Leveraging Big Data to Develop Next Generation Demand Side Management Programs and Energy Regulations

Daniel Young, Mike McGaraghan, Nate Dewart,
Peter Borocz and Forest Kaser, Energy Solutions
Pat Eilert, Pacific Gas and Electric Company
Daniel Hopper, Southern California Edison

ABSTRACT

Big data – the concept of developing and analyzing increasingly large sets of data – is becoming more common as the tools for data collection become increasingly powerful. In the energy efficiency industry, through the application of custom web-crawling software, it is now possible to collect massive amounts of data to support improved analyses of integrated demand side management (IDSM) initiatives. Many online retailers are already set up to facilitate this type of ongoing data collection through their websites' application programming interface (API) specifications, thereby facilitating significant opportunities for incentive programs and energy codes and standards. This paper discusses potential uses in this field for big data by first presenting a review of existing federal efficiency regulation research that has explored various product performance-price relationships through the effect of *learning* on manufacturer practices and technological innovation, and associated *experience curves* used to account for decreasing production costs over time. We then present the results of big data-driven analysis for a recent computers efficiency standards proposal, and in more detail, an ongoing project for LED lamps. We conclude by discussing the potential for leveraging big data to help inform energy efficiency advocacy efforts by greatly improving the accuracy of key planning metrics, such as incremental measure costs and efficiency distributions for given products. In sum, big data may prove to be a game-changing tool for the energy efficiency industry to maximize the savings potential for the next generation of DSM initiatives.

Introduction

The development of integrated demand side management (IDSM) programs and energy efficiency codes and standards relies on models that estimate the energy savings and cost-effectiveness of the proposed measures. Ideally, these models would be calibrated to account for the dynamic nature of measure production costs, product pricing, incremental measure cost (IMC), product performance, and naturally occurring market adoption rates (NOMAD). Most products experience significant market changes over time (whether rapid or gradual) in several or all of these categories, and the forecasting of these key inputs into future analysis periods by standards analysts and program planners has been constrained by available data and methods for collecting this data. These market shifts result from a number of factors including improved manufacturing processes and industry *learning*, availability of raw materials, increased industry competition, or fluctuating market demand, and can render data from traditional collection methods inaccurate within a few months. Moreover, traditional data collection methods are often targeted to one specific purpose and are therefore limited in market breadth, resulting in a cumbersome and time-consuming task of reconciling the results into one useful data set. For example, many existing consumer electronic resources that could be used for IDSM program research purposes, such as CNET.com, are designed to provide information on model cost, while others like ENERGY STAR® collect data on energy consumption (but only for a subset of more efficient products).

Fortunately, the tools for data collection are becoming increasingly powerful, and the development and analysis of large datasets, or use of *big data*, can greatly expand the types of analysis available. By consolidating the collection of price points and multiple product attributes in real-time and gathering the data in increasingly large quantities, this big data method can dramatically improve efforts to quantify price-performance relationships and enable much more reliable and defensible forecasts for product performance and pricing over time. These analysis tools can be useful for many different applications within the energy efficiency field to optimize program and standards design and increase energy savings potential.

This paper briefly describes an evolution of the use of *experience curves* and the introduction of the big data concept within the energy efficiency standards context and discusses some other potential uses; it then concludes by highlighting one web crawler model designed by the authors to collect data for incremental measure cost (IMC) and performance data for light emitting diode (LED) replacement lamps.

Recent Research Methods Focusing on Dynamic Market Analysis

Over the years, there have been several studies designed to understand the effect of learning on manufacturing costs over time. The general principle assumes that as an industry gains more experience making a given product, it will be able to make the product at a lower cost. This idea was first pioneered by the Boston Consulting Group (BCG) in the mid-1960s in their analysis of the semiconductor industry. BCG found that a leading manufacturer in the industry saw the per unit cost of manufacturing fell by about 25% each time it doubled its cumulative total production volume (The Economist 2009). A similar study by RAND Corp. in the late-1970s investigated the experience curve associated with the cost of building nuclear power plants (Mooz 1978).

Since this original research on experience curves, several studies have explored this concept through a review of Department of Energy appliance efficiency rulemakings. Several studies have compared the predicted price increases associated with standards to the actual observed price changes (DOE 1989; Greening et al. 1996; Nadel 2000; Dale et al. 2009; Mauer et al 2013). Though multiple factors besides industry learning may be responsible, these studies generally reached the conclusion that actual product prices after standards took effect were lower than what had been predicted by DOE during its standards rulemakings. This indicates that past standards cost-effectiveness analyses likely overstated the relative costs of energy efficiency standards. These studies provided a solid foundation for DOE's later analyses and introduction of a methodology for calculating experience curves, as discussed below.

