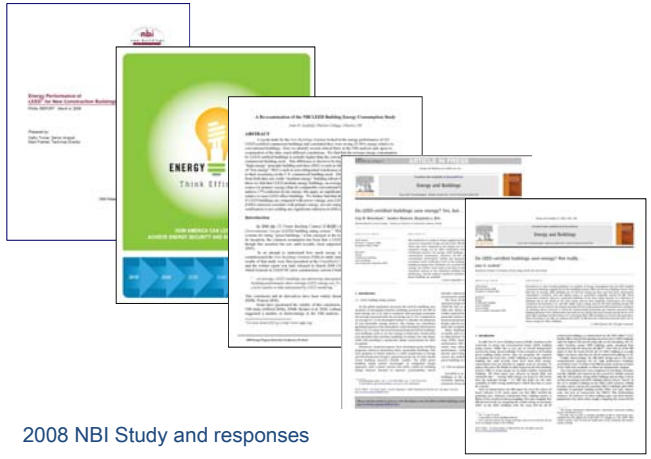



*2015 IEPEC Conference — Long Beach, California*
*August 13, 2015*

## Outline

- Why do I have concern about ENERGY STAR scores?
- ENERGY STAR Methodology
- Main conclusion about ENERGY STAR scores
  Score is ad hoc – loosely related to energy consumption
  EPA regression is largely irreproducible
- Today's focus: external validation of regression
  Of 9 models based on CBECS data, 6 fail miserably, 1 passes, 2 marginal
- Suggested future direction for ENERGY STAR





2008 NBI Study and responses



## Testimony before the House Subcommittee on Investigations and Oversite

*John H. Scofield – May 8, 2012*

"The closest thing to a scientifically-based green building rating system of which I am aware is the ENERGY STAR building score. It isn't very sexy, but it is based upon:
1) metered energy data,
2) primary (or source) energy Consumption, and
3) requires data validation by a third party."

ARTICLES

## New Analysis: LEED Buildings are in Top 11th Percentile for Energy Performance in the Nation

Published on 13 Nov 2012 | Written by Ashley Katz | Posted in Media


*LEED-certified buildings in the analysis had average ENERGY STAR score of 89 San Francisco building stock among top performers*




Energy and Buildings 67 (2013) 517–524

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

## Efficacy of LEED-certification in reducing energy consumption and greenhouse gas emission for large New York City office buildings

John H. Scofield*

*Department of Physics & Astronomy, Oberlin College, Oberlin, OH 44074, USA*

ARTICLE INFO

*Article history:*
Received 23 April 2013
Accepted 18 August 2013

*Keywords:*
Energy
Commercial buildings
Green buildings
LEED-certification
Energy use intensity
Source energy
ENERGY STAR
Energy benchmarking
Green house gas emission

ABSTRACT

In this paper 2011 energy consumption, green house gas (GHG) emission, and ENERGY STAR *Energy Performance Rating* (EPR) data for 953 office buildings in New York City are examined. The data were made public as a result of New York City's local law 84. Twenty-one of these office buildings were identified as LEED-certified, providing the opportunity for direct comparison of energy performance data for LEED and non-LEED buildings of the same type, time frame, and geographical and climate region. With regard to energy consumption and GHG emission the LEED-certified buildings, collectively, showed no savings as compared with non-LEED buildings. The subset of the LEED buildings certified at the *Gold* level outperformed other NYC office buildings by 20%. In contrast LEED *Silver* and *Certified* office buildings underperformed other NYC office buildings. The average EPR for the LEED buildings was 78, 10 pts higher than that for all NYC office buildings, raising questions about the validity and interpretation of these EPR's. This work suggests that LEED building certification is not moving NYC toward its goal of climate neutrality. The results also suggest the need to re-examine some aspects of ENERGY STAR's benchmarking tool.





## Background

- 2009 IEPEC paper debunking NBI LEED-building study
  LEED buildings, on average, show no primary energy savings
- 2012 testimony before Congressional Subcommittee
  embraced ENERGY STAR building certification
- 2012 study of energy use by LEED-certified NYC Office buildings
  LEED offices don't save energy but have higher ENERGY STAR scores
- 2012 ASHRAE case study on Adobe Headquarters buildings
  all have scores of 99 or 100 yet average source EUI = 350 kBtu/sf
- 2012 USGBC Press Release claiming LEED buildings save 39%
  average ENERGY STAR scores for LEED buildings was 89

ENERGY STAR scores suggest above average energy efficiency
Energy data demonstrate no savings

Is the ENERGY STAR score flawed?

