

Approaches for Identifying Consumer Preferences  
for the Design of Technology Products: A Case  
Study of Residential Solar Panels

by

Heidi Qianyi Chen

B.S., Cornell University (2010)

Submitted to the Department of Mechanical Engineering  
in partial fulfillment of the requirements for the degree of

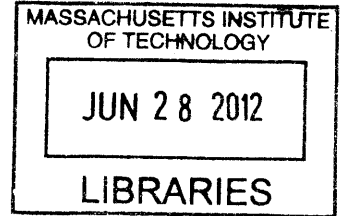
Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

© Massachusetts Institute of Technology 2012. All rights reserved.



ARCHIVES

Author .....  
Department of Mechanical Engineering  
May 17, 2012

Certified by .....  
Maria C. Yang  
Assistant Professor of Mechanical Engineering & Engineering Systems  
Thesis Supervisor

Accepted by .....  
David E. Hardt  
Professor of Mechanical Engineering  
Chairman, Department Committee on Graduate Students



# Approaches for Identifying Consumer Preferences for the Design of Technology Products: A Case Study of Residential Solar Panels

by

Heidi Qianyi Chen

Submitted to the Department of Mechanical Engineering  
on May 17, 2012, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Mechanical Engineering

## Abstract

This thesis investigates ways to obtain consumer preferences for technology products to help designers identify the key attributes that contribute to a product's market success. A case study of residential solar PV panels is conducted in the context of the California, USA market within the 2007-2011 time span. First, interviews are conducted with solar panel installers to gain a better understanding of the solar industry. Second, a revealed preference method is implemented using actual market data and technical specifications to extract preferences. The approach is explored with three machine learning methods: Artificial Neural Networks, Random Forest decision trees, and Gradient Boosted regression. Finally, a stated preference self-explicated survey is conducted, and the results using the two methods compared. Three common critical attributes are identified from a pool of 34 technical attributes: power warranty, panel efficiency, and time on market. From the survey, additional non-technical attributes are identified: panel manufacturer's reputation, name recognition, and aesthetics. The work shows that a combination of revealed and stated preference methods may be valuable for identifying both technical and non-technical attributes to guide design priorities.

Thesis Supervisor: Maria C. Yang

Title: Assistant Professor of Mechanical Engineering & Engineering Systems



## Acknowledgments

Graduate school would not have been possible without my amazing advisor. Maria, thank you for giving me a chance to get into the world of design and research even though I did not have any experience coming in. Thanks for all the time you've set aside for our (almost) weekly meetings. You have been so supportive, flexible and encouraging, walking me through all the new demands of grad school. Thank you for being a wonderful advisor, and above all, a wonderful person.

Many thanks also go to Tomonori Honda, for being so amazing at what he does, and for patiently teaching me all the statistics and machine learning methods that are present in this thesis. Thanks for all the early morning Skype meetings and late night emails that have gone into making these papers work.

To the people in Ideation: Justin, Jesse, Geoff, Anders, Lisa, Bo, Francisco, Catherine and Mieke. Thanks for making grad life so much fun, and for all the random conversations and helpful advice that you've so generously given. I will miss you.

My appreciation also goes to the Joseph Huntington Memorial Fellowship and the Center for Scalable and Integrated Nanomanufacturing (SINAM), for supporting my work and keeping me warm, fed and clothed for these two years.

Lastly, thanks to my loving family, my boyfriend Scott, SP small group, Leslie Regan and all the friends I've made here at MIT: It would not have been possible without you.



# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
<b>2</b>	<b>Related Work</b>	<b>15</b>
2.1	Stated and Revealed Preferences . . . . .	15
2.2	Residential Solar Photovoltaic (PV) Panels . . . . .	17
2.3	Gap in Literature . . . . .	18
2.4	Notes . . . . .	18
<b>3</b>	<b>Interviews with Solar Panel Installers</b>	<b>21</b>
3.1	Core Questions . . . . .	21
3.2	Methodology . . . . .	22
3.3	Results . . . . .	22
3.3.1	Installation Process . . . . .	22
3.3.2	Choice of PV Panels . . . . .	23
3.3.3	Factors that Limit Adoption of PV in Residential Homes . . . . .	24
3.4	Conclusions . . . . .	25
<b>4</b>	<b>Revealed Preference Analysis from Technical Attributes</b>	<b>27</b>
4.1	Dataset . . . . .	27
4.1.1	Data Filtering . . . . .	28
4.1.2	Support Vector Machine to Find Cutoff Point . . . . .	28
4.1.3	Variance Inflation Factor Calculation to Reduce Multicollinearity . . . . .	30
4.2	Methodology . . . . .	31

4.2.1	Critical Attribute Determination . . . . .	31
4.2.2	Artificial Neural Network Regression . . . . .	34
4.2.3	Random Forest Regression . . . . .	36
4.2.4	Gradient Boosting Machine . . . . .	37
4.3	Results: Critical Attributes . . . . .	38
4.4	Comparison of Methods . . . . .	41
4.5	Conclusions . . . . .	42
<b>5</b>	<b>Stated Preference Analysis from PV Installer Survey</b>	<b>45</b>
5.1	Survey Objectives . . . . .	45
5.2	Methodology . . . . .	46
5.3	Survey Respondent Demographics . . . . .	48
5.4	Findings . . . . .	53
5.4.1	Importance Rankings of Technical Attributes . . . . .	53
5.4.2	Relative Importance of Technical and Non-Technical Attributes	55
5.4.3	Aesthetic Preferences . . . . .	59
5.4.4	Reputation and Service Quality . . . . .	61
5.5	Conclusions . . . . .	66
<b>6</b>	<b>Discussion</b>	<b>67</b>
6.1	Limitations of Study . . . . .	70
6.2	Future Directions . . . . .	71
<b>A</b>	<b>Interview and Survey Questions</b>	<b>73</b>
A.1	Interview Questions for Photovoltaic Panel Installers in Boston, MA .	73
A.1.1	Job Overview . . . . .	73
A.1.2	Company . . . . .	73
A.1.3	Customer Decision Process . . . . .	73
A.1.4	Homeowners . . . . .	74
A.1.5	Installer perspective . . . . .	75
A.2	Online Survey Questions for Californian Installers . . . . .	76



# List of Figures

4-1	Cumulative market share of panels . . . . .	29
4-2	SVM classification performance for choice of cutoff point . . . . .	30
4-3	Flowchart of methodology . . . . .	33
4-4	MSE and R fitting over 10 attributes using 1-8 neurons . . . . .	34
4-5	ANN bootstrapping error validation . . . . .	35
5-1	Gender split of survey respondents . . . . .	49
5-2	Age of survey respondents . . . . .	49
5-3	Years of experience in the solar industry . . . . .	50
5-4	Years in current job position . . . . .	50
5-5	Area of involvement with client . . . . .	51
5-6	Client PV sector . . . . .	51
5-7	Manufacturer affiliation . . . . .	52
5-8	Importance ranking of technical attributes . . . . .	54
5-9	Relative importance of technical and non-technical attributes: Open ended question . . . . .	56
5-10	Relative importance of technical and non-technical attributes: Rating question . . . . .	58
5-11	Aesthetic preferences for PV panel categories . . . . .	60
5-12	Name recognition of panel manufacturers . . . . .	62
5-13	Reputation of panel manufacturers . . . . .	62
5-14	Service level of panel manufacturers . . . . .	63



# List of Tables

4.1	Attribute definition list. (A) Initial specs used for SVM analysis (B) Expanded specs used for VIF analysis (C) Final reduced specs used for 3 regression methods . . . . .	32
4.2	Correlation table for important attributes found by ANN . . . . .	35
4.3	Top attributes' Random Forest variable importance values over 100 runs	37
4.4	Top attributes' GBM relative influence values over 100 runs . . . . .	38
4.5	Important attributes across the 3 methods . . . . .	39
4.6	R <sup>2</sup> values for the 3 methods . . . . .	41
5.1	PV panel appearance groups . . . . .	59
5.2	Aesthetic ranking values from survey . . . . .	60
5.3	Manufacturer reputation survey ratings, rating of 3 is excellent, 1 is poor	64
5.4	Manufacturer service level survey ratings, rating of 3 is excellent, 1 is poor . . . . .	65



# Chapter 1

## Introduction

Within a firm there is a constant need to rise above the competition and gain market success. In order to achieve this goal, firms are constantly trying to find ways to appeal to the customer by determining their changing needs, wants, values and behavior and designing for them [1]. However, in the fast paced world of product development, there is a limit on the time and resources that can be allocated to product development. Hence, the identification of key product attributes that contribute to a product's market success is crucial, especially in the early stages of product development where the conceptual design phase can account for a large percentage of the overall manufacturing cost [2]. It is in the interest of both designers and stakeholders to know how to optimally allocate resources in order to increase the likelihood of market success.

This is especially so for technology-push products that are still within the early part of the technology diffusion S-curve, where there is slow uptake of the technology and product features have not fully matured [3]. Since the technology is still considered new, only early adopters have warmed to the product, and there is potential for much feature and market expansion. Knowing what the customer wants at this stage is necessary for the product to bridge the chasm between the early adopters and the early majority, allowing the product to flourish in the market [4].

This thesis proposes an approach to extracting consumer preferences by determining critical attributes using the established revealed preference framework [5, 6, 7, 8],

and drawing on advances in computational intelligence and machine learning to support the analysis. This is compared with the results from a more traditional stated preference survey method.

The main research questions are:

1. Can revealed consumer preferences be obtained from market data and engineering specifications using machine learning methods?
2. Is there agreement among the machine learning methods that suggest the validity of the data and methods?
3. How does the revealed preference approach compare with a stated preference self-explicated survey method?

We present a case study of residential solar photovoltaic panels in the California market to illustrate our methodology.

This thesis starts with a literature review of the area of stated and revealed preferences in Chapter 2. Next, in Chapter 3, interviews were conducted with residential solar PV installers in the Boston, MA area to get a first hand understanding of the solar industry and the processes involved in setting up a PV system in a household. We then use a combination of engineering specification data obtained from solar panel data sheets and real market data from the California Solar Initiative database to extract revealed preferences using machine learning methods in Chapter 4. A stated preference survey was conducted with Californian solar installers, and a comparison of results from stated and revealed preference methods made in Chapter 5. Finally, Chapter 6 serves as a summary and conclusion of our findings.

# Chapter 2

## Related Work

### 2.1 Stated and Revealed Preferences

Much work has been done within the academic community to determine consumer preferences using choice modeling. These can be broken down into 2 main categories: stated preference methods which measure consumers' explicit preferences over hypothetical alternatives, and revealed preference methods which extract preferences from actual market data [9].

Over the years, stated preference methods have gained ground in the marketing community due to their flexibility and ease of implementation. Popular survey based stated preference methods include self-explicated methods like Kelly's repertory grid [10, 11], Self-Explicated Method (SEM) [12] and the Single Unit Marketing Model [13] among others, requesting consumers to rank or rate various product attributes. Another group of stated preference methods where relative preferences are obtained include MaxDiff [14], and conjoint analysis [15, 16], which ask consumers to choose between different products which have varying attributes. Multiple hybrid models that incorporate both self-explicated and relative preferences also exist. Non-survey based methods include focus groups and field observations, which require considerable time, expertise and resources to carry out, and may be hard to quantify.

The potential problem with these stated preference methods is that consumers often exhibit preference inconsistencies, constructing their preferences along the way,

or changing their preferences due to some shift in the phrasing of the questions [17]. Research on the accuracy of consumers' predictions show a disconnect between preferences obtained during preference elicitation and actual decision making [18]. Stated preference methods have also come under considerable criticism because of the belief that consumers react differently under hypothetical experiments compared to when they are faced with the real market situation [19, 20].

In comparison, revealed preference methods could be a better reflection of purchase behavior than stated preference methods as they take into account external factors like third party influences that might affect the consumer's decision. This has been expressed in the economics and decision making literature to be especially important if the consumer's choice is based heavily on the recommendation of a more experienced expert, as a result of complexity inherent in the product, or limited personal experience [21]. However, revealed preference methods have been difficult to implement due to several factors. These include the high cost of collecting large sets of relevant data, limited technological knowledge, problems with multicollinearity, and the inability to test new variables [22]. As technology has improved and computer processing has become increasingly fast, efficient and cost effective, it has become feasible to reevaluate these methods. Furthermore, more companies are keeping digital records of product sales, making data collection less of a burden than before. Machine learning methods that are capable of dealing with multicollinearity involving regression and classification can now be applied on large sets of marketing data, overcoming the issue with multicollinearity that several academics have identified, and allowing for the identification of key attributes in an efficient way. Finally, the inability to test new variables still poses a significant challenge, as the new variables may be outside the data range, and involve extrapolation outside the range used to create the model. This can be dealt with by a careful use of stated preference methods in combination with the revealed preference framework.

Revealed preference methods have been used widely in economics research, but little has been done in the area of design applications. Similar work that has been done in the joint field of product design and machine learning include: Agard and Kungsiak's



work on data mining for the design of product families [23], where algorithms were used for customer segregation; Ferguson et al's work on creating a decision support system for providing information from later to earlier stages in the design process [24]. A good overview of other applications of computational intelligence in product design engineering can be found in Kusiak's 2007 review [25].

## 2.2 Residential Solar Photovoltaic (PV) Panels

In recent years, the US government has been encouraging the development of renewable power, providing the solar industry with increased funding for the development of solar panels for residential, commercial and utility deployments. Residential installations in particular have gained attention due to their use of otherwise "dead" space, utilizing area on rooftops or facades for the panels. Generous subsidies and rebates have been put into place in order to encourage homeowners to adopt the technology. Despite this, the industry is still considered by many to be in the early stages of the technology diffusion S-curve, with few homeowners choosing to purchase PV systems for their properties.

This thesis proposes the view of considering solar panels as a product rather than a technology. Products differ from technology in that they may be described by both qualitative and quantitative characteristics, and are designed to appeal to consumers. Much of the current academic engineering focus on solar panels has rightly been on the physical science and technology behind the application, improving the performance of the conversion of sunlight to electricity, assessing and increasing the reliability and durability of the system [26]. This is critical for spurring increases in the demand for large scale facilities installations. At the same time, it is important to convince consumers to purchase a PV system at the residential level where decision makers are spread out and individual households have different requirements. There is limited academic literature on understanding consumer needs in order to increase adoption. Existing research is centered on identifying characteristics of adopters [27], discovering their motives for acquiring a PV system [28], determining barriers to adoption [29],

and understanding the link between attractive factors of PV systems [30]. However, these studies are limited to stated preference studies, and do not include real market data or technical specifications.

At this moment, the industry is also facing an oversupply condition, a result of an increase in global manufacturing with little corresponding increase in the demand on the side of the consumer. Especially now that the survival of companies are at stake, funding is tight and profits are low, there is a need to focus available resources on high priority features that will lead to more consumer purchasing to increase the firm's market share and maintain profits. The state of the industry thus lends itself to our study.

## 2.3 Gap in Literature

This thesis sets itself apart in the design community by focusing on revealed preferences instead of stated preferences as a means to extract consumer purchasing preferences. Compared to existing data mining methods, we take data from widely available sources instead of from within the company, combining real market data and engineering specifications from data sheets in order to determine a set of critical attributes that can be prioritized to boost a firm's competitiveness. The result of machine learning algorithms are also compared to validate their effectiveness.