Historically, DOE appliance standards analyses have calculated an IMC for each proposed standard level, and projected that IMC to remain constant in the future. In its most recent analysis, DOE attempted to apply the lessons learned from past research on experience curves to develop a methodology for projecting a decrease in IMC over time. In a study completed in 2012, DOE analyzed the potential impact of experience curves on the final selection of energy efficiency standard levels for five rulemakings. DOE found that in all of these rulemakings, factoring in experience curves to scale down incremental costs over the analysis period would have resulted in much higher NPV estimates across the board; in some cases, comparing against the constant IMC assumption, a higher standard level would appear to be the most cost-effective option (Desroches et al. 2012).

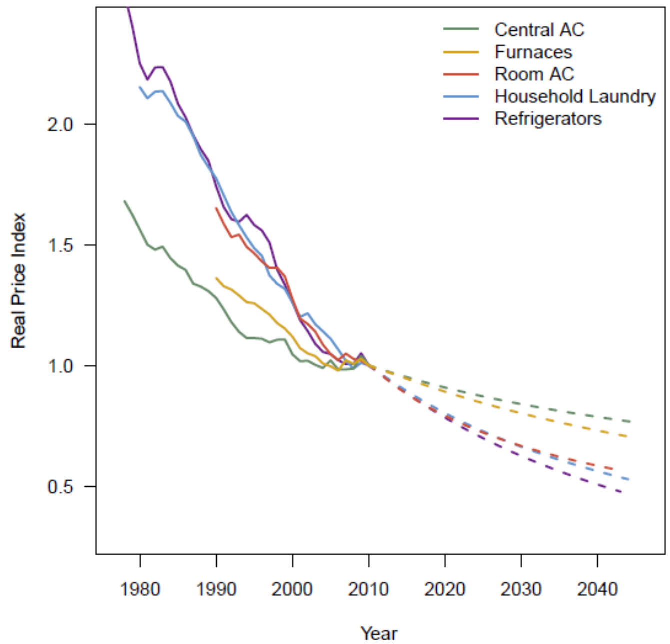

Figure 1. Real Price Index for five products analyzed by DOE.
Source: Desroches et al. 2012.

Figure 1 illustrates the decrease in *real price index* for five products analyzed in the DOE study. These curves, created by using the Consumer Price Index (CPI) to adjust the Producer Price Index (PPI) for inflation, represent the learning rate for each product, based on an index of real prices over the past several decades. To estimate learning rates for a given product using this methodology, the required inputs are cumulative shipments and average product cost (ideally, adjusted for increases in quality to account for changes in product utility or other improvements in "baseline" performance of the product type) over time, since the product's inception. Cumulative shipments are commonly available for many products through industry group databases, such as those kept by AHAM and AHRI. To estimate average product cost over time, DOE relied on the US Bureau of Labor Statistics' PPI. PPI exists for thousands of products and industries, and can be used in conjunction with cumulative shipments data to develop manufacturing cost experience curves for a variety of products. A strength of this approach is that it draws from a large body of existing research on learning rates to estimate the impact of incorporating decreasing IMC in energy efficiency standards cost-effectiveness analysis. The data is relatively easy to obtain and provides a first order approximation of the expected reduction of IMC over time. DOE's methodology is highly repeatable, and relatively simple to incorporate into future standards analysis.

However, DOE's key assumption is that the IMC decreases at the same rate as the average manufacturing cost of the product. This assumption is considered conservative, since higher efficiency products likely rely on newer technologies or higher quality components, which are expected to become less expensive at a faster pace than the rate of decrease for the product class as a whole. In other words, pricing of higher efficiency products is likely to come down more quickly than the pricing of lower efficiency products. In this way, DOE's study does not actually target IMC. To use the learning curve method to truly analyze IMC, it would be necessary to research the effect of learning on individual product components, narrowed down to the finest component pieces possible. These experience curves for individual components could then be aggregated to produce experience curves specific to products of different performance levels (for example, a unique learning curve might be generated for front loading clothes washers that is distinct from the learning curve for top loading washers, or for washers in general). This level of research would be exceptionally difficult and cumbersome with traditional data collection methods. The merits of bringing in new data collection techniques for modeling the dynamics of appliance price efficiency distributions are currently being explored by the researchers behind the DOE study (see Van Buskirk 2013), and are discussed in depth in the sections below.

Big Data

The concept of big data, or developing and analyzing increasingly large sets of data, is becoming more common as the tools for data collection become increasingly powerful through advances in information technology. The ability to collect wide swaths of data on an ongoing basis and the versatility of customizable software offers enormous potential for improving the analysis used for standards development and IDSM programs. For a given product, big data has the potential to provide an accurate assessment of the current market, and if maintained, can build a foundation to forecast trends over time. Big data can be especially useful not only for the analysis of IMC but also performance, where the objective is to understand how national or statewide trends change over time based on sales of hundreds or even thousands of different products from different retailers. One such big data collection model described below presents a method for using automated web crawler tools to track real-time price data over a substantive period.