## Work behind this paper

- Three years of investigation
- EPA public *Technical Methodology* documents
- discussions with EPA staff
- discussions with LBNL and ORNL building scientists
- filed 30+ *Freedom of Information Act* (FOIA) requests
  hundreds of records including non-public documents and PM data
- fully replicated 12 current EPA building models
  regression, cumulative EER distribution, Lookup Tables, etc.
  external (where possible) and internal validation of all models
- Replication of some older models (Office, K-12 School)

Preliminary results presented at 2014 ACEEE Summer Study on Buildings



## ENERGY STAR benchmarking score

- Key metric is building source energy use intensity (EUI): **$e = E/A$**
- One could compare **e** with **e's** for national stock of similar buildings
- But this does not adjust for external factors that impact energy use
- Central ENERGY STAR contribution:
  - identify external factors **$\{x_1, x_2, \ldots\}$** that impact energy use
  - perform multivariate linear regression to extract **$\{a_0, a_1, a_2, \ldots\}$**
  - calculate predicted EUI, **$p = a_0 + a_1x_1 + a_2x_2 + \ldots$**
  - define a building's *Energy Efficiency Ratio* **$EER = e/p$**
- EER is basis for a building's energy efficiency ranking
  - lower EER achieved by reduced e
  - lower EER achieved by raising p (through external factors)

How reliable is the EPA's regression?
Would a different choice of parameters yield different p?
What is the uncertainty in p?



## ENERGY STAR building models

| ENERGY STAR Building Models | Latest revision | Data Source | Dataset n | Dataset N | Regression $R^2$ | Regression m |
|---|---|---|---|---|---|---|
| Residence Hall/Dormitory* | 2004 | CBECS 1999 | 79 | 35,000 | 88% | 4 |
| Medical Office* | 2004 | CBECS 1999 | 82 | 87,000 | 93% | 5 |
| Office/Finance/Bank/Court | 2007 | CBECS 2003 | 498 | 250,000 | 33% | 9/11 |
| Retail | 2007 | CBECS 2003 | 182 | 152,000 | 71% | 9 |
| Supermarket/Grocery | 2008 | CBECS 1999/2003 | 83 | 24,000 | 51% | 7 |
| Hotel | 2009 | CBECS 2003 | 142 | 54,000 | 37% | 6 |
| K-12 School | 2009 | CBECS 2003 | 353 | 300,000 | 27% | 11 |
| House of Worship | 2009 | CBECS 2003 | 269 | 250,000 | 37% | 8 |
| Warehouse | 2009 | CBECS 2003 | 277 | 190,000 | 40% | 8 |
| Senior Care | 2011 | Industry survey | 553 | 31,000 | 43% | 10 |
| Hospital | 2011 | Industry survey | 191 | 4,500 | 22% | 4 |
| Multifamily Housing | 2015 | Industry survey | 322 | 160,000 | 24% | 5 |

* older models perform regression on LnE and do not utilize CBECS weights

**Today look at seven models based on 2003 CBECS data.**
**Note typical regression $R^2$ ranges from 30-50%**



## Why is validation necessary?

- Regression applies only to a dataset
- Frequently assume that results are more general
- Suppose regression predicts future stock prices?
- Best method external validation
  obtain second dataset
  see if regression accurately predicts new result
- Internal validation more common

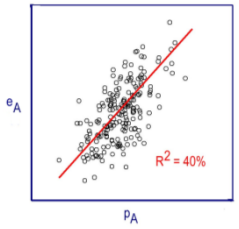
## External Validation Method

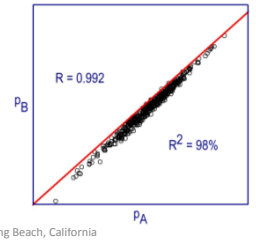
- Begin with the EPA's model data from 2003 CBECS (dataset A)
- Perform the model regression → yields $R_A^2$ and $\{a_0, a_1, ..., a_m\}_A$
- Obtain a second, equivalent data from 1999 CBECS (dataset B)
- Perform the model regression → yields $R_B^2$ and $\{a_0, a_1, ..., a_m\}_B$
- Combine datasets A + B
- Use $\{a_0, a_1, ..., a_m\}_A$ to predict EUI ($p_A$) for combined dataset
- Use $\{a_0, a_1, ..., a_m\}_B$ to predict EUI ($p_B$) for combined dataset
- Expectation is that $p_A = p_B$ for each of the samples
- Graph $p_B$ vs $p_A$ to see how well they agree
- Calculate correlation coefficient R
- $R^2 \geq 90\%$ indicates good agreement