## 2.4 Notes

For this thesis we make use of market share as a reflection of market success, even though the definition of market success varies widely in literature [31]. Market share was chosen as it is publicly available, unlike customer satisfaction levels, revenue or profits which are usually kept within the company and are difficult or costly to measure. It has also been discovered to be the most useful customer-based measure for the success of line extensions of existing products [32].

In our discussions, we make a distinction between technical and non-technical

attributes. We define technical attributes as quantitative attributes that affect engineering design, like dimensions, weight, and efficiency. They are listed in engineering data sheets, and can be easily measured. Non-technical attributes are qualitative attributes that are difficult to measure because they are affected by subjective judgments. These non-technical attributes include aesthetics, company relationship, reputation, and name recognition, amongst others.

The aim of this thesis is to see if there is a correlation between photovoltaic panel attributes, and their success in the market, measured by market share. In this way, designers will be able to better optimize design priorities that lead to product success.



# Chapter 3

## Interviews with Solar Panel Installers

At the core of this thesis is the notion that solar panels are products that are designed with the customer in mind, a practice known as user-centered design. Hence, in order to gain a better understanding of the solar industry, in depth interviews with solar photovoltaic (PV) installers in Massachusetts were conducted at the start of the research cycle.

### 3.1 Core Questions

The purpose of the interviews was to build up a base understanding of the PV industry and to determine if it makes sense to look at design methods for solar PV panels in the early stage of the design phase. Core questions included:

1. What is the process that homeowners have to go through in order to install solar panels on their homes?
2. What are the factors that limit adoption of solar panels?
3. Who makes what decisions during the installation process?
4. What are the attributes that are important when choosing between solar panels?

## **3.2 Methodology**

Interview questions were drafted and can be referred to in Appendix A.1. Solar PV installers in the Boston, Massachusetts area were sourced, and contacted by email. Installers were chosen over homeowners and manufacturers because they deal with both parties, and are present from the point of the homeowner decision to try renewable energy to the completed installation. Out of the 22 installers contacted, 2 agreed to be interviewed in person. Interview times and locations were set up to suit the interviewees. Each interview took approximately 1 hour.

## **3.3 Results**

Responses from the two installers have been condensed and summarized below.

### **3.3.1 Installation Process**

The process toward a solar installation for a typical homeowner starts with an email or telephone call. The installer looks at satellite imagery of the house, and talks with the homeowner about his or her electrical bill, and the approximate cost of setting up a PV system. If the homeowner is still interested, within a week, the installer visits the site to check its suitability for a solar installation. This involves measuring the amount of light the roof receives, calculating the percentage of the roof that is obstructed from direct sun by tall trees or other buildings. Subsequently, the installer crafts a proposal of the installation, including the choice of PV panels and inverter, installation layout and other related factors. Typically, about 1 month passes between the site visit and the homeowner approving the proposal. The installer then applies for state rebates and subsidies, a process that can take up to 60 days for approval.

Upon approval of the rebate application, it takes about 2 weeks to obtain the materials needed and schedule the installation. The actual installation process only takes about 2 days. However, the process does not end there. Inspections by various town building and electrical departments to produce permits need to be scheduled

and carried out. The whole process typically takes about 4-5 months.

### 3.3.2 Choice of PV Panels

Most homeowners know little about PV technology, and rely on installers to make educated decisions regarding the details of the installation. Hence, the choice of the PV panel to use actually lies with the installer instead of the homeowner.

*“Homeowners mostly call saying: I have no idea if this is a good idea, can you tell me more?”*

*“We do all the recommendations. Homeowners usually don’t come in with a knowledge base. They come in with needs, with homes that are a certain way: I have a standing seam metal roof and I want something that is easy to install and looks good..”*

*“I tell them: Let the professionals handle it. It is too hard to be a solar expert.”*

Central considerations for the homeowner include:

1. Size of the system

*“... balance cost and how much [the PV system] going to make, so it produces most of the power needed to cover the electric bill”*

2. Return on investment

*“There are the financially conscious ones, who know their electricity consumption by month for the last few years, are good at figuring out investment returns, ROI, payback.”*

3. Aesthetics

*“Aesthetics are central to homies (customers who care about home improvement)... who love to have nice homes, the latest and greatest.”*

*“What do the panels look like? Black frames versus silver. Even within that, some panels just look better than others and have a different aesthetic”*

*“... need to make it integrate with the home, work around the roof”*

4. Country of origin, reputation of PV panels

*“Where is this made? Does it have a good reputation?”*

Central considerations for the installer include:

1. Ease of installation

*“Easy to install is key. Make installation fast and easy.”*

2. Cost of panel

3. Relatively good efficiency

*“Making panels cheap, relatively easy to install, relatively good efficiency. That will be huge. That will be really huge.”*

4. Good relationship with panel manufacturers

*“Warranty support, deliverability, experience, willingness to work with company, avoiding channel conflict.”*

### **3.3.3 Factors that Limit Adoption of PV in Residential Homes**

1. It is difficult to encourage homeowners to install a PV system in their homes; there needs to be a lot of patience and understanding to work through issues.

*“It takes the same amount of time to sell a 5kW as it does a 2MW, homeowners take time to decide.”*

2. Technology needs to improve to take advantage of homes with more sunlight obstruction.



*“Not every house is oriented the same way, can’t do every house. There’s trees, landscaping, shading.”*

*“Increasing the potential number of sites available through better conversion.... increasing the number of viable homes with trees and landscaping...”*

### 3. Instability of politics, legislation

*“Politics and legislation affects subsidies and incentives.”*

*“It is important to make incentives more stable”*

## 3.4 Conclusions

By interviewing PV installers, we learned that decision making regarding PV panel choice is made predominantly by PV installers, not homeowners. In most cases, homeowners simply take the recommendation of the installers. This decision is a complex one made by weighing many attributes, including both technical attributes (e.g. efficiency, cost, size) and non-technical attributes (e.g. installer relationship with manufacturer, aesthetics).



# Chapter 4

## Revealed Preference Analysis from Technical Attributes

Armed with knowledge from the interviews presented in Chapter 3 that the purchase decision is made by weighing various attributes of the PV panels, we move on to determine which technical attributes are important in this decision. Since technical attributes can be measured and designed for, companies are particularly interested in them, as improving these attributes may result in better performance on the market.

In this chapter, we perform a revealed preference analysis to extract preferences from a combination of market data and engineering specifications taken from the PV industry in California. The approach is explored using three machine learning methods: Artificial Neural Networks, Random Forest decision trees, and Gradient Boosted regression.

### 4.1 Dataset

The working dataset published in September 7, 2011 from the California Solar Statistics California Solar Initiative incentive application database [33] served as the thesis' source of market data. The data is considered representative of the USA solar consumption, as California is the current leading producer of solar power in the United States, accounting for 44% of the total grid-connected PV cumulative installed ca-

capacity through quarter 3 of 2011 [34]. The working database includes all incentive applications from January 2007 to November 2011 made in California, hence includes both successful subsidized installations and unsuccessful incentive applications made by a variety of consumers. It was assumed that unsuccessful incentive applications did not result in a PV installation.

#### **4.1.1 Data Filtering**

The data was filtered to include only residential installations with a completed installation status, excluding applications that are from the commercial, government or non-profit system owner sector, as well as those that were canceled or pending. This was done in order to concentrate on the small scale PV systems that were actually installed during the 2007-2011 time-frame. Installations with more than 1 PV module type were filtered out, as the effective cost calculations cannot be done. Finally, new panels introduced during the past year were removed, as they are still too new and the market has not had adequate time to respond. After filtering, the data set was reduced from 73,514 to 32,896 installed systems with a total of 586 panel types, mostly due to filtering out non-residential systems. Filtering out systems with more than 1 PV panel type accounted for less than 0.8% of the total number of systems, and the effect of neglecting them in the subsequent calculations was taken to be negligible.

#### **4.1.2 Support Vector Machine to Find Cutoff Point**

From this dataset, the quantity installed of each panel was calculated as a proxy for market share, and the panels ranked by that metric. Since a large portion of the market is controlled by a small subset of the 586 panels, as shown in Fig. 4-1, further analysis was required to find a cutoff point to focus the further analysis on the panels that are considered the most successful in the open California market. An established binary linear classifier called Support Vector Machine (SVM) [35], was chosen. 200 panels picked at random were sourced for their technical specifications from manufacturer data sheets to form a specifications dataset. This dataset included

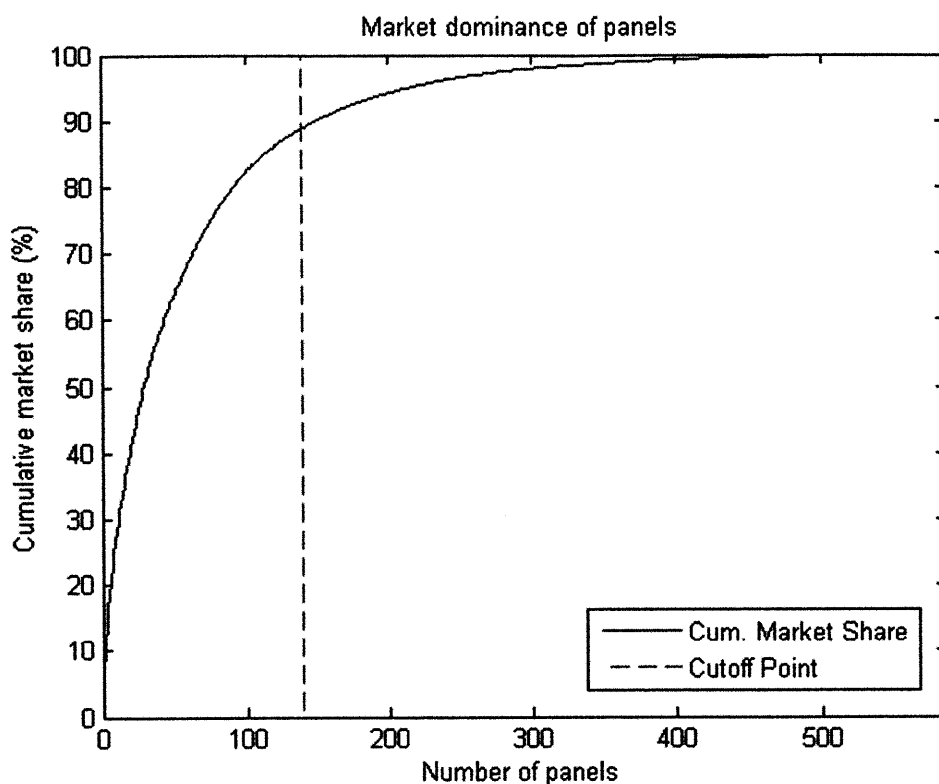


Figure 4-1: Cumulative market share of panels

common markers for technical performance, including attributes like efficiency tested at Standard Test Conditions (STC), rated power, and power warranty. These 22 markers were chosen as initial distinguishing attributes of the technical performance of each panel, as shown in Tab. 4.1A.

SVM took the set of PV panel data and categorized the panels into 2 groups by multiple attributes, including the 22 attributes stated in Table 4.1A and the panel's market share. A linear kernel was applied as it best suited the data spread. Figure 4-2 shows a noticeable drop-off in SVM classification performance at the top 140 panels, so that was chosen to be the cutoff point. This subset was determined to control 88.9% of the California market.

In the same way as the previous step, the 140 panels with the highest market share were identified by their model number and sourced for their technical specifications. From a combination of panel datasheets and marketing material, an expanded

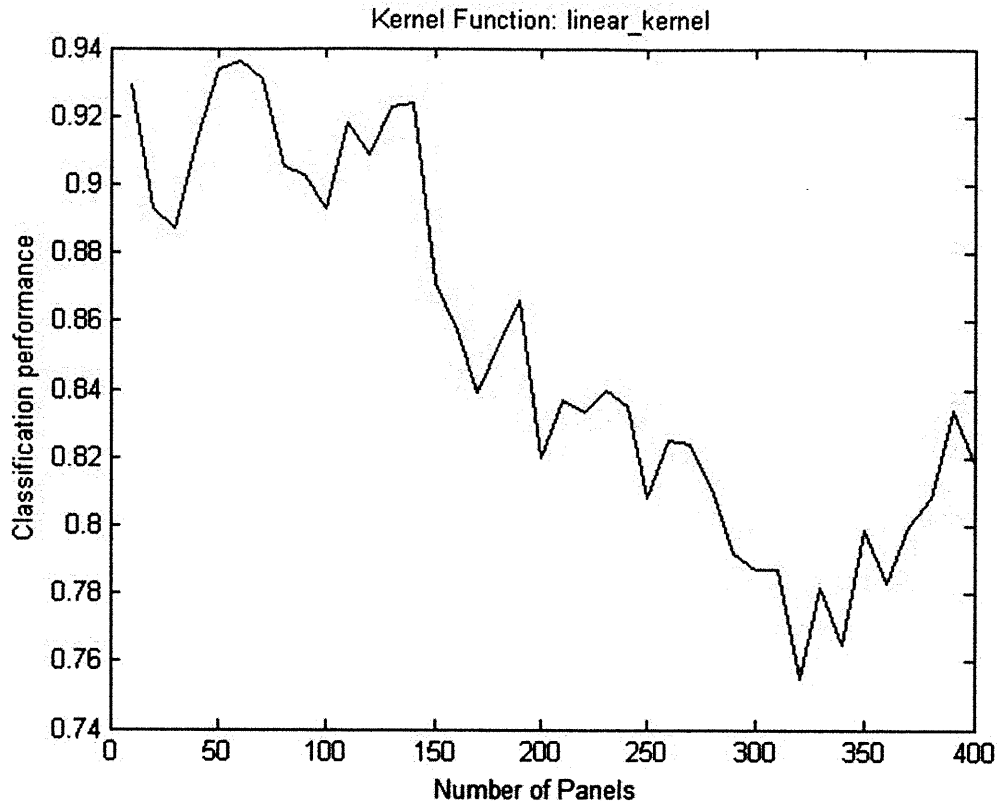


Figure 4-2: SVM classification performance for choice of cutoff point

list of 34 attributes was identified (Tab 4.1B). This expanded list adds distinguishing characteristics of the panels, like appearance, packaging and environmental characteristics to the initial 22 attribute list, and is a more comprehensive collection of technical attributes.

### 4.1.3 Variance Inflation Factor Calculation to Reduce Multicollinearity

As expected, the expanded list of attributes exhibited a high degree of multicollinearity, meaning that the attributes were highly correlated. This is a problem as it decreases the accuracy of the model. To reduce parameter correlation between the attributes and improve the multiple regression model, the redundant attributes were identified using a Variance Inflation Factor (VIF) calculation, which quantifies the

severity of multicollinearity in an ordinary least squares regression analysis. This method was chosen because of the ease of comparing multicollinearity between attributes. The VIF for each attribute was calculated using Eqn. 4.1 by holding it as the dependent variable and performing regression with the rest of the attributes as independent variables.

$$VIF_k = \frac{1}{1 - R_k^2}, \quad R_k^2 = 1 - \frac{MSE_k}{\sigma_k^2} \quad (4.1)$$

where  $k$  is the attribute number from 1-34,  $MSE$  is the mean squared error,  $R^2$  is the coefficient of determination, and  $\sigma$  is the standard deviation. Attributes with high VIF values of more than 20 were removed from the specifications list [36], as shown in Tab. 4.1C. A total of 8 attributes were removed, leading to a reduced list of 26 attributes.