Custom Web Crawlers

Web crawlers are specialized tools that are programmed to track specific product information on retailer websites up to several times per day. Many existing web crawler services such as CamelCamelCamel.com and PriceGrabber.com cater to consumers, tracking price trends for specific models. None of these existing tools, however, provide the exact level and precision of data that would be most valuable to an IMC and performance data analysis. Customized web crawlers can be designed to pull more granular data needed for energy efficiency measure analysis and to do so at regular intervals. In some cases, online retailers provide Application Program Interfaces (APIs) to allow interested parties easier access to collecting data from their websites without interfering with the main sites that serve typical customers. Essentially, these APIs expose underlying product databases to simplify the web crawling process. Initial tool development and ongoing maintenance costs are greatly reduced if an appropriate API is available. If an API is not provided, a web crawler must be programmed to include *screen scraping* capability to extract the appropriate data from the retail site. In such cases, it may be necessary to acquire appropriate permissions from online vendors before using the web crawler-collected data for the purposes of product price/quality trend analyses in IDSM program research and development.

At a minimum, a useful web crawler will collect the following product attributes for a large fraction of the market: retailer, brand, model number, price (including regular price and sale price), and basic product specifications. If needed, web crawlers can be designed to gather significantly more product specific or performance data. It may also be desirable to attempt to collect data regionally, which would be particularly useful for utility incentive programs or codes and standards programs. For example, with some online retailers, such as Home Depot, online prices are displayed based on the assumed zip code of the user browsing the website. Web crawlers can be programmed to search from any zip code, so it is possible to collect and compare prices from all over the country, and potentially even internationally. However, for retailers that do business primarily or exclusively online, prices generally do not vary based on location of the end user, so state-specific data would not be available.

Other factors can impact strategy for web crawler design as well. Knowledge of utility rebate programs that may be reflected in online pricing (and how these rebate offerings change over time) is important to capture in the design of a web crawler tool. Some understanding of product cycles in specific industries can improve the way data is collected and the types of analyses that are performed, as can an understanding of retailer sales strategies. For example, stores such as Walmart advertise “Every Day Low Prices” (EDLP) on a full spectrum of products being sold, while some stores generally sell at higher prices overall, but advertise special markdowns that result in cheaper prices for certain products for shorter periods. Focusing on the EDLP retailers may be more useful in terms of IMC analyses, as their prices are subject to less volatility and are expected to be more consistent from store to store, though including other stores may help display the wide range prices that may be offered in a particular market over time.

For some products, retailer websites may not specifically list product energy efficiency or other performance metrics that may be of interest from the energy efficiency measure perspective, such as power factor. In these cases it is often possible to link the product pricing data obtained by a web crawler with product performance data available through other online databases (such as the ENERGY STAR® qualified product list or other similar industry resources). To some extent, this model-matching strategy can be programmed to happen automatically, although it can be a challenge to achieve some model matches given the inconsistency with which retailers list manufacturer model numbers.

Early Application of Big Data, Web Crawlers and Experience curves

The use of big data and experience curves (with a different methodology than developed by DOE) was implemented in a recent emerging technology report for computers (PG&E 2012) and a subsequent standards proposal analysis submitted to support California’s Appliance

Standards rulemaking for computers (CA IOUs 2013). This standards analysis was one of the California Investor Owned Utilities' Codes & Standards (IOU C&S) team's first forays into big data methods, as it utilized a data collection method similar to the web crawler described in more detail below.

Price trends for discrete computer components — central processing units (CPUs) and graphics processing unit (GPUs) and hard drives (both traditional hard drive disk (HDD) and more advance, solid state drives (SSD)) — were obtained from a number of sources, including online computer component retailers (Newegg.com, BestBuy.com, Amazon.com), processor manufacturer MSRPs (e.g. Intel and AMD), and third-party computer hardware reviewers (e.g. TomsHardware.com). These components were identified as the highest priority in terms of computer energy consumption and were the most readily accessible. Power supply unit (PSU) data was also a high priority, but collected through other, non-web crawler methods. Products were released between 2006 and the present, and prices were tracked on a quarterly basis for the first 2 - 2.5 years of their release.

Once the data was collected, historical prices were used to calculate the Compounded Annual Growth Rates (CAGR) for each component (negative values) (see Table 1). In the case of hard drives, since there was a discrete delineation between the baseline (HDD) and high-efficiency (SSDs), separate CAGRs were developed. The estimated future IMC is a result of extending this CAGR to the future date when standards will become effective, and then subtracting the resulting prices between baseline and high-efficiency components, thereby representing more realistic future pricing as demonstrated by the existing literature.