## Results for Simulated Building Data

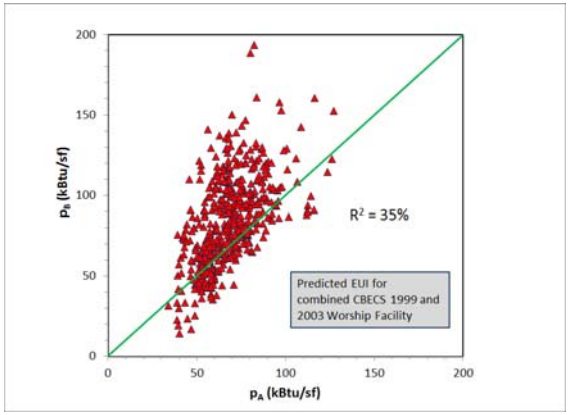
- construct a simulated building stock:
  $e = a_0 + a_1x_1 + ... + a_5x_5 + \text{noise}$ (large noise term)
- generate 10,000 rows of data $\{e, x_1, ..., x_5\}$
- randomly select 200 of these as **dataset A** and another 200 for **dataset B**
- follow validation procedure to compare $p_A$ and $p_B$





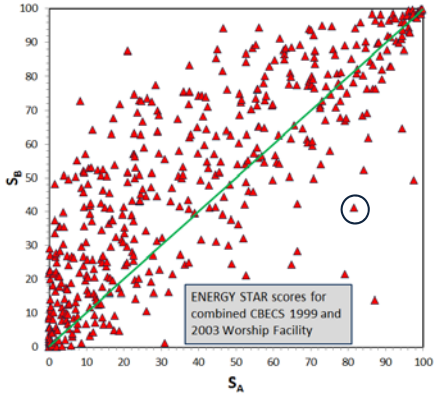

## Worship Facilities




- Two regressions produce very different predicted EUI
- Little correlation between $p_A$ and $p_B$



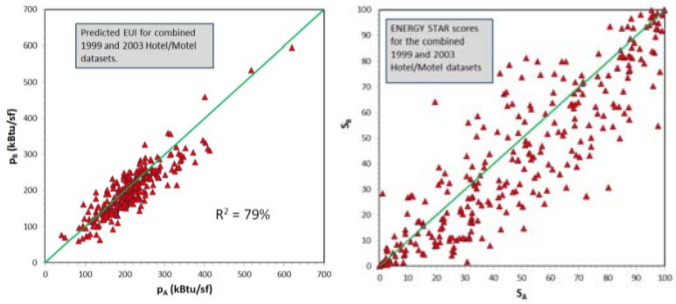
## Worship Facilities – impact on ES scores




$S_A = 82$, $S_B = 40$

- $p_A$ and $p_B$ produce two different ENERGY STAR scores $S_A$ and $S_B$
- rms differences between $S_A$ and $S_B$ are 18 points

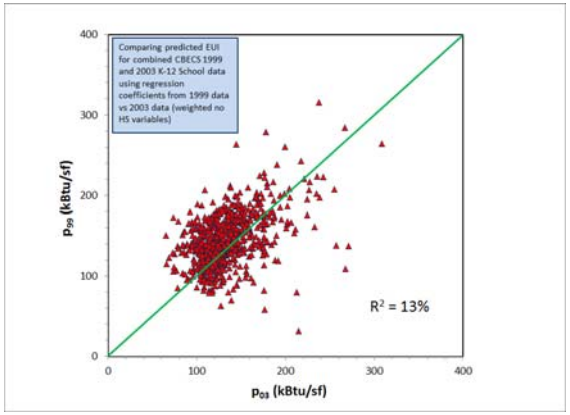
## Hotels/Motels




- Correlation between $p_A$ and $p_B$ is not great, though better than for Worship model
- Still impact on ENERGY STAR scores is significant: $dS_{rms}$ = 16 pts