## 4.2 Methodology

### 4.2.1 Critical Attribute Determination

An overview of the methodology is presented in Fig. 4-3. A set of 3 computational machine learning regression methods were used to determine the important technical attributes that most influence market share. These methods were chosen over others as they are known in the machine learning community to be robust. However, other methods like SVM regression and Elastic Nets could have been used to achieve the same purpose.

1. *Artificial neural network (ANN) regression*

ANN regression is a non-linear statistical data modeling that models complex relationships between inputs and outputs in a network of synapses and neurons. [37].

2. *Random Forest regression*

Random Forest regression is an ensemble of unpruned regression trees created by

Table 4.1: Attribute definition list. (A) Initial specs used for SVM analysis (B) Expanded specs used for VIF analysis (C) Final reduced specs used for 3 regression methods

Properties	Specifications	A- SVM	B-VIF	C- Reg.	Definition
Electrical	Rated power (W)	x	x		Power output of panel at STC
	Power variance (-%)	x	x	x	Negative power output tolerance
	Power variance (+%)	x	x	x	Positive power output tolerance
	Power at PTC (W)	x	x		Power output of panel at PTC
	Power ratio: PTC/STC	x	x	x	Power output ratio: PTC/STC
	Efficiency at STC (%)	x	x	x	Panel efficiency at STC
	Fill factor at STC		x	x	Fill factor of panel at STC
Physical	Weight (kg)	x	x		Total weight of panel
	Weight per W (kg/ W)	x	x	x	Weight of panel per Watt of rated power output
	Area of panel (m <sup>2</sup> )	x	x		Area of panel
	Cell Number	x	x	x	Number of PV cells in panel
	Frame color (black/ not black)	x	x	x	Color of panel frame
	Thickness (mm)	x	x	x	Thickness of panel
	Length (mm)		x	x	Length of panel
	Width (mm)		x	x	Width of panel
	Appearance (even/ un-even)		x	x	Visual surface evenness of panel
	Cardboard free packaging			x	Panel packaging contains no cardboard
	Optimized packaging		x	x	Panel packaging optimized for least waste
Lead-free			x	Solder used in panel is lead-free	
Tile			x	Panel in form of roof tiling	
Certifications	IEC 61215 / IEC 61646	x	x	x	IEC PV design qualification and type approval
	IEC 61730	x	x	x	IEC PV module safety qualification
	UL 1703	x	x	x	UL Standard for safety of flat-plate PV panels
	CE Marking	x	x	x	Compliance with European conformity requirements
	ISO 9001	x	x	x	ISO Quality management standard
	ISO 14001	x	x	x	ISO Environmental management standard
	NEC 2008		x	x	NEC Safe installation of electrical equipment standard
	Safety class II @ 1000V IEC 61701 UL 4703		x	x	Double insulated appliance standard IEC PV Salt mist corrosion standard UL PV cable standard
Warranty	Workmanship Warranty (years)	x	x	x	Workmanship warranty
	Power warranty (% power warranted years)	x	x	x	Power warranty, calculated for comparison by taking area of the % warranted by years warranted curve
Economics	Effective Cost/W (\$/W)	x	x	x	Post subsidy system cost per Watt of rated power output
	Time on market (years)	x	x	x	Length of time panel has been on the market



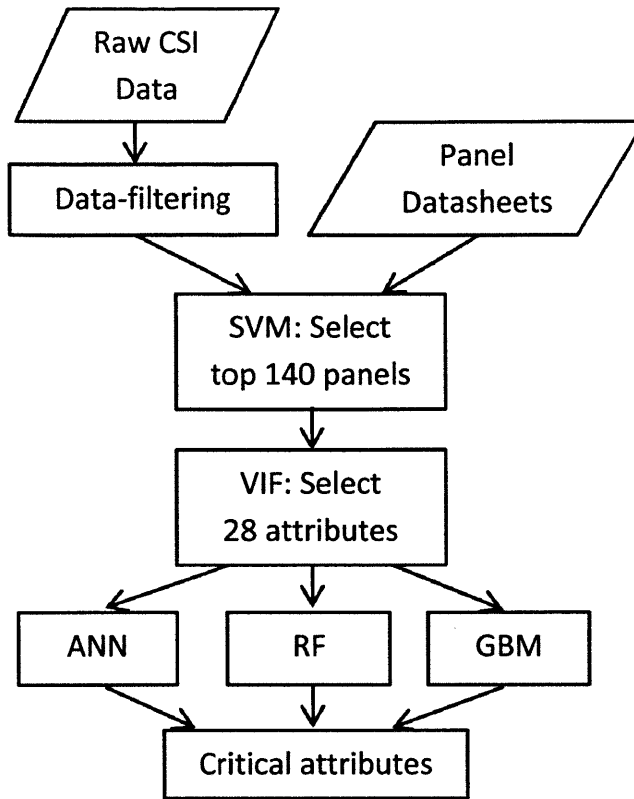


Figure 4-3: Flowchart of methodology

bootstrap samples of the data with random feature selection in tree induction. It makes predictions by aggregating the predictions of the ensemble [38].

### 3. Gradient Boosting Machine (GBM)

The gradient boosting machine is an algorithm that generalizes the decision tree prediction model by allowing optimization of an arbitrary differentiable loss function [39, 40].

The common set of important attributes found using these models is then taken to be the set of critical technical attributes. The rationale behind taking the intersection of the important attributes is that the different approaches have different assumptions, weaknesses and strengths. Random Forest and GBM are decision tree based algorithms, which are robust to outliers in data points and deal well with irrelevant predictor attributes. ANN does not perform as well on the above characteristics, but

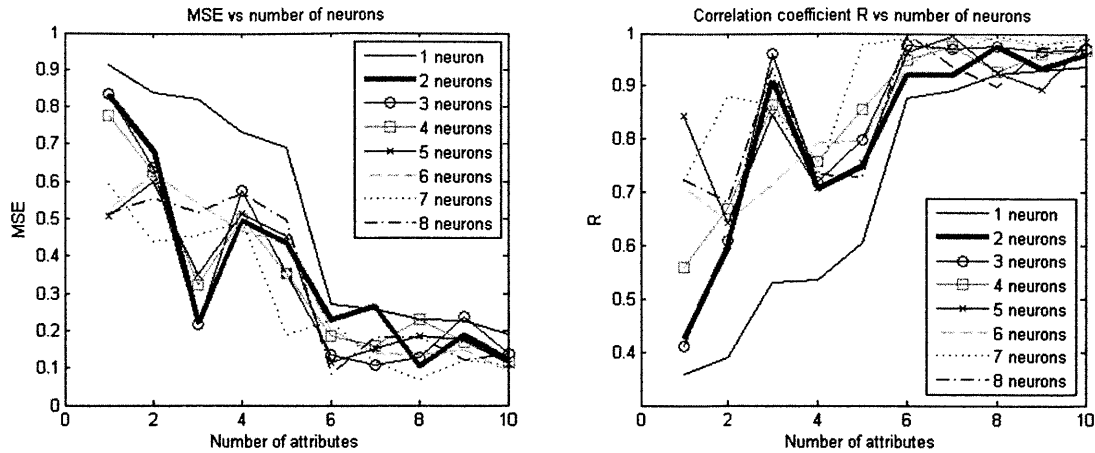


Figure 4-4: MSE and R fitting over 10 attributes using 1-8 neurons

is better at capturing non-linear and complex combinations of predictor attributes. For example, attributes A and B may not be important when taken alone, but may be significant when a combination of both is present. Random Forest and GBM may not consider A and B to be important attributes, but ANN will. Additionally, ANN and GBM may have some issues with over fitting, but Random Forests is more robust and will not over fit easily. All the algorithms picked can naturally handle both continuous and categorical predictor attributes, which is essential because the attribute list contains both binary and continuous data. They are also able to deal with incomplete data sets with some missing entries.

## 4.2.2 Artificial Neural Network Regression

A supervised feed forward Artificial Neural Network (ANN) fit was done in the MATLAB environment [41]. In order to determine the neural network architecture with an optimal number of neurons which gives the best fit without over fitting, the variance in the performance of fit with increasing neurons was tested. The number of neurons used for fitting was increased systematically from 1 to 8, using the top 10 attributes that mapped the best to market share. Each test was done with 300 trials to ensure that the global optimum was obtained, as MATLAB's neural network toolbox uses random initialization, which could affect the final result.

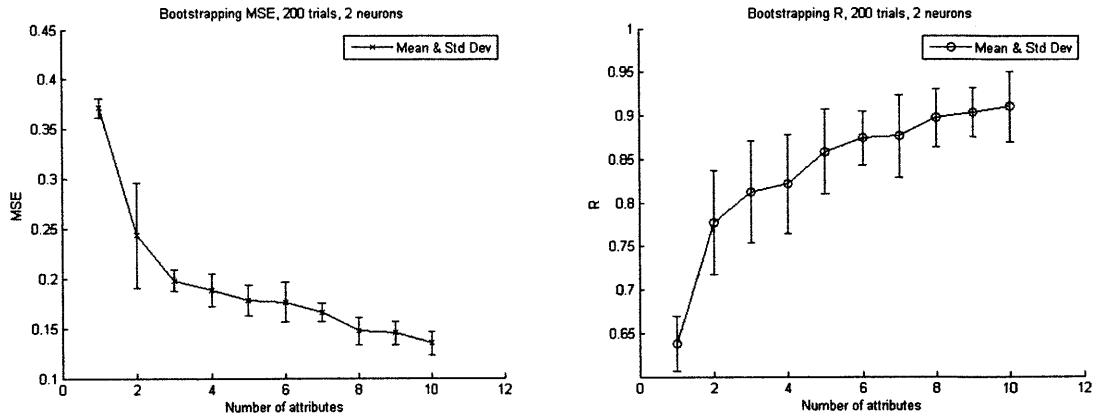


Figure 4-5: ANN bootstrapping error validation

Table 4.2: Correlation table for important attributes found by ANN

<b>R</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
Power Warranty	1.00									
Efficiency at STC	-0.43	1.00								
Time on market	-0.17	-0.19	1.00							
NEC 2008	0.05	-0.06	-0.26	1.00						
Cell Number	-0.07	0.16	0.03	-0.21	1.00					
Power Variance (+%)	-0.10	0.13	0.27	0.29	-0.22	1.00				
Safety class II	0.04	-0.01	-0.30	-0.09	0.19	-0.43	1.00			
Power ratio: PTC/STC	-0.07	0.26	-0.09	-0.07	0.17	-0.12	0.16	1.00		
Effective Cost/W	-0.14	0.28	0.23	0.10	0.26	0.36	-0.13	0.07	1.00	
Length	-0.15	0.00	0.51	0.06	-0.29	0.44	-0.24	0.00	0.16	1.00

For each neuron number, the corresponding mean squared error (MSE) and correlation coefficient R fits were obtained, and these were aggregated to form a graph of MSE and R fits using varying numbers of neurons as shown in Fig. 4-4. The optimal number of 2 neurons was selected, as it has a comparable MSE and R value to other neural networks with a higher number of neurons.

Using this optimal number of neurons for fitting, a new neural network model that maps each of the attributes to market share was built. Each optimization was run over 300 trials to ensure accurate optimization on the global minimum. MATLAB's parallel processing toolbox was used to run 3 processes simultaneously to speed up the analysis. The best model with the lowest MSE and highest corresponding R was picked to be the first element for the set of important attributes.

The second important attribute was chosen by creating new neural network models that map each attribute plus the first important attribute to market share. This was repeated until adding a new attribute did not reduce the MSE, resulting in a total of 6 important attributes. Further testing was conducted to ensure that the model is robust using correlation tables and bootstrapping methods. The corresponding bootstrapping values of MSE and R are displayed in Fig. 4-5. The correlation table of the important attributes is shown in Tab. 4.2.

### 4.2.3 Random Forest Regression

The Random Forest regression was performed using the `randomForest` statistical package created by Liaw and Wiener for the R Project environment based on the original Fortran code by Breiman and Cutler [42]. Since the Random Forest algorithm is robust to over fitting, very little tuning was required. The built in variable importance permutation calculation was used to identify critical attributes. 10,000 trees were grown and 3 variables were randomly sampled as candidates at each split. A lot of trees were necessary to get stable MSE and stable estimates of variable importance, as each input row needed to be predicted many times. The choice of 3 variables sampled at each split was decided by trying alternatives from 2-16 and choosing the best result with the lowest MSE.

Table 4.3: Top attributes' Random Forest variable importance values over 100 runs

<b>Attributes</b>	<b>Mean</b>	<b>Std Dev</b>
Power warranty	27.05	0.85
Time on market	25.21	0.73
Power variance (-%)	21.49	0.75
Fill factor at STC	20.40	1.00
IEC 61215 (crystalline) or IEC 61646 (thin film)	17.76	0.76
ISO 9001	17.51	0.88
Weight	13.79	1.15
Safety class II @ 1000V	13.67	0.81
Power variance (+%)	12.79	0.91
Efficiency at STC	10.48	0.95

100 regressions were done and an average of the importance values was taken, shown in Tab. 4.3. It was observed that although the variable importance measures varied slightly from run to run, the ranking of the importances was stable. Due to the nature of the method, cross validation was unnecessary as it generates an internal unbiased estimate of the generalization error as the forest building progresses. The importance threshold was chosen to be the absolute of the lowest importance value, resulting in a total of 10 important attributes.

#### 4.2.4 Gradient Boosting Machine

The Gradient Boosting Machine (GBM) was similarly performed in the R statistical environment using the `gbm` package written by Ridgeway based on extensions to Friedman's gradient boosting machine [43]. The learning rate, `shrinkage`, was set to 0.001, for the slowest rate but the highest accuracy. Other variables affecting the optimization, the maximum depth of variable interactions `interaction.depth`, the minimum number of observations in the trees' terminal nodes `n.minobsinnode` and the fraction of the training set observations randomly selected to propose the next tree in the expansion `bag.fraction`, were also varied systematically to obtain the optimum result with the lowest fitted MSE.

At each run, 4000 trees were grown with 10 fold cross validation. The number of trees grown was chosen to be very high to be sure that the optimal number of

Table 4.4: Top attributes' GBM relative influence values over 100 runs

Attributes	Mean	Std Dev
Fill factor at STC	23.91	0.64
Efficiency at STC	18.25	0.61
Power Variance (-)	13.43	0.56
Power ratio: PTC/STC	11.82	0.43
Effective Cost	6.72	0.19
Power warranty	5.05	0.30
IEC 61215 (crystalline) or IEC 61646 (thin film)	2.89	0.33
Weight	2.70	0.14
Time on market	2.37	0.29
Length	2.08	0.07

trees lies within the tested range. After each run, the function `gbm.perf` was used, which estimates the optimal number of trees using the data from the cross validation performed. The result at this number of trees is extracted and used.

The relative influence was then calculated by permuting one predictor variable at a time and computing the associated reduction in predictive performance. The computed relative influence was normalized to sum to 100. The mean of these relative influences over 100 regressions was then taken, shown in Tab. 4.4. The importance threshold was chosen to be a relative influence of 2, after which the relative influence values for the rest of the attributes holds steady around 1. This resulted in a total of 10 important attributes.