Table 1. Experience curves for key computer components based on first 2-2.5 years of product availability

Component	CAGR
CPU	-10%
HDD – Magnetic, 3.5”	-11%
HDD – Magnetic, 2.5”	-20%
HDD – Solid state	-28%
PSU	-4%
GPU	-15%

Source: CA IOUs 2013

Case Study: Big Data in Action

The Pacific Gas and Electric Company (PG&E) C&S team recently developed a proposal for standards options for LED lamp quality and the CEC is currently considering the adoption of multiple performance metrics beyond energy efficiency. This measure is significantly more complicated than typical measures where the primary focus is generally limited to the relationship between price and efficiency. In this proposal, the relative impacts of various performance parameters on product pricing will be key indicators of what parts of the proposed requirements are cost-effective. This measure is also a challenge due to the high pace of change in the LED market, where new products are constantly being developed as the technology continues to mature. As such, it is an excellent opportunity to pioneer new analytical approaches and research strategies, including the application of experience curves and big data collection techniques. This case study presents work recently completed by the PG&E C&S team to analyze the relationship between performance and price for LED lamps, and describes new research underway to examine how these relationships could be changing over time.

2012 LED Lamp Price Study

In 2012, the PG&E C&S Team began work on a Title 20 standards proposal to establish minimum product quality, taking into consideration several performance metrics. To develop potential standard levels, it was important to better understand what aspects of lighting quality are most costly, and the costs to consumers associated with different aspects and levels of quality. In summer and fall of 2012, the PG&E C&S team conducted a statistical study of LED lamp prices and characteristics. The study sought to answer these specific questions for LED lamps:

- Are there any statistically significant relationships between key lighting performance metrics and price?
- Which metrics have the greatest statistical linkage with price?
- What is the estimated magnitude of the effect of influential metrics on price?

To evaluate these questions, the C&S team manually collected lamp price and performance characteristics from various online lamp vendors, constructed a model of lamp price based on performance characteristics, and conducted a multiple regression analysis to evaluate and refine the model. A high level summary of the methodology and results is provided in the following sections.

First, the C&S Team identified over 700 unique price points for over 500 unique lamp models, including omni-directional and directional lamps. Prices were identified for 247 different PAR lamps, 147 A lamps, 49 MR lamps and a smaller number of products for several other lamp shapes (BR, Candle, G, and others) through web-based research and manual data collection. The team then collected data on a large number of performance metrics for each product for which price data was collected. Table 2 shows the extent to which data was available for the targeted metrics for the three most prevalent lamp shapes. In addition to the performance metrics listed in Table 2, a note was made if the product was ENERGY STAR-qualified and if the product was marketed as dimmable or not (so this information is considered to be available for 100% of products).

Table 2. Performance data availability by lamp shape (Number of products for which data was available for each metric); 2012 manual collection methods

Attribute	A	MR	PAR	All Lamps
Watts	147	49	247	443
Lamp Shape	147	49	247	443
Lumens	127	49	247	423
Lumen Maintenance	125	49	247	421
Correlated Color Temperature (CCT)[1]	123	49	247	419
Warranty	97	27	247	371
CRI	107	49	206	362
Power Factor	28	13	230	271
Beam Angle	NA	47	89	182
Voltage	57	49	5	111
R9[2]	0	0	21	21
Chromaticity Consistency Bins[3]	0	0	21	21
Zonal Lumens[4]	0	0	0	0
Harmonic Distortion	0	0	0	0

Key:

Metric was available for 80-100% of products	(green)
Metric was available for 50-80% of products	(light green)
Metric was available for 20-50% of products	(light orange)
Metric was available for less than 20% of products	(orange)

The C&S team then conducted a multivariable regression analysis to evaluate and refine a model to predict product price as a function of lamp performance. The model that was established was a good fit to the data with an adjusted R^2 of 0.7; it had a statistically significant slope ($p<0.001$), included only individual effects with statistically significant slopes, and yielded homoscedastic (normally distributed) residuals. While the model appeared to be a good fit according to the visual distribution of residuals, no data was held back to allow cross-validation. Future efforts will include cross-validation or bootstrapping to evaluate the generalizability of the model.

[1] A dimension of chromaticity used to describe the color appearance of a light source.
[2] A metric describing the ability of a light source to accurately render objects of a deep color.
[3] A measure of the variation in light color within various samples of a specific lighting product.
[4] A measure of the luminous intensity of a light source in different directions.

2012 Regression Results Discussion

A model based on only four basic performance characteristics, lamp shape, wattage, color rendering index (CRI), and ENERGY STAR qualification status, explained 70% of the observed variability in price. The model predicted that ENERGY STAR qualification would increase lamp price by 21%, whereas each five CRI units would increase lamp price by 6% and each increase of one watt would increase lamp price by 5%. Interestingly, certain performance metrics that appeared superficially correlated with price did not demonstrate statistically significant independent effects on price when corrected for the influence of other metrics. For example, lumens did not demonstrate a significant influence on price, independent of the effect of wattage. Similarly, correlated color temperature (CCT), lumen maintenance (L70), warranty length, and power factor did not demonstrate statistically significant independent influences on price after correcting for the influence of the other factors. Likewise, improvements in other key performance metrics such as efficacy and dimmability did not appear to increase price significantly, if at all. The model resulted in the following equation that describes the impact of lamp shape, ENERGY STAR qualification status, CRI, and wattage on lamp price:

$$Price = e^{(a\times shape\,+\,b\times ES\,+c\times(CRI-CRI_{mean})+d\times(Watts-Watts_{mean})+constant)}$$

where

shape	= the shape of the product (A, PAR, MR, BR, R, G, Candelabra);
ES	= the ENERGY STAR qualification status of the product (1 or 0);
CRI	= the CRI of the product;
CRI_{mean}	= the mean CRI for all products;
Watts	= the wattage of the product;
$Watts_{mean}$	= the mean wattage of all products;

and

a, b, c, d, and constant are the parameter estimates derived from the regression model. The values of the parameter estimates are as follows:

a = -0.465
b = 0.189
c = 0.013
d = 0.045
constant = 3.598

This analysis was instrumental in understanding the interactions between price and various performance metrics and was used to inform the cost-effectiveness evaluations presented in the CASE Report. However, collecting the data used in this analysis was exceedingly tedious and time consuming. Though useful for producing a single snapshot in time of the LED lamps market, it would not be practical to manually collect this type of data on an ongoing basis.

Bringing in Big Data

In winter 2013-2014, the C&S Program developed and rolled out a retailer-based web crawler tool that utilizes both screen-scraping methods and retailer provided APIs (Application Programming Interfaces) to automatically capture product pricing data for all LED products being sold at select online retailers. Prices are gathered on a weekly basis and stored in a database that enables users to track product price fluctuations and trends over time. The project team chose the following list of retailers from which to gather product price information:

- acehardware.com
- bestbuy.com
- bulbamerica.com
- bulbs.com
- costco.com
- homedepot.com
- lowes.com
- walmart.com
- 1000bulbs.com

Price data is collected for over 40 LED replacement lamp varieties (i.e. shapes), including both directional and non-directional replacement lamps, downlights/recessed retrofit kits, and linear LED tubes. Data are also collected for a wide variety of base types including screw-base and pin-base. For certain online retailers, data are obtained using multiple zip codes to account for the fact that prices offered online may vary based on the location of the consumer. All told, every week a total of over 3,000 unique price points are collected, corresponding to more than 1,000 unique LED lamp models, offered from over 50 different manufacturers.

In addition to real-time price collection, the tool links pricing data to product performance data. Wherever possible, the tool maps the model numbers collected to those in the Lighting Facts Database to obtain lamp performance information that can be linked to the price points collected. Where a model match to Lighting Facts Database is unsuccessful, the tool relies upon performance information captured from the retailer. This linkage of product performance data allows users to view product price data not just by lamp shape and size, but by any specific performance characteristic available (for example one can compare prices of lamps of different color temperature, different beam angles, different lumen outputs, different wattages, etc.). This process essentially automates the data collection process that had been entirely manual in 2012, so that now it can be performed each week, and with a much larger set of products, in a fraction of the time.

The following two figures provide an example of the type of data that is currently being collected by the web crawler. These figures show pricing for LED replacement lamps (all lamp shapes), first by efficacy, and then by wattage, and they can help draw conclusions about market pricing with respect to these metrics. Each of these figures displays over 25,000 individual price points collected between November 2013 and April 2014.

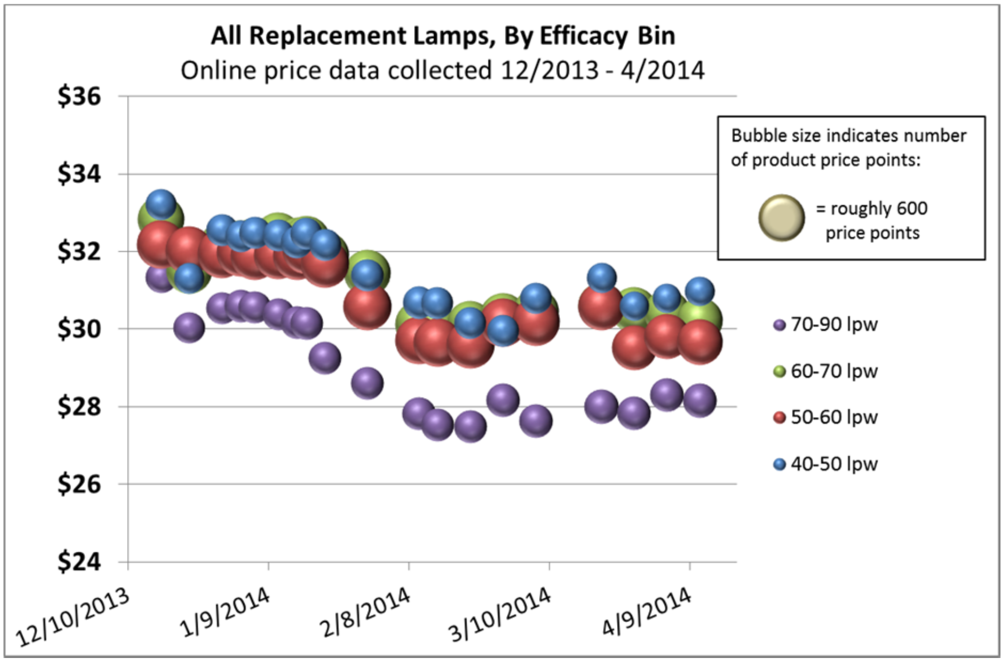

Figure 2. LED replacement lamps, average online prices by efficacy bin.