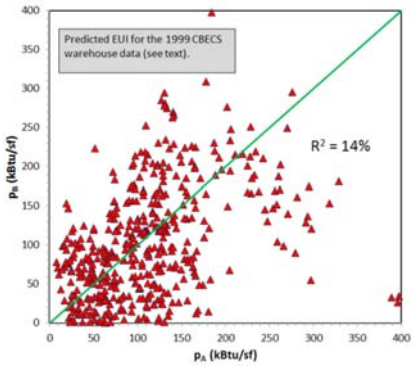
## K-12 Schools




- Revised EPA model to eliminate High School variables
- Little correlation between $p_A$ and $p_B$



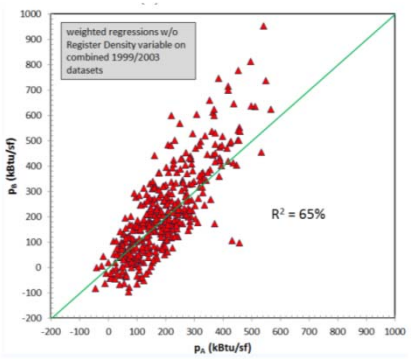
## Warehouses




- Revised 2003 Warehouse model data to include self-storage facilities
- Very little correlation between $p_A$ and $p_B$

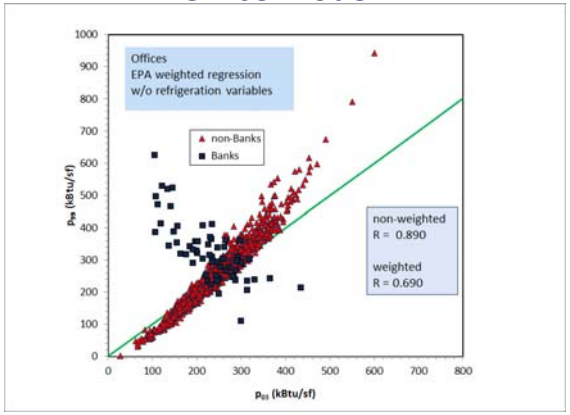


## Retail Stores




- Revised model by eliminating *Register Density* variable (see paper)

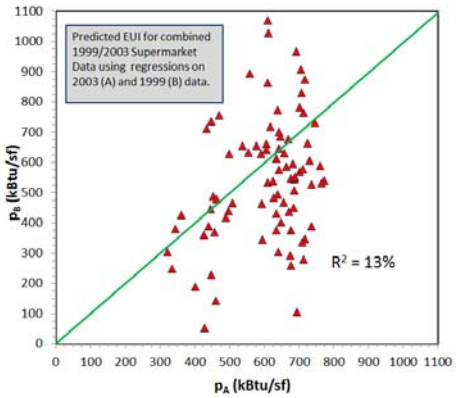
## Office Model




- Office model modified to remove refrigeration variables
- without Bank data R = 0.97 and $R^2$ = 95%



## Supermarket/Grocery Store




- This dataset is particularly problematic
- In the 2014 ACEEE demonstrated equal model generated from random numbers



## What are the implications?

- ENERGY STAR scores for four building types (*Worship*, *Grocery*, *Warehouse*, *K-12 Schools*) are based on regressions that have no general validity. Scores for these buildings are meaningless.
- Previously (ACEEE paper) I have shown the same thing for *Medical Office* and *Residence Hall* models
- Regressions for two other models have marginal validity (Hotels/Motels and *Retail Stores*).
- Only the *Office* model (much more data) demonstrates external validity – and that is without banks.
- No reason to believe that the remaining three models are better no external data available to test validity



## Directions for ENERGY STAR?

- Data collection aspect of Portfolio Manager useful
- Return to non-weighted regressions
- Regressions based on physical models
- Revisit all regressions and eliminate any variables that do not individually demonstrate validity
- Use CBECS 2012 data to test existing models before developing new ones
- If necessary – abandon use of p altogether and focus only on EUI.

Any claims of energy savings based on ENERGY STAR scores have no scientific basis. They should be ignored.