### 4.3 Results: Critical Attributes

The summary of important attributes found from each method is shown in Tab. 4.5. The critical attributes are taken to be the important attributes that are common to all 3 methods, and form the feature set of concern. The rank ordering of the feature set is not considered to be important, as variations in the machine learning methods will cause differences in the rank ordering of the attributes.

The critical attributes found across all 3 methods are:

1. *Power warranty*

Table 4.5: Important attributes across the 3 methods

Rank	ANN	Random Forest	GBM
1	Power warranty	Power warranty	Fill factor at STC
2	Efficiency at STC	Time on market	Efficiency at STC
3	Time on market	Power variance (-%)	Power variance (-%)
4	NEC 2008	Fill factor at STC	Power ratio: PTC/STC
5	Cell number	IEC 61215 / IEC 61646	Effective Cost/W
6	Power variance (+%)	ISO 9001	Power warranty
7	Safety class II @ 1000V	Weight	IEC 61215 / IEC 61646
8	Power ratio: PTC/STC	Safety class II @ 1000V	Weight
9	Effective Cost/W	Power variance (+%)	Time on market
10	Length	Efficiency at STC	Length

Measure of power output performance guaranteed by the manufacturer over a period of time

2. *Efficiency at Standard Testing Conditions (STC)*

Measure of performance of a panel

3. *Time on Market (years)*

Length of time panel has been on the market, measure of market maturity of product

At first glance, the critical attributes found are reasonable. Power warranty is linked to consumer confidence, as well as the reliability of the solar panel. Efficiency is a reflection of the performance of the technology, in this case the panel's ability to convert sunlight into electricity. Time on market is a measure of market maturity and durability of the panel.

It is important to note that the relationships between the critical attributes and market share derived from the machine learning algorithms do not imply causation. For example, the power warranty might not be the direct reason why customers prefer a certain panel over another, it might instead be a reflection of increased consumer confidence in the manufacturer's quality that results in increased market share. On the other hand, if there is no relationship, the attribute is not an important factor in the purchase decision.

Furthermore, the presence of all 3 critical attributes found does not guarantee market success for the product. The panel might have a good power warranty, high efficiency, and a long market lifetime, and still perform poorly on the market. Other non-technical factors like service quality, country-of-origin, and manufacturer reputation may play important roles in the purchase decision that are not reflected in this part of the study. They will be taken into account in the next chapter, and in future work. What the analysis does show is that the panels need to have competitive levels of these critical attributes in order to have a chance at succeeding in the market. Hence, the list of critical attributes can be seen as "must-have" attributes that designers should not neglect in the product development phase.

It is of value to note the factors that do not show up as important attributes in any of the methods. Interestingly, reduced waste in packaging, lead-free solder and the ISO 14001 environmental management standard fail to appear as important. The possibility that a consumer might miss these factors is low, because manufacturers heavily promote them as differentiating features, and they are displayed in large font at prominent places on the panel datasheets and advertising material. Since these are the only 3 factors in our analysis that reflect additional design thought on the eco-friendliness of the product, it can be inferred that consumers do not consider the environmental impact of non-core aspects of solar panels to be important when making their purchase decision. This is the opposite result of what is expected from using a stated preference method. This is a common problem in stated preference methods, with consumers responding differently in hypothetical situations than in actual market conditions. Homeowners who purchase PV systems frequently think



Table 4.6:  $R^2$  values for the 3 methods

	ANN	RF	GBM
$R^2$	0.851	0.889	0.944

of themselves as more environmentally conscious than the average population. However, previous research findings support our finding, showing that inconsistencies exist within "green" consumption areas, where environmentally conscious consumers will not necessarily buy more "green" energy products [44].

Unexpectedly, effective cost per Watt only appears as 5th in the GBM list of important attributes and 9th in the ANN list, although cost is frequently considered by many to highly influence the purchase decision. This result is a limitation of our study, as due to constraints in collecting data, we used the total cost of the PV system, which includes not only the panels, but also the inverter, labor, and installation costs, minus the state subsidy that was applied. This effective cost might not have been a factor of consideration when choosing between different panels. For a more accurate reflection of how cost influences this decision process, the panel price per Watt should have been used, but this data was unavailable in the California Solar Statistics database, and thus was not considered in this study.

## 4.4 Comparison of Methods

Some agreement between the various machine learning algorithms can be seen in Table 4.5. Only 3 attributes are common, 9 attributes occur twice, and 3 attributes only occur once. The different predictions are due to the noise in the data, which is an inherent problem when dealing with real data. The internal structure of the methods also differ, meaning the methods perform regression in differing ways. Although Random Forest and GBM are both decision tree based methods, because the learning approach differs, the important attributes found could be inconsistent. ANN has a completely distinct internal structure from the decision tree based methods, causing

the important attributes found to be different. The combination of noisy real data and differing internal structures of the methods results in limited agreement.

A comparison of the accuracy of the models in predicting market share using the important attributes is shown in the  $R^2$  goodness-of-fit values reflected in Tab. 4.6, where  $R^2$  is calculated by Eq. 4.2.

$$R^2 = 1 - \frac{MSE}{\sigma_{MS}^2} \quad (4.2)$$

where MSE is the mean squared error, and  $\sigma_{MS}$  is the standard deviation of market share.

Table 4.6 indicates that all the models perform relatively well, with GBM being the most accurate. Ideally, Random Forest and GBM should have similar performance, because they are both decision tree based algorithms. The difference lies in how they optimize decision trees using ensemble approaches. Random Forest is usually more robust to internal parameter choice and performs well for wider varieties of parameters. Meanwhile, GBM tends to do better than Random Forest when the internal parameters are optimized carefully, as in this case. This highlights the need to carefully tune and test the parameters of machine learning methods before using the results.

With regard to computation time, GBM and Random Forest took a similar amount of time to run. ANN took a much longer time to train properly, although this might have been partly due to the difference in platform, with MATLAB running slower than R.

## 4.5 Conclusions

In this chapter, we proposed a machine learning approach for revealing consumer preferences for residential PV panels. We found 3 critical attributes that designers can prioritize for the optimization of time and resource allocation for the product development cycle: PV panel power warranty, efficiency, and length of time on the market.

It appears that consumer preferences can be extracted successfully from marketing data and engineering specifications using the methods we attempted. However, this is limited by the presence of data that is used in the analysis, as shown by the absence of cost in the list of critical attributes because panel cost data could not be obtained.



# Chapter 5

## Stated Preference Analysis from PV Installer Survey

In the previous chapter, we showed that consumer preferences can be extracted from a combination of market data and engineering specifications using machine learning methods. However, as pointed out by prior research in the field, the revealed preference method has the limitation that only the set of attributes that are present in the data can be tested. There is a possibility that there may be other critical attributes that are not present within this dataset which are an important part of the homeowner purchasing decision process.

In this chapter, we perform a stated preference analysis using a self-explicated survey method to serve as a comparison with of our previous revealed preference work, as well as to gather data that will be used in future work to boost the revealed preference method with additional attributes that are not available from data sheets.

### 5.1 Survey Objectives

The research objectives of the survey are to determine:

1. The importance of technical attributes of PV panels.
2. The relative importance of technical (e.g. efficiency) and non-technical at-

tributes (e.g. reputation) that affect the purchase decision between PV panels.

3. The aesthetic preferences between different PV panel styles.
4. The reputation and service quality provided by several top solar manufacturers.

Surveys were sent out to PV panel installers instead of homeowners as it was understood from conducting interviews with PV panel installers (Chapter 3) that homeowners typically do not have adequate knowledge in the field to be able to make their own decisions on panel choice. Instead, installers will recommend a PV panel type to the homeowner, according to their aesthetic and price needs, and homeowners will simply approve or reject the proposal. Since the decisions are made predominantly by PV installers, they were the target of this survey.

The survey was only sent out to certified PV system installers within the state of California, to match the geographical area of the market data used in the revealed preference analysis (Chapter 4).

## 5.2 Methodology

Survey questions were drafted to meet the research objectives of the survey, and can be found for reference in Appendix A.2. A range of open ended, ranking and rating questions were used. Special attention was paid to organization, randomization of response alternatives and word choice of questions to reduce the effects of priming and social desirability bias. Priming occurs when respondents are exposed to a stimulus which causes them to give increased weight in subsequent judgment tasks [45], while social desirability bias causes respondents to answer questions based on an accepted social norm, rather than truly expressing their own opinions [46]. Both result in poor accuracy of survey results, thus the survey was carefully crafted to minimize these effects.

Internet-based surveys were chosen over telephone and paper surveys because it was faster and cheaper to distribute the surveys online to a relatively large population. The ease of completion and return of an internet-based survey also offers an added

benefit over paper surveys that have to be mailed back. The web based survey development tool SurveyMonkey was picked as the survey creation and distribution medium, as it offered an easy to understand user interface for the survey taker.

Personalized survey invites were sent to the intended respondents to encourage higher participation rates [47]. No personal identifying data was collected, and each respondent was ensured the confidentiality of their responses. The total length of the survey was capped so that the approximate time it would take an average respondent to complete the survey would be about 15 minutes. This was done in order to reduce survey fatigue, and to encourage more survey takers to start the survey.

A first pilot survey was sent to 10 researchers and graduate students who had experience writing survey questions, to serve as a check for the wording and order of the questions so that data collection errors were minimized. The same pilot survey was sent to 2 Massachusetts-based PV installers, to ensure that the questions were worded using terms common to the solar industry. Changes were made, and a second pilot survey was sent to a randomly selected group of 51 California-based PV installers by email. The second pilot group was given a chance to be entered into a sweepstakes for a \$25 Amazon.com gift card upon completion of the survey. The respondents were given a time period of 2 weeks from initial request to survey closing, with 1 reminder email sent a week after the initial email invitation. A total of 3 responses were received, for a response rate of 5.9%.

After slight modifications to the wording of certain questions, instructions, and the survey email invite, the survey invite was sent out by email to 2094 listed PV system installers within the state of California. Respondents were entered into a sweepstakes for a \$100 Amazon.com gift card upon completion of the survey. The survey was carried out over a period of 3 weeks, from January 27 to February 18, 2012, with 2 reminder emails spaced apart by a week each. A total of 138 responses were received, for a response rate of 6.6%. In total, the time taken from start of survey creation to analysis was about 3 months.

## 5.3 Survey Respondent Demographics

The PV system installers that responded to the survey were predominantly male, with only 10.5% female installers compared to 89.5% males, as shown in a pie chart in Fig. 5-1. We suspect this reflects the gender imbalance within the profession itself, and less due to factors associated with the survey. All respondents were 21 years of age or older, with the largest group (33.8%) falling within the age range of 50-59.

Of this group, most respondents (73.9%) indicated that they have 1-10 years of working experience within the solar industry (Fig. 5-3), and 65.6% have been in their current position for about as long (Fig. 5-4). 9%, or 12 survey respondents reported more than 20 years of experience in the field.

All respondents had multiple areas of involvement with the client, with a good spread among the interested areas (Fig. 5-5). This ranged from the first contact with the residential client, including marketing and initial site evaluation, all the way through the installation process to monitoring and maintenance. The "Other" responses included management, project financing and customer support.

Although the survey was targeted at installers who do residential installations, 15.7% of the respondents were in companies that dealt primarily with clients in the commercial, non-profit and utility-scale sectors (Fig. 5-6). The other 84.3% dealt predominantly with residential clients. Lastly, 37.0% of installer companies were a distributor/dealer for a specific solar panel manufacturer, while 63% operated independently of manufacturers (Fig. 5-7).

From the data, the average profile of the survey respondents is a male, in his forties or fifties, with about 5 years of experience in the solar industry. He has experience dealing with the residential client in multiple areas, and deals mostly with residential PV installation projects.