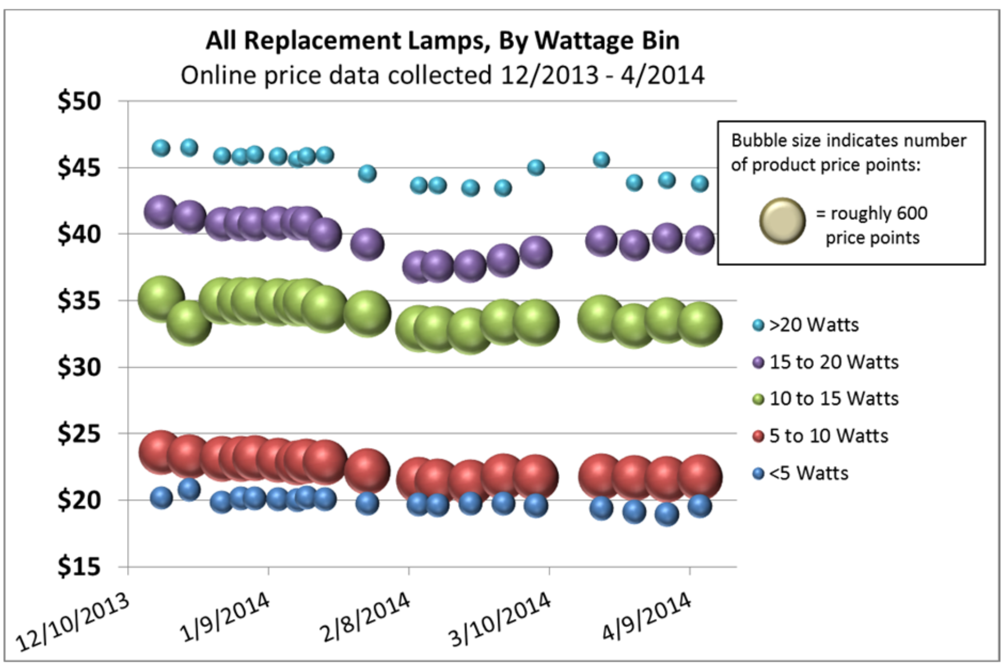

Figure 3. LED replacement lamps, average online prices by wattage bin.

2014 Regression Results

Relying on the new data collected via web crawling tools, it is now possible to conduct additional analyses to identify any statistical relationships between price and product performance based on data collected in 2014. Though the graphs above suggest a strong correlation between wattage and price, and virtually no correlation between efficacy and price, these are simple, two-variable analyses and do not necessarily prove a statistical relationship. Using the influx of big data acquired through this process, the C&S team set out to determine whether the availability of additional product characteristics could be used to build a more powerful statistical model than the model generated in 2012.

The 2012 analysis was limited by the data that had been collected, but with significantly more data to draw from, it was possible to run additional statistical tests to determine if other factors are greater influencers of price. As shown in Table 3 below, the web crawler was able to gather performance data for more lamp attributes, and for hundreds more products than were gathered in 2012. Even for performance attributes that are very seldom reported, there were generally 50 – 100 products found with data on that attribute.

Table 3. Performance Data Availability by Lamp Shape (Number of products for which data was available for each metric); 2014 Web Crawler

Attribute	A Shape	Small Diameter Directional	Large Diameter Directional	Decorative	All
Lamp Shape	245	402	1043	137	1827
Watts	245	399	1029	136	1809
Lumen Maintenance	244	396	1033	135	1808
Lumens	239	371	1033	132	1775
Efficacy (lpw)	239	370	1022	132	1763
CCT	239	382	992	132	1745
Dimmable (Y/N)	222	348	936	119	1625
CRI	227	360	955	106	1648
Input Voltage	168	356	904	95	1523
Beam Angle	NA	348	797	NA	1181
Warranty	124	243	618	87	1072
Candlepower[5]	NA	239	458	NA	700
Energy Star Qual. (Y/N)	123	122	378	70	693
Product Weight	12	7	76	0	95
Power Type (AC/DC)	20	26	33	5	84
Power Factor	17	23	4	5	49
R9	16	23	4	5	48
Duv[6]	16	19	4	5	44

[5] A measure of the illuminating power of a light source (expressed in candelas).
[6] Another dimension of chromaticity (in addition to CCT) used to describe the color appearance of a light source.