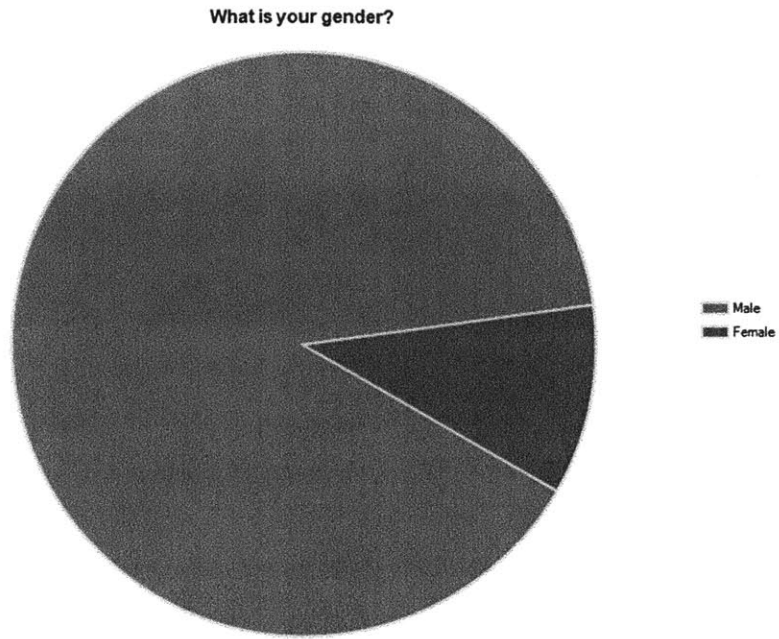


Figure 5-1: Gender split of survey respondents

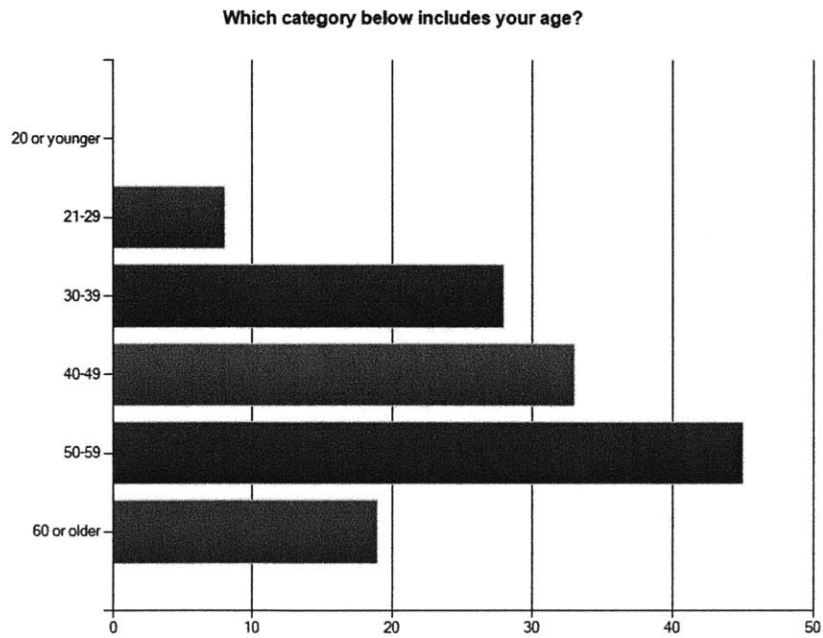


Figure 5-2: Age of survey respondents

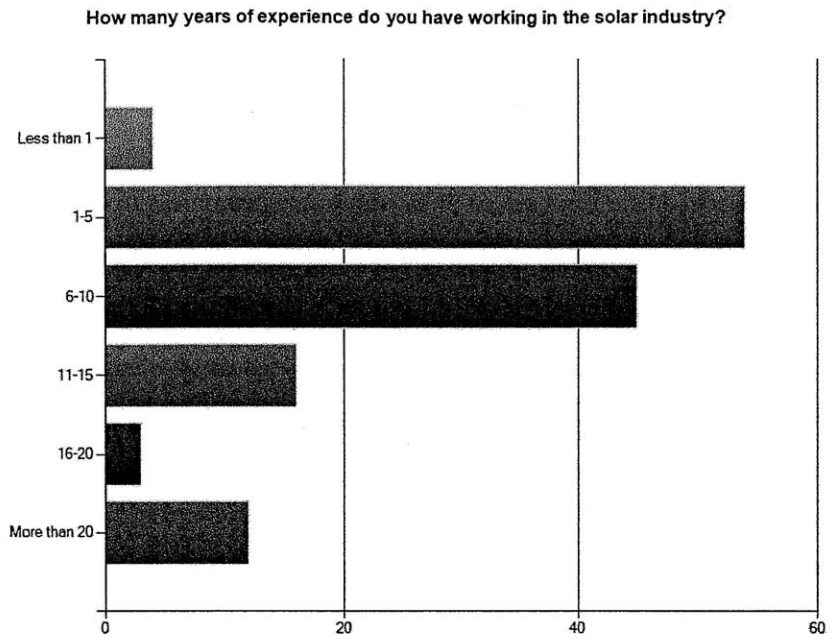


Figure 5-3: Years of experience in the solar industry

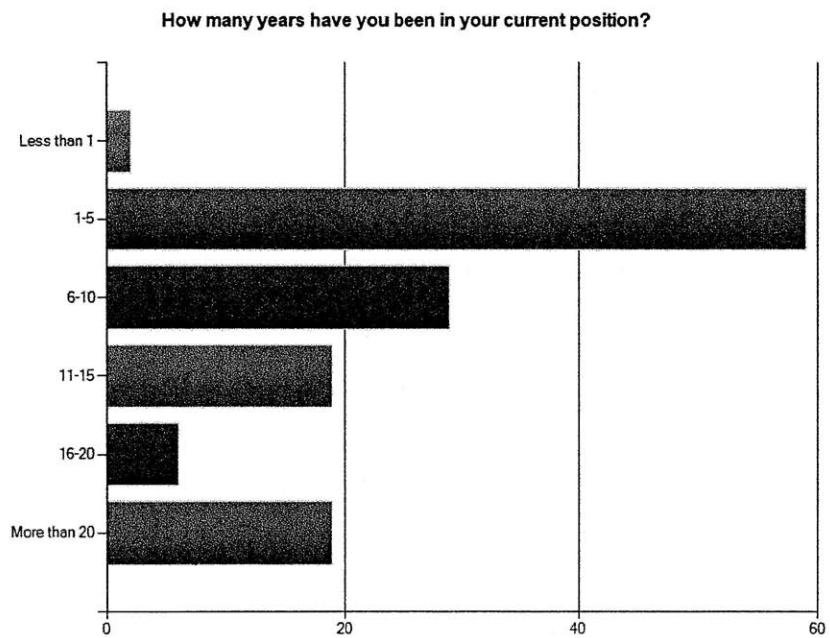


Figure 5-4: Years in current job position

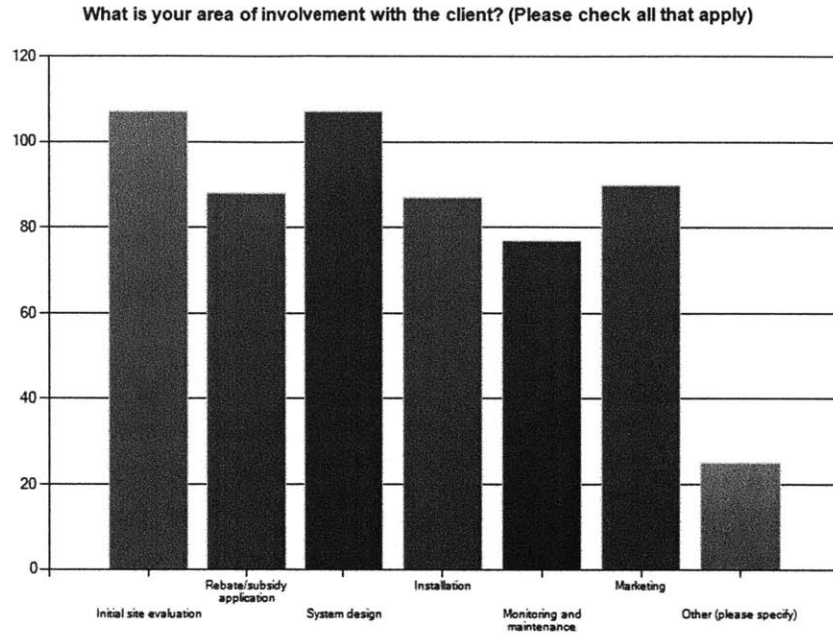


Figure 5-5: Area of involvement with client

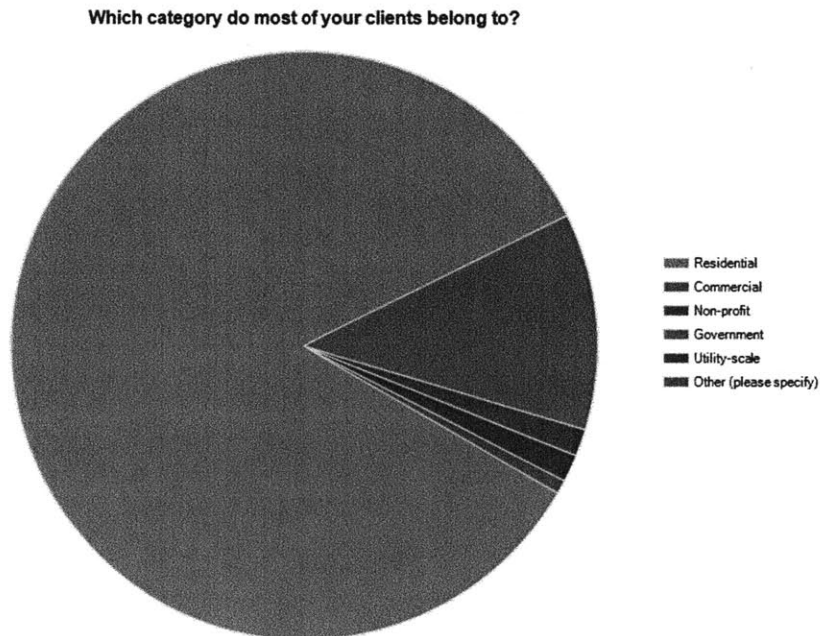


Figure 5-6: Client PV sector

Is your company a distributor/dealer for a specific manufacturer(s)?

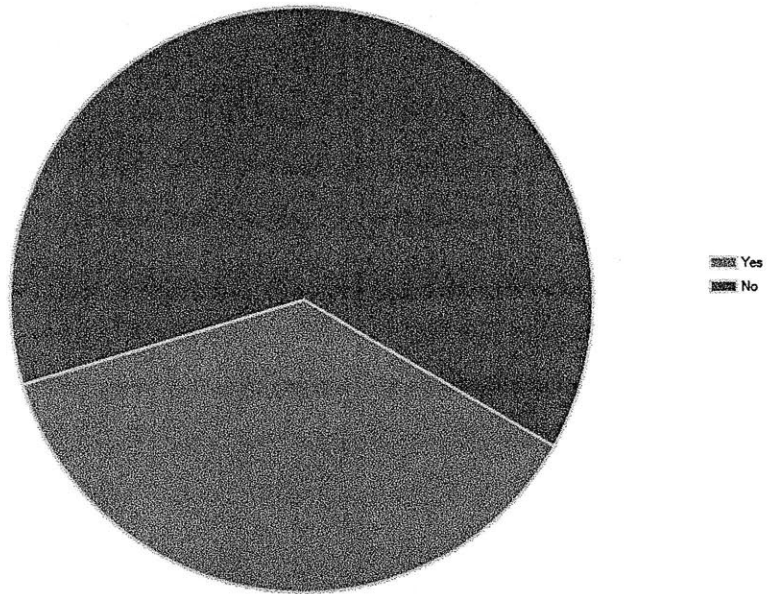


Figure 5-7: Manufacturer affiliation

## 5.4 Findings

### 5.4.1 Importance Rankings of Technical Attributes

Figure 5-8 shows the importance ranking of the technical attributes from the survey. Respondents were asked to pick 6 of the most important technical factors of a PV panel model and rank them in order of importance. The list corresponded to factors in the revealed preference analysis (Chapter 4). The importance of the attributes can be obtained by calculating the weighted sum, as shown in Eqn. 5.1.

$$A = \sum(w_i * c_i), \text{ for } i = 1, 2, \dots, 6 \quad (5.1)$$

where A is the weighted sum, w is the ranking weight, c is the count, and i is the rank.

As expected, price per Watt was the attribute with the highest ranking, being mentioned by 99 out of 136 survey respondents as being one of the 6 most important technical factors of a PV panel model. This is followed by PV panel efficiency, power warranty and product maturity. Weight was the attribute with the lowest relative importance.

This finding is consistent with the revealed preference analysis conducted in Chapter 4, where the 3 critical attributes found to most influence market share were PV panel efficiency, power warranty and time on the market. Panel price was not included in the analysis as that data was not available.

The consistency in results serves as a validation for our previous analysis. However, it points to the possibility of obtaining misleading results when important data is not included in the analysis. As panel price data was not available, it did not come up as a critical attribute, even though it is the most important attribute that installers consider when making a PV panel recommendation to a homeowner. This is a big problem that afflicts revealed preference analysis- intuitively the researcher or designer might know that an attribute is important, but if the data relating to that attribute is unavailable, the revealed preference analysis becomes unreliable. This is

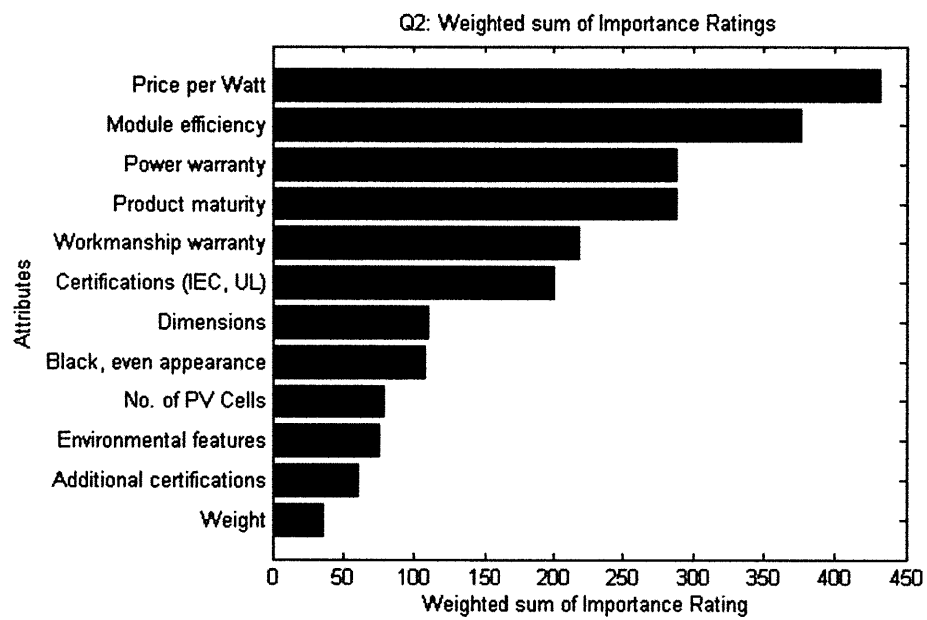


Figure 5-8: Importance ranking of technical attributes

the reason why most marketing research relies heavily on stated preference analysis. As data collection becomes increasingly automated and comprehensive, this problem should diminish over time. However, researchers should still exercise caution while relying solely on revealed preferences.

#### **5.4.2 Relative Importance of Technical and Non-Technical Attributes**

In the survey, two types of questions regarding the relative importance of technical and non-technical attributes were asked survey respondents: an open ended, and a rating question.

**Open ended** List the 5 most important things you consider when choosing between PV panel models to recommend to a homeowner.

**Rating** How important are the following overall factors when recommending a PV panel to a homeowner?

The open ended question was asked at the very start of the survey to minimize priming effects caused by the options provided in the rating question. Using the formula for the weighted sum in Eqn. 5.1, the importance rankings calculated can be seen in Fig. 5-9, along with the number of times each attribute was mentioned by respondents. Cost far outweighed the rest of the factors by a large margin, with 77.5% of respondents mentioning "cost" or "price" as 1 of the 5 most important factors of consideration. Aesthetics was mentioned by 40.6% of respondents, followed closely by warranty, efficiency, country of origin and quality. These same 6 factors were also given the highest importance ranks.

From this list of attributes, it is interesting to note that certain factors that were overlooked. None of the installers considered any environmentally friendly features, like optimized packaging, lead-free solder, or a recycling program to be one of the top 5 important features. This is in line with the results found in the revealed preference analysis, where it was found that the presence of additional eco-friendly features

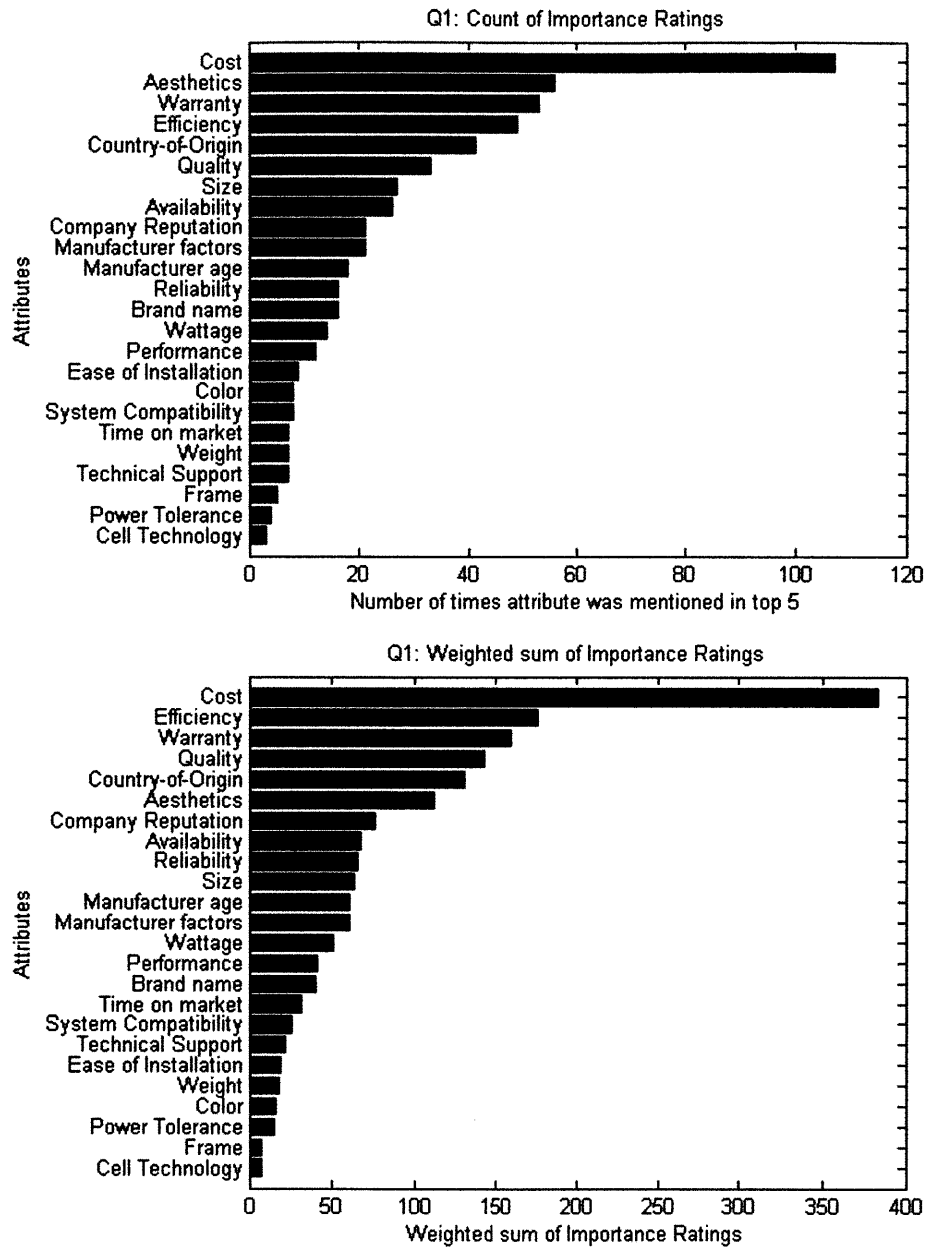


Figure 5-9: Relative importance of technical and non-technical attributes: Open ended question



did not affect the panel type's market share, even though they were thought to be important by the manufacturers as a distinguishing factor to be marketed.

Similarly, only 2 installers mentioned PV panel certifications, like IEC 61215/ IEC 61646 and UL 1703, which are standards for PV design qualification, type approval and safety. This might be because all panels that can be sold within California must by law have these certifications, and they are no longer a distinguishing factor in the opinion of the Californian PV installers.

Comparing the results of this question to the importance rankings of technical attributes shown in Fig. 5-8, there seems to be some corroboration. There was no explicit mention of optional certifications, like IEC 61701, the standard for salt mist corrosion, or ISO 9001, the standard for quality management. The number of cells on a PV panel was also not mentioned by installers. These match up with the technical attributes with low importance rankings.

However, the weight of the panel was an attribute that 5.0% of installers brought up as one of the 5 most important attributes, but was ranked as the least important technical attribute. Hence, it appears to be unimportant to the majority of installers, but something that a select few care about a lot. This is an example of preference heterogeneity, which is the extent to which individual preferences vary across consumers [48]. The reason for this discrepancy might be because the installers have to manually carry the panels up on the roof during the installation, which can be a difficult procedure for some if the panels are heavy. For these installers, the weight of the panel might matter a lot, but for other more able bodied installers, or installers who do not personally carry out the installation, the weight of the panel might not be a big factor of consideration.

In the rating question, when asked to rate the importance of technical and non-technical attributes when recommending PV panel to a homeowner on a 5 part scale from very unimportant to very important, installers rated the attributes as shown in Fig. 5-10. Reliability was ranked the highest, above price per Watt and the panel manufacturer's reputation. Once again, weight of the panel was rated as the least important, scoring an average of 3.06, only slightly higher than a "neutral" 3.

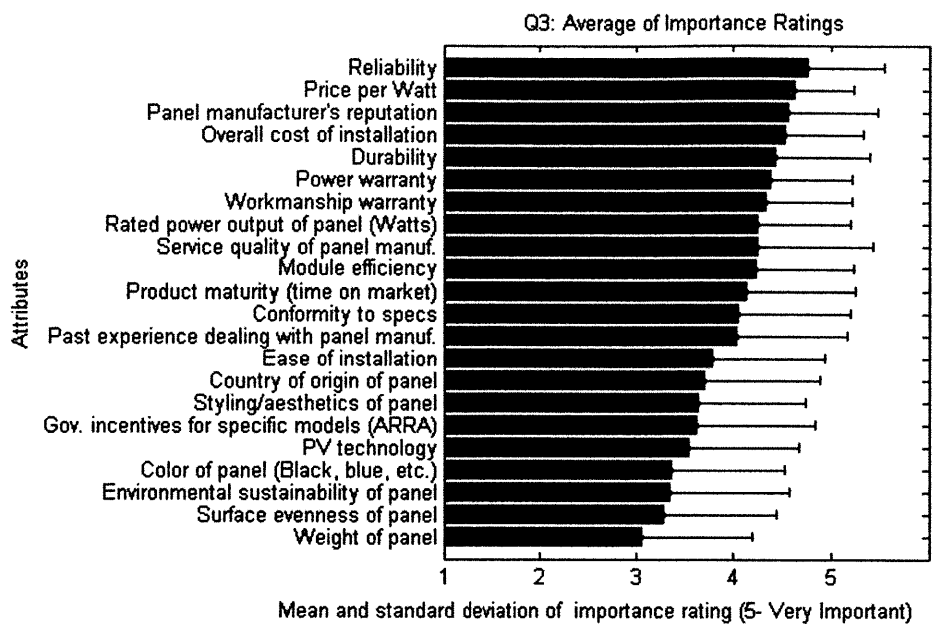
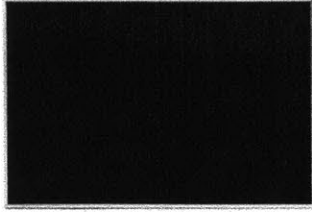
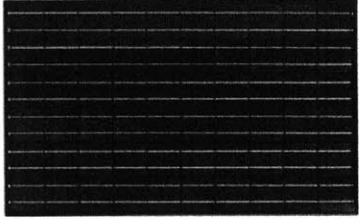
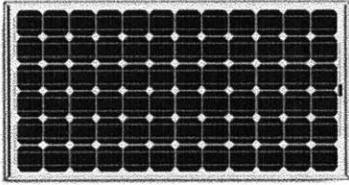
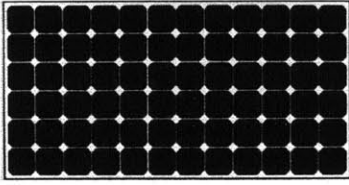


Figure 5-10: Relative importance of technical and non-technical attributes: Rating question

Comparing the results between the two question types, several differences surface. Reliability of the panel obtained the highest importance ratings, but was only mentioned by 11.6% of the respondents as one of the top 5 factors. Panel module efficiency was 10th in importance on the ratings question but mentioned by 35.5% of respondents, coming in second in importance in the open ended question. Aesthetics was 16th on the ratings question, but 6th on the open ended one. The discrepancy between the ranking of attribute importance from the two questions points to internal inconsistencies associated with stated preference methods. This preference inconsistency is present when different phrasings of a question are shown to achieve different results, making it difficult to determine the actual preference of the respondent.

A consistent finding from the two questions is the relative importance of certain non-technical attributes compared to the technical attributes. The panel manufacturer's reputation, country of origin, and aesthetics seem to be important attributes that play into the installers' decision to recommend a panel type to a homeowner.

Table 5.1: PV panel appearance groups

	Light frame color	Dark frame color
Even surface appearance		
Uneven surface appearance		

### 5.4.3 Aesthetic Preferences

The panels were grouped into 4 different categories based on their frame color and surface appearance, as shown in Table 5.1. This categorization was based on data from the top 140 panels studied in the previous chapter (Chapter 4). For each category, 3 example panels were selected. The survey was set up so that each respondent was randomly given a set of panels, one from each category. Hence, each example panel was seen by approximately 33.3% of the survey respondents. The panels were presented without their categories stated, and were known to respondents simply as "Panel A", "Panel B", "Panel C" and "Panel D".

Respondents were first asked to rank the panels in order of attractiveness, then

Table 5.2: Aesthetic ranking values from survey

**Attractiveness**

	Dark frame, uneven	Dark frame, even	Light frame, even	Light frame, uneven
Mean Ranking	1.736	2.760	2.128	3.376
Standard Deviation	0.973	0.880	0.971	0.882

**Visual Compatibility with Roof**

	Dark frame, uneven	Dark frame, even	Light frame, even	Light frame, uneven
Mean Ranking	1.640	2.832	2.048	3.480
Standard Deviation	0.880	0.883	0.920	0.765

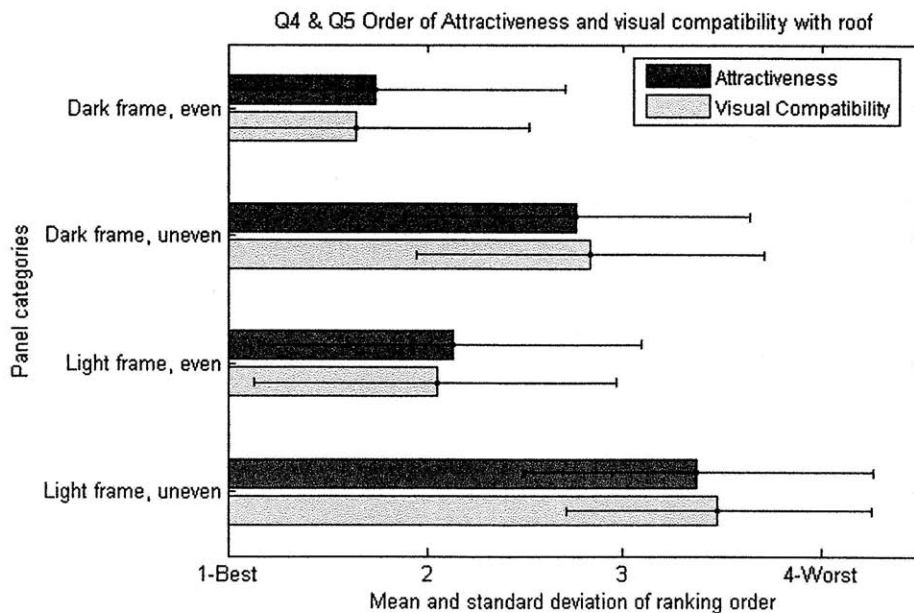


Figure 5-11: Aesthetic preferences for PV panel categories

in order of visual compatibility on a roof. The results can be seen in Tab. 5.2 and Fig. 5-11.

It appears that installers prefer panels with an even surface appearance over panels with an uneven surface appearance. There was also a slight preference for dark colored frames over light colored frames. A good correlation between the level of attractiveness and visual compatibility with roofs means that installers associate a panel that blends in well with the roof as being more attractive.

This result suggests that using a black back sheet over a white one is preferable, and will increase the attractiveness of the panel substantially. It is in line with panel manufacturers' move toward sleeker, all black panels with a smooth appearance and black frame for residential installations, despite the slight decrease in panel efficiency.

The similarity of preferences across all 3 sample panels from each category suggests that the aesthetic non-technical attribute is comprised of a combination of two technical attributes: frame color and surface appearance. This brings to attention that even though people consider aesthetics to be a subjective, non-quantifiable measure of attractiveness, in this case it can be broken down into technical attributes that can be quantifiable in the development process. Designers and developers can then tweak these attributes in order to achieve an improved aesthetic appeal.

#### **5.4.4 Reputation and Service Quality**

In the next section, survey respondents were asked to rate 22 PV panel manufacturers corresponding to the top 140 panels found in the previous chapter by reputation and service quality. They were given 4 options: Excellent, good, poor and no opinion. The results are displayed in Figures 5-12, 5-13, 5-14, and Tables 5.3 and 5.4 .

An interesting observation is that higher name recognition does not always correspond to a better reputation. Two of the top ten most recognized manufacturer names were rated by installers to have a below average reputation. However, reputation and service quality seem to be closely related, with manufacturers with a better service quality rating also having a better rated reputation.

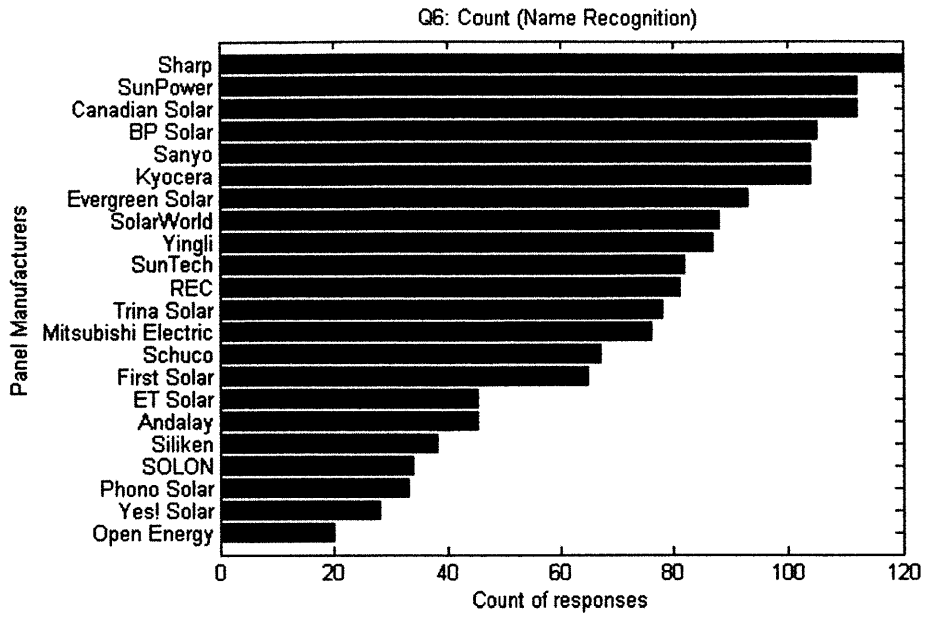


Figure 5-12: Name recognition of panel manufacturers

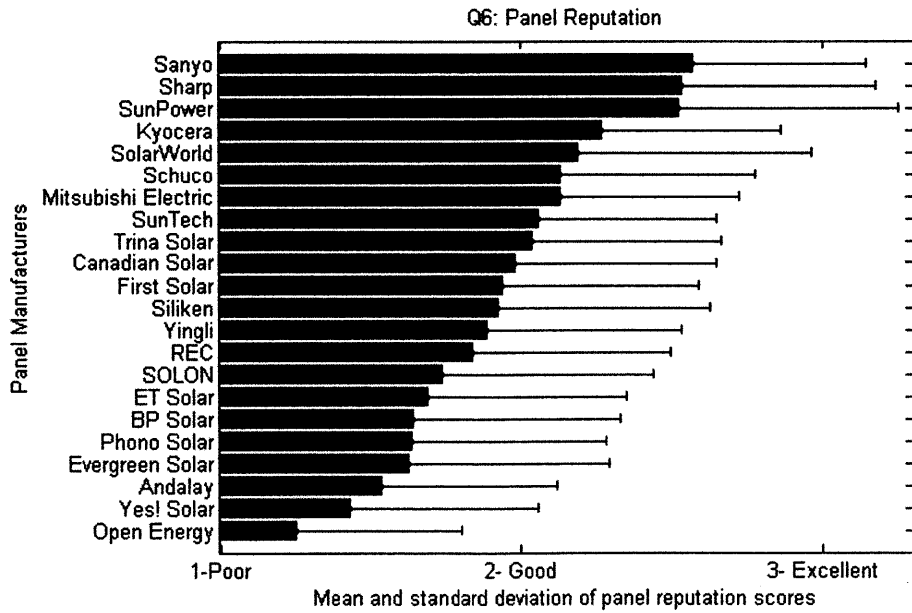


Figure 5-13: Reputation of panel manufacturers

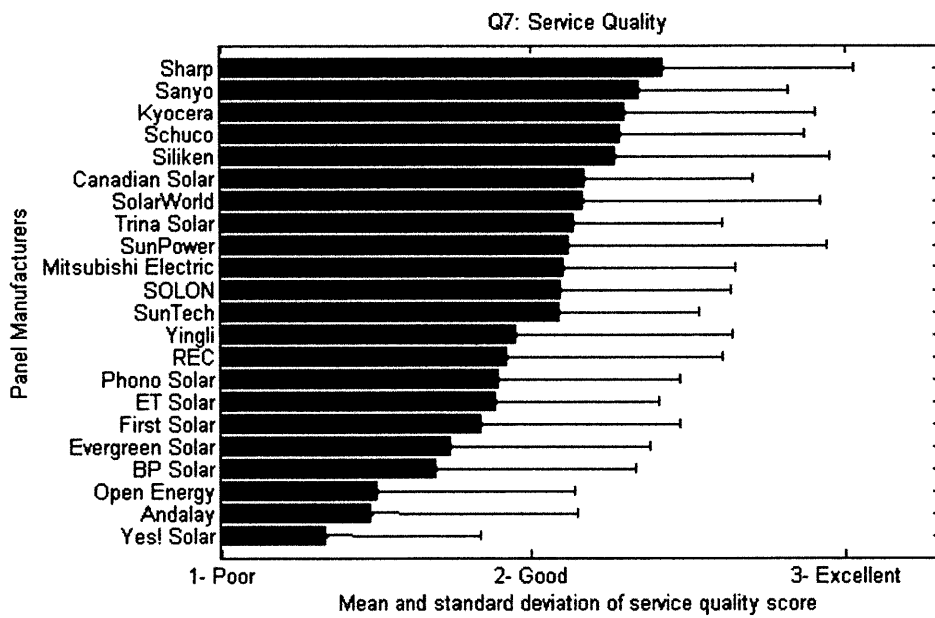


Figure 5-14: Service level of panel manufacturers

Table 5.3: Manufacturer reputation survey ratings, rating of 3 is excellent, 1 is poor