Key:

Metric was available for 80-100% of products	
Metric was available for 50-80% of products	
Metric was available for 20-50% of products	
Metric was available for less than 20% of products	

The large influx of data allowed the C&S team to improve on the 2012 model in two significant ways. First, enough data now exists to assess the impact of manufacturer/brand as an independent variable (there are more brands represented, and more price points per brand). Second, enough data now exists to develop individual models for each lamp type, rather than one model that applies to all lamp types. Initial results have yielded positive results for four distinct models – one for A-lamps, one for small diameter directional lamps (e.g. MR16s), one for large diameter directional lamps (e.g. BR30, PAR38), and one for decorative lamps. All of these models have similar or improved explanatory power from the 2012 model, and all appear to be good fits to the data. Table 4 below provides a summary of each model, the number of products included in the final run, the explanatory power of the model, and the key product attributes impacting the model.

Table 4. Summary of statistical models developed with 2014 Web-Crawler LED lamp price data

Lamp Shape	Number of Price Points in Final Model	Adjusted R^2 Value	P Value	Primary Attributes Impacting Price
A Lamps	235	0.648	< 0.0001	Brand, Watts, CRI, Efficacy
SDDL	114	0.814	< 0.0001	Brand, Lumens, Energy Star Qualification
LDDL	244	0.804	< 0.0001	Brand, Lumens, Dimmability, Warranty
Decorative	123	0.71	< 0.0001	Brand, Watts, Dimmability

The results of these models suggested that brand alone appears to be one of the biggest drivers of price for all lamp types, a conclusion not available as a result of the 2012 analysis. Below is a graphical depiction of the small diameter directional lamp model, which showed the highest adjusted R^2 value of the four. This graph includes estimated prices for a brand in the 20th percentile in terms of cost vs. another brand in the 80th percentile. Again, one of the most interesting outcomes of this model is what it says about product performance attributes that *don't* impact price in a significant way. In this case, efficacy, lifetime, CRI, dimmability, warranty, beam angle and many other metrics were not found to improve the statistical significance of the model.

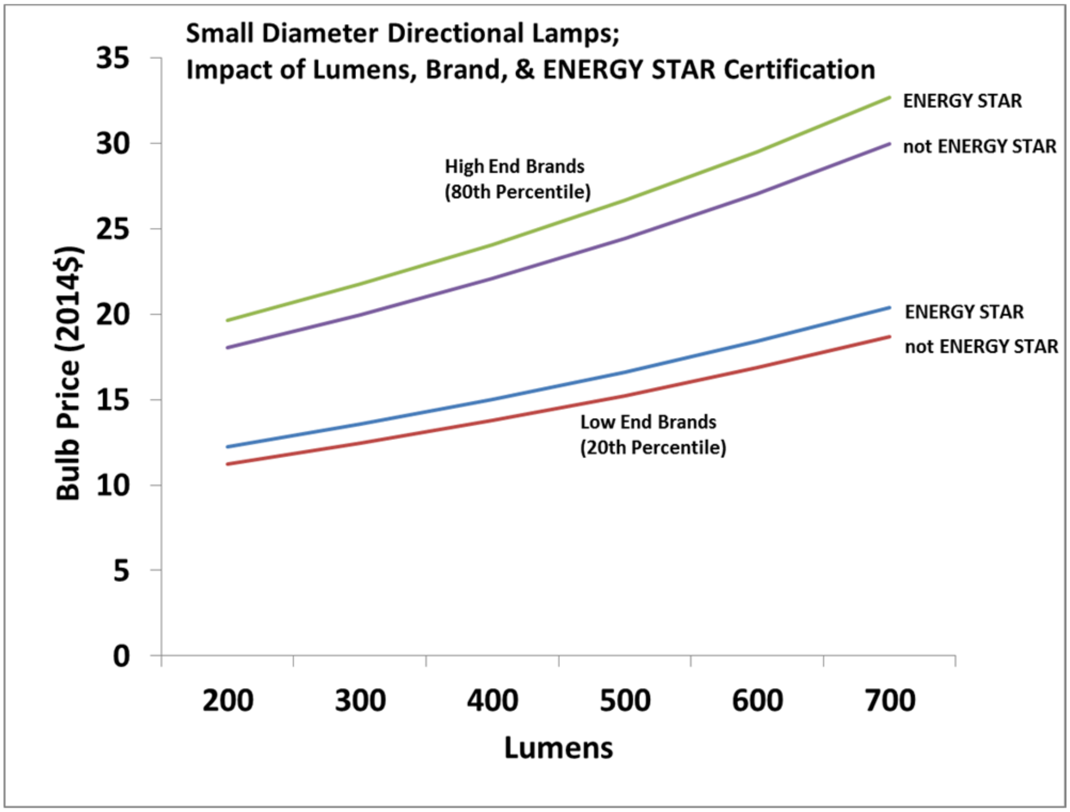

Figure 4. Graphical representation of the statistical model developed to estimate pricing for small diameter directional lamps; Based on 2014 Web-Crawler Data.