<b>Manufacturer</b>	<b>Count</b>	<b>Mean</b>	<b>Std Dev</b>
Andalay	45	1.533	0.588
BP Solar	105	1.638	0.695
Canadian Solar	112	1.982	0.671
ET Solar	45	1.689	0.668
Evergreen Solar	93	1.624	0.674
First Solar	65	1.938	0.659
Kyocera	104	2.269	0.595
Mitsubishi Electric	76	2.132	0.596
Open Energy	20	1.250	0.550
Phono Solar	33	1.636	0.653
REC	81	1.840	0.661
Sanyo	104	2.577	0.569
Schuco	67	2.134	0.649
Sharp	120	2.542	0.634
Siliken	38	1.921	0.712
SolarWorld	88	2.193	0.771
SOLON	34	1.735	0.710
SunPower	112	2.527	0.722
SunTech	82	2.061	0.595
Trina Solar	78	2.038	0.633
Yes! Solar	28	1.429	0.634
Yingli	87	1.885	0.655



Table 5.4: Manufacturer service level survey ratings, rating of 3 is excellent, 1 is poor

<b>Manufacturer</b>	<b>Count</b>	<b>Mean</b>	<b>Std Dev</b>
Andalay	23	1.478	0.665
BP Solar	65	1.692	0.635
Canadian Solar	67	2.164	0.539
ET Solar	25	1.880	0.526
Evergreen Solar	50	1.740	0.633
First Solar	24	1.833	0.637
Kyocera	62	2.290	0.611
Mitsubishi Electric	40	2.100	0.545
Open Energy	16	1.500	0.632
Phono Solar	18	1.889	0.583
REC	36	1.917	0.692
Sanyo	68	2.338	0.477
Schuco	29	2.276	0.591
Sharp	82	2.415	0.608
Siliken	23	2.261	0.689
SolarWorld	56	2.161	0.757
SOLOON	11	2.091	0.539
SunPower	78	2.115	0.821
SunTech	49	2.082	0.449
Trina Solar	38	2.132	0.475
Yes! Solar	9	1.333	0.500
Yingli	39	1.949	0.686

## 5.5 Conclusions

The results from the survey conducted amongst PV panel installers in California validate the critical attributes found in the previous chapter (Chapter 4), but show that non-technical attributes are also important in the purchasing decision. This presents a problem to the revealed preference analysis, as non-technical attributes cannot be easily and accurately quantified to be measured for the purposes of comparison.

The survey also suggests that some of the non-technical attributes may be influenced heavily by certain technical attributes that may be measurable. The findings suggest that aesthetics may be decomposed into two technical attributes- the frame color and the evenness of the surface. This opens up an area of future work for the decomposition of other non-technical factors.

# Chapter 6

## Discussion

In this thesis, we investigated ways to obtain consumer preferences for technology products to help designers identify the key attributes that contribute to a product's market success. A case study on residential solar PV panels was done in the context of the California market within the 2007-2011 time span. The research was done in 3 stages. First, we conducted interviews with solar installers to gain a better understanding of the solar industry, then we implemented two methods to obtain consumer preferences: a revealed preference method based on collected data, and a stated preference method using a survey. The results from these two methods were compared, and our findings reported.

The research questions we set out in Chapter 1 can be answered as follows:

1. *Can revealed consumer preferences be obtained from market data and engineering specifications using machine learning methods?*

In Chapter 4, we showed that sets of consumer preferences can be revealed by machine learning analysis of market data and engineering specifications. We obtained a group of three critical attributes that most influenced market share: power warranty, PV panel efficiency, and the time the panel has spent on the market.

This result was compared with data from a stated preference self-explicated sur-

vey method in Chapter 5, and the findings were validated. However, due to the unavailability of pricing data for the PV panels over the time period, price of the panel per Watt was not included in the revealed preference analysis, and did not show up as a critical attribute. This is a problem of the revealed preference analysis, as the result is only as good as the data that is available. Without a comprehensive collection of data, the results may be misleading. We foresee this this problem will be diminished over time, as increasing use of digital records mean that data availability is on the rise.

2. *Is there agreement among the machine learning methods that suggest the validity of the data and methods?*

Yes, there seems to be some agreement among the machine learning methods used. Out of the 26 attributes tested in Chapter 4, 12 attributes were ranked in the top 10 most important attributes by 2 out of 3 machine learning methods, with 3 attributes occurring thrice and 9 attributes occurring twice. This shows a good consistency between the results from the machine learning methods.

However, the rank order of the important attributes are not the same across the 3 methods. We think this is due to the way the different methods handle the data optimization when there is a significant amount of the noise in the data. This noise is inevitable when dealing with real data, and poses a challenge to the researcher. This makes it difficult to determine which rank order is necessarily correct. There is a need to look deeper into the machine learning method we have explored in order to determine the differences in which the methods handle the data.

3. *How does the revealed preference approach compare with a stated preference self-explicated survey method?*

The revealed preference method has several advantages over stated preference methods. First, revealed preference methods are not affected by preference construction or reversal, caused when people construct their preference only when asked, instead of retrieving their preferences from a pool of preexisting preferences. This means that people might report one thing, but may not necessarily act in the same way when faced with a real situation. This is not a problem for revealed preferences as consumers have already made their decision.

If data is available, the revealed preference method is quick to perform, and can be done in about a week with a regular computer, faster if cluster methods are used. The method can use both marketing and engineering data, and does not require engineering specifications to be translated into a form that the consumer can understand. This means that more engineering specifications can be tested, and directly applied to engineering design in a more accurate and efficient manner.

However, there are several areas that the designer or researcher needs to take caution in. The result of revealed preference analysis is only as good as the data that is used in creating the model. If the input data is missing some important attributes, in our case the price data, the results will be misleading. This requires designers to have an intuition for the important attributes, or for the input dataset to be comprehensive. An increase in data however, will increase the time taken for the analysis to complete, which may be a trade off to the designer.

We also found in the survey in Chapter 5 that non-technical attributes may be an important part of the decision making process when choosing between PV panels. These non-technical attributes include panel manufacturer's reputation, name recognition, and service, among others. It is difficult for these attributes to be used in a revealed preference analysis as they are subjective measures which are difficult to measure accurately. This subjectivity means that there will be increased noise introduced into the dataset, which might lead to a poor

result.

Hence, it seems logical to combine the revealed and stated preference methods in order to overcome the difficulties faced by each of the methods. The stated preference method can be used to boost the revealed preference method so that non-technical attributes can be included in the analysis. This will be attempted in future work.

## 6.1 Limitations of Study

Only data from California's PV market within 2006-2011 time period has been explored. These results might not hold when a larger geographical market is studied, as the preferences of people in different locales and climates might be different.

The lack of available PV panel data has also limited our revealed preference analysis. We were only able to obtain the technical data from publicly available specification sheets, and did not have access to other technical data like temperature response, or performance in diffused light, so the analysis did not reflect these considerations.

As noted by literature, accurate consumer preferences are very difficult to measure, making it difficult to validate the accuracy of our findings. Our method is limited to cross-method validation that compares our revealed preference approach to a self-explicated survey method. Additional validation using other established methods such as conjoint analysis may yield slightly different results.

Another challenge to our method of finding the set of critical attributes is the loss of importance rankings during the last part of the analysis, where the critical attributes are taken to be the set of attributes common to the 3 machine learning methods. Where in the survey results an order of importance can be established, it is suppressed in our revealed preference method. This is a potential problem for designers if the critical attributes found are not independent, and where optimizing one attribute would mean a trade off with another critical attribute. More work could be done with the machine learning methods in order to establish a rank order of the attribute importances.

## 6.2 Future Directions

From the Chapter 5, we found that the non-technical attribute "aesthetics" may be broken down into two technical attributes. It would be interesting to determine if other non-technical attributes exhibit the same correlation with technical attributes. Future work could explore the correlation between the technical and non-technical attributes, to see if a decomposition of the non-technical attributes may be found. It is possible that designers may be able to change technical attributes to achieve an increase in a non-technical attribute, which may then affect the performance of the product on the market.

Since both revealed and stated preference methods have their strengths and weaknesses, in future work a combined model using both should be explored, using stated preferences to boost the revealed preference framework [49, 50, 51]. This would aid designers as they would be able to include both technical and non-technical attributes in their analysis and subsequent design decision making.





# Appendix A

## Interview and Survey Questions

### A.1 Interview Questions for Photovoltaic Panel Installers in Boston, MA

#### A.1.1 Job Overview

- Can you tell me more about what you do, about your job title and primary responsibilities?
- In what capacity do you interact with homeowners directly?

#### A.1.2 Company

- How do homeowners find out about your company?
- How do you get new customers?

#### A.1.3 Customer Decision Process

- Pretending I'm a homeowner, and I come to you saying: I want solar panels in my house. How would you advise me? Condensed version with marketing material. Can you walk me through the steps that a homeowner undergoes to get solar power installed?

- What happens at each step?
- To what extent does the homeowner take an active role in the installation process?
- What are the choices that need to be made during consultation and design? solar panel (size, tech) (go through 1 by 1), How do you decide this? Who decides this?
- What is the range of solar panels that you offer? How do you choose the manufacturer/ solar panels?
- Does the designer or the homeowner make the choice on what type and make of solar panel? Do they want this freedom?
- Do homeowners get to choose the type of solar panel they want?
- What factors do homeowners consider when they are choosing a panel?
- How important are the details of a particular technology? (ie. silicon vs CdTe vs CIGS, toxicity, need for direct sunlight)
- What types of mounting hardware does your company offer? Is the choice of mounting dependent on the type of solar panel?
- On your web page/ From your experience, are homeowners concerned about the visual impact of the solar panels on their roof? Do they communicate their concern before, during or after the installation? What are the specific factors that they are concerned about? (prompt: glare, size, visibility, color, contrast with roof and surrounding architecture)

#### A.1.4 Homeowners

- Can you characterize homeowners that you deal with into groups?

- How well informed are they regarding the installation process? Photovoltaic technology? Financial support (subsidies)? What are these group's major concerns?
- From your interactions with users, can you give me examples of why some homeowners decide to convert to solar power? What are their primary motivating factors?
- Do some homeowners have strange or different motivations?
- Can you recount an experience where a homeowner was difficult to work with?

### **A.1.5 Installer perspective**

(Now that we are reaching the end, I want to hear your opinions on some things)

- Some companies are looking into solar shingles. What is your opinion on that?
- What advantages of solar power do you find the most compelling in persuading more people to convert to solar power?
- Is there anything that you would change about the physical system?
- What do you think are the critical factors that need to change in order for solar power to be more widely adopted?

2 final questions:

- Out of all the things we have talked about, first, is there anything that we did not cover?
- Now, as I go back to my office and stare at the data, what is the one thing that I should take away?

## A.2 Online Survey Questions for Californian Installers

### Study in Photovoltaic Panel Recommendations by Certified Installers

#### 1. Introduction

You have been asked to participate in a research study conducted by the Ideation Lab from the Mechanical Engineering department at the Massachusetts Institute of Technology (MIT). The purpose of this study is to gain a deeper understanding of how solar installers make photovoltaic panel recommendations for residential solar installations.

You have been selected as a possible participant in this study because of your expertise in advising homeowners in the selection and installation of photovoltaic panels.

This survey is voluntary. You have the right not to answer any questions, and to stop the survey at any time for any reason. We expect that the survey will take about 15 minutes to complete. Your responses will be kept confidential and no personal identifying data will be requested. At the end of the survey, you will be given a chance to enter our sweepstakes for a \$100 Amazon gift card.

Please contact Heidi at [ideation-research@mit.edu](mailto:ideation-research@mit.edu) with any questions or concerns.

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

**\*1. Clicking below indicates that I have read the description of the study and agree to participate.**

I agree to participate

## Study in Photovoltaic Panel Recommendations by Certified Installers

2.

**2. List the 5 most important things you consider when choosing between photovoltaic panel models to recommend to a homeowner.**

1 (most important)	<input type="text"/>
2	<input type="text"/>
3	<input type="text"/>
4	<input type="text"/>
5 (least important)	<input type="text"/>

## Study in Photovoltaic Panel Recommendations by Certified Installers

3.

**3. From the list below, please select 6 of the most important technical factors of a photovoltaic panel model, and rank them from 1 to 6 in order of their importance (1 being the most important factor). The remaining factors do not need to be ranked.**

	1 (most important factor)	2	3	4	5	6 (least important factor)
Module efficiency	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Black, even appearance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Weight per Watt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Workmanship warranty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Additional certifications (NEC 2008, ISO 9001, ISO 14001)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Certifications (IEC, UL)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of photovoltaic cells in module	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dimensions (length, width, height)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Power warranty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Environmental features (Lead free, optimized packaging, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Price per Watt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Product maturity (Years on market)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Study in Photovoltaic Panel Recommendations by Certified Installers

4.