Additional Analyses Enabled by Influx of Big Data

The research completed in 2012 and the analysis described above offer singular snapshots in time of price-performance relationships for LED lamps. As described above, the C&S team performed a similar analysis using the 2014 data to attempt to understand how the market has changed. By creating separate regression models at specific intervals of time and observing how the explanatory power of the variables in the model change with each iterative model, it may be possible to uncover trends among each variable in the model. With additional datasets to draw from, these trends may become clearer. Moreover, collecting data weekly and on an ongoing basis could also allow the team to include time itself as an independent variable in a regression analysis. This could potentially allow for statistical models to forecast how relationships between price and performance are expected to change in the future. Another approach for factoring time into the analysis is to use the "date added" as another field in the regression analysis. In other words, prices and product features are time-stamped to the date on which the product was first detected by the web crawler or the date it was first added to the Lighting Facts Database. This would allow the analysis to distinguish between products that have been on the market for several years from those that were only recently introduced, and also to determine whether the length of time the product is on the market has a statistically significant impact on price.

Conclusion

Big data coupled with advanced analytics and algorithms has great potential for optimizing energy standards and IDSM program development. As previously discussed, recent research has shown that existing methods for measure analysis have not adequately accounted for trends in manufacturer learning and other variables leading to price declines; this in turn results in sub-optimal programs and standards that do not realize the full extent of cost-effective savings that might otherwise be achievable. Recent efforts by energy efficiency advocate organizations, DOE, and other researchers have made significant inroads toward developing reliable methods for forecasting future price-efficiency relationships and price-performance relationships. As presented in this paper, the time to develop these models can be reduced, and their accuracy can be improved, through the analysis of larger, more comprehensive datasets.

In addition to helping to create more accurate forecasts of future product price-efficiency and price-performance relationships, big data may also be used to develop estimates of naturally occurring market adoption (NOMAD) of higher efficiency products, which has historically been a very challenging task. The same techniques being applied to create forecasts of manufacturer learning, and the resulting impacts on the incremental cost of higher efficiency products, can be used to predict general trends in product performance. Again, leaning on the power of increasingly large datasets, collected on an ongoing basis, it will be possible to develop more accurate models for projecting how product efficiency, or any other performance metric, improves over time. As described in the LED lamps case study presented above, a number of methods are currently being explored for using statistical models to analyze large, web crawler-compiled datasets of product cost and performance.

An accurate projection of incremental measure cost and expected NOMAD is crucial to developing effective energy codes and standards and IDSM programs with optimized savings. When information is lacking, and future trends in product development are not accounted for, suboptimal programs can leave significant energy and consumer cost savings on the table. The ideas discussed in this paper offer a methodology for taking advantage of web crawling technology to build increasingly powerful datasets that can be analyzed to enhance the effectiveness of future IDSM programs.

References

CA IOUs (California Investor Owned Utilities). 2013. Codes and Standards Enhancement Initiative for PY 2013. Title 20 Standards Development: Analysis of Standards Proposal for Computers. Docket #12-AEER-2A.

Dale, Larry, Camille Antinori, Michael McNeil, James McMahonand K. Sydny Fujita. 2009. "Retrospective Evaluation of Appliance Price Trends." Energy Policy 37. (p. 597-605).

Desroches, Louis-Benoit, Karina Garbesi, Colleen Kantner, Robert Van Buskirk, Hung-Chia Yang. 2012. "Incorporating Experience Curves in Appliance Standards Analysis." Lawrence Berkeley National Laboratory. September 4.

DOE (U.S. Department of Energy). 1989. Technical Support Document: Energy Efficiency Standards for Consumer Products: Residential Refrigerators, Refrigerator-Freezers and Freezers. Washington, DC: U.S. Dept. of Energy.

Economist, The. 2009. "The experience curve." September 14. http://www.economist.com/node/14298944.

Mauer, J., deLaski, A., Nadel, S., Fryer, A. and Young, R. 2013. "Better Appliances: An Analysis of Performance, Features, and Price as Efficiency Has Improved." Report A132. May 2013. http://www.appliance-standards.org/sites/default/files/Better_Appliances_Report.pdf.

Mooz, William E. 1978. "Cost Analysis of Light Water Reactor Power Plants." Prepared for the Department of Energy. R-2304-DOE. June 1978. http://www.rand.org/content/dam/rand/pubs/reports/2008/R2304.pdf.

Nadel, Steven. 2000. "Comments of the American Council for an Energy-Efficient Economy on DOE's Proposed Minimum Efficiency Standards for Residential Central Air Conditioners and Heat Pups". Presentation to DOE, November 16. Washington, DC: American Council for an Energy-Efficient Economy.

PG&E (Pacific Gas and Electric Company). 2012a. Cost-Effective Computer Efficiency. Prepared by Ecova. http://www.etcc-ca.com/reports/cost-effective-computer-efficiency.

Van Buskirk, Robert. 2013. "Modelling the dynamics of appliance price-efficiency distributions." ECEEE Summer Study proceedings.