### 4. How important are the following overall factors when recommending a photovoltaic panel to a homeowner?

	Very important	Important	Neutral	Unimportant	Very unimportant
Government incentives for specific models (ARRA)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall cost of installation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Module efficiency	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Weight of panel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Panel manufacturer's reputation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conformity of panel to specifications sheet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Country of origin of panel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Surface evenness of panel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Color of panel (Black, blue, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Power warranty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reliability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rated power output of panel (Watts)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Past experience dealing with panel manufacturer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of service provided by panel manufacturer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Photovoltaic technology (Monocrystalline Si, CdTe thin film, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ease of installation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Price per Watt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Durability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Product maturity (time on market)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Environmental sustainability of panel (lead-free, sustainable packaging)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Styling/aesthetics of panel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Workmanship warranty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

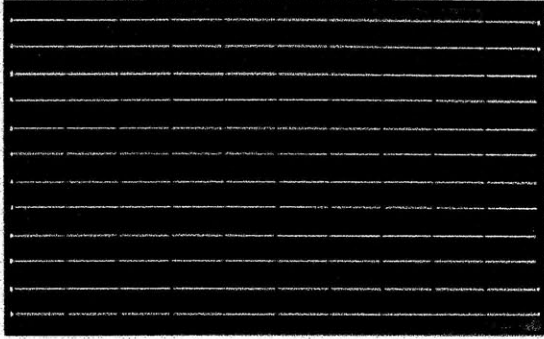
Page 4

# Study in Photovoltaic Panel Recommendations by Certified Installers

## 5.

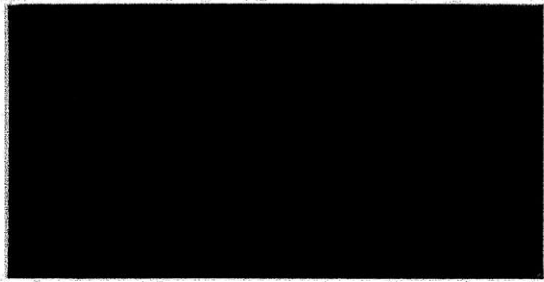
Look at the following photos of panels and consider their appearance/styling.

- **Panel A:**



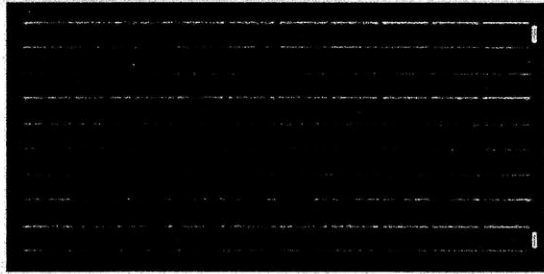
Respondents: 33.34%

- **Panel A:**



Respondents: 33.33%

- **Panel A:**

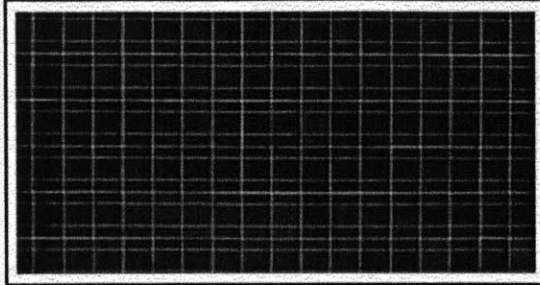


Respondents: 33.33%



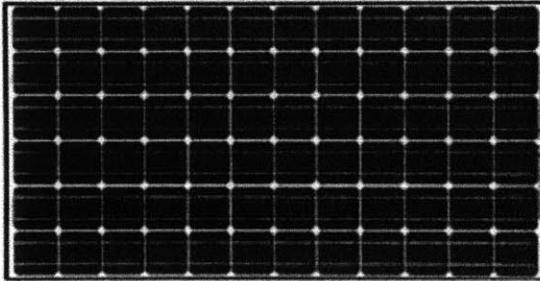
## Study in Photovoltaic Panel Recommendations by Certified Installers

- **Panel B:**



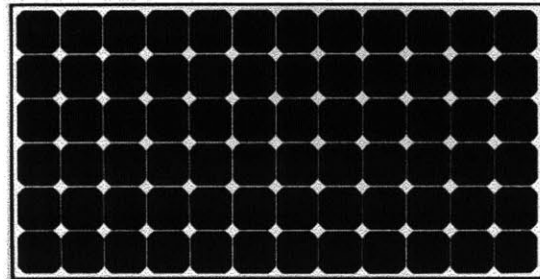
Respondents: 33.34%

- **Panel B:**



Respondents: 33.33%

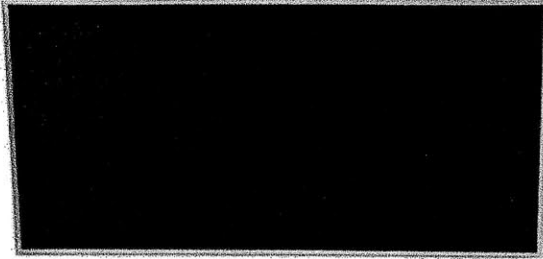
- **Panel B:**



Respondents: 33.33%

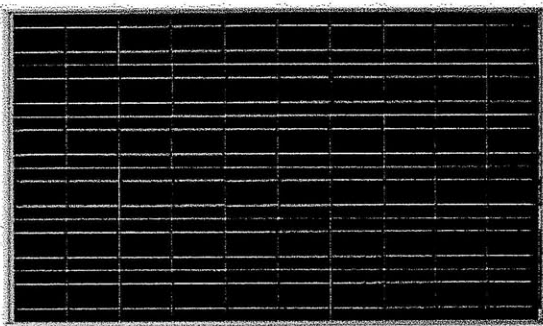
## Study in Photovoltaic Panel Recommendations by Certified Installers

- **Panel C:**



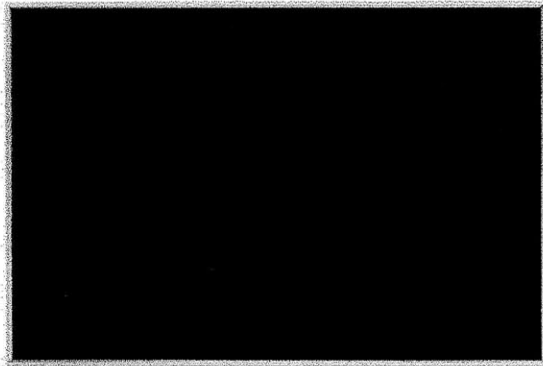
Respondents: 33.34%

- **Panel C:**



Respondents: 33.33%

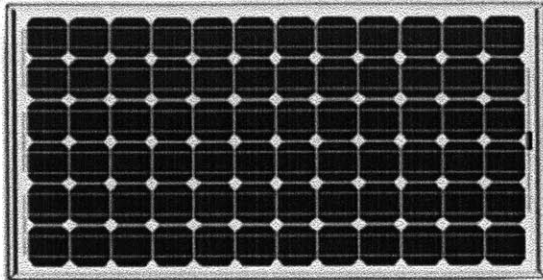
- **Panel C:**



Respondents: 33.33%

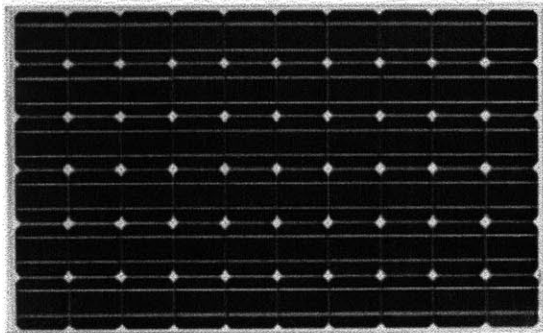
# Study in Photovoltaic Panel Recommendations by Certified Installers

- **Panel D:**



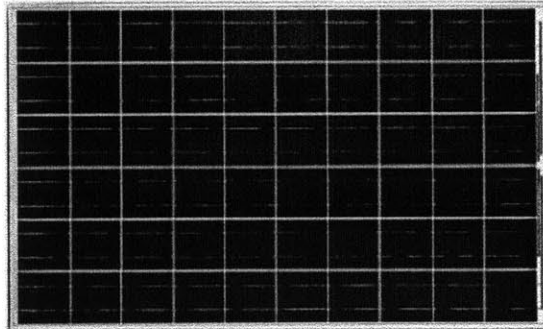
Respondents: 33.34%

- **Panel D:**



Respondents: 33.33%

- **Panel D:**



Respondents: 33.33%

## Study in Photovoltaic Panel Recommendations by Certified Installers

### 5. Having trouble displaying the images on your computer?

Check this box if you cannot view pictures of 4 panels, and skip the next 2 questions (Q6 & Q7)

### 6. Please rank these panels in order of attractiveness, from 1 being the most appealing to 4 being the least appealing.

	1 (most appealing)	2	3	4 (least appealing)
Panel A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Panel B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Panel C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Panel D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### 7. Please rank these panels in order of visual compatibility on a roof, from 1 being the most visually compatible to 4 being the least visually compatible.

	1 (most compatible)	2	3	4 (least compatible)
Panel A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Panel B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Panel C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Panel D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Study in Photovoltaic Panel Recommendations by Certified Installers

6.

### 8. How would you describe the company reputation of the following photovoltaic panel manufacturers?

	Excellent	Good	Poor	No opinion
Andalay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BP Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Canadian Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ET Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evergreen Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
First Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kyocera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mitsubishi Electric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open Energy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Phono Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
REC	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sanyo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Schuco	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sharp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Siliken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SolarWorld	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SOLON	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SunPower	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SunTech	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trina Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yes! Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yingli	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Study in Photovoltaic Panel Recommendations by Certified Installers**

**9. From your experience, how would you describe the level of service provided by the following photovoltaic panel manufacturers?**

	Excellent	Good	Poor	No experience dealing with this manufacturer
Andalay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BP Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Canadian Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ET Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evergreen Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
First Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kyocera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mitsubishi Electric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open Energy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Phono Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
REC	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sanyo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Schuco	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sharp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Siliken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SolarWorld	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SOLON	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SunPower	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SunTech	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trina Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yes! Solar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yingli	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Study in Photovoltaic Panel Recommendations by Certified Installers

### 7. Demographics

**10. What is your gender?**

- Male  
 Female

**11. Which category below includes your age?**

- 20 or younger  
 21-29  
 30-39  
 40-49  
 50-59  
 60 or older

**12. How many years of experience do you have working in the solar industry?**

- Less than 1  
 1-5  
 6-10  
 11-15  
 16-20  
 More than 20

**13. How many years have you been in your current position?**

- Less than 1  
 1-5  
 6-10  
 11-15  
 16-20  
 More than 20

## Study in Photovoltaic Panel Recommendations by Certified Installers

### 14. What is your area of involvement with the client? (Please check all that apply)

- Initial site evaluation
- Rebate/subsidy application
- System design
- Installation
- Monitoring and maintenance
- Marketing
- Other (please specify)

### 15. Which category do most of your clients belong to?

- Residential
- Commercial
- Non-profit
- Government
- Utility-scale
- Other (please specify)

### 16. Is your company a distributor/dealer for a specific manufacturer(s)?

- Yes
- No



## Study in Photovoltaic Panel Recommendations by Certified Installers

8.

**17. Thank you for participating in our survey! If you would like more information about the study, please contact [ideation-research@mit.edu](mailto:ideation-research@mit.edu).**

**If you are willing to be contacted for further questions for our research study, please enter your email address here. Additionally, if you have comments regarding any part of the survey please write them below.**

**IMPORTANT:** Click the "Done" button below to end the survey and proceed to the sweepstakes entry page for the chance to win a \$100 Amazon gift card. The information you provide for the sweepstakes will not be associated with your survey responses.



# Bibliography

- [1] Drucker, P., 1994. "The theory of the business". *Harvard Business Review*, **72**(5), Oct., pp. 95–104.
- [2] Ulrich, K. T., 2011. *Product design and development*.
- [3] Geroski, P., 2000. "Models of technology diffusion". *Research Policy*, **29**(4-5), Apr., pp. 603–625.
- [4] Rogers, E., 1984. *Diffusion of innovations*. The Free Press, New York.
- [5] Samuelson, P. A., 1938. "A note on the pure theory of consumer's behaviour". *Economica*, **5**(17), Feb., pp. 61–71.
- [6] Little, I. M. D., 1949. "A reformulation of the theory of consumer's behaviour". *Oxford Economic Papers*, **1**(1), Jan., pp. 90–99.
- [7] Samuelson, P. A., 1948. "Consumption theory in terms of revealed preference". *Economica*, **15**(60), Nov., pp. 243–253.
- [8] Houthakker, H. S., 1950. "Revealed preference and the utility function". *Economica*, **17**(66), May, pp. 159–174.
- [9] Szenberg, M., Ramrattan, L., and Gottesman, A. A., 2006. *Samuelsonian economics and the twenty-first century*. Oxford University Press, Nov.
- [10] Mark, E., 1980. "The design, analysis and interpretation of repertory grids". *International Journal of Man-Machine Studies*, **13**(1), July, pp. 3–24.
- [11] Tan, F. B., and Hunter, M. G., 2002. "The repertory grid technique: A method for the study of cognition in information systems". *MIS Quarterly*, **26**(1), Mar., pp. 39–57.
- [12] Netzer, O., and Srinivasan, V., 2011. "Adaptive self-explication of multiattribute preferences". *Journal of Marketing Research*, **48**(1), p. 140156.
- [13] Marder, E., 1999. "The assumptions of choice modelling: Conjoint analysis and SUMM". *Canadian Journal of Marketing Research*, **18**, pp. 3–14.

- [14] Cohen, S., 2003. "Maximum difference scaling: Improved measures of importance and preference for segmentation". In Sawtooth Software Conference Proceedings, Sawtooth Software, Inc, Vol. 530, pp. 61–74.
- [15] Green, P. E., Carroll, J. D., and Goldberg, S. M., 1981. "A general approach to product design optimization via conjoint analysis". *The Journal of Marketing*, **45**(3), July, pp. 17–37.
- [16] Green, P. E., and Srinivasan, V., 1990. "Conjoint analysis in marketing: New developments with implications for research and practice". *The Journal of Marketing*, **54**(4), Oct., pp. 3–19.
- [17] MacDonald, E. F., Gonzalez, R., and Papalambros, P. Y., 2009. "Preference inconsistency in multidisciplinary design decision making". *Journal of Mechanical Design*, **131**(3), Mar., pp. 031009–13.
- [18] Horsky, D., Nelson, P., and Posavac, S., 2004. "Stating preference for the ethereal but choosing the concrete: How the tangibility of attributes affects attribute weighting in value elicitation and choice". *Journal of Consumer Psychology*, **14**(1 & 2), p. 132140.
- [19] Cummings, R., Brookshire, D., Schulze, W., Bishop, R., and Arrow, K., 1986. *Valuing environmental goods: an assessment of the contingent valuation method*. Rowman & Allanheld Totowa, NJ.
- [20] Kahneman, D., and Knetsch, J. L., 1992. "Valuing public goods: The purchase of moral satisfaction". *Journal of Environmental Economics and Management*, **22**(1), Jan., pp. 57–70.
- [21] Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C., 2008. "How are preferences revealed?". *Journal of Public Economics*, **92**(89), Aug., pp. 1787–1794.
- [22] Adamowicz, W., Louviere, J., and Williams, M., 1994. "Combining revealed and stated preference methods for valuing environmental amenities". *Journal of Environmental Economics and Management*, **26**(3), May, pp. 271–292.
- [23] Agard, B., and Kusiak, A., 2004. "Data-mining-based methodology for the design of product families". *International Journal of Production Research*, **42**(15), pp. 2955–2969.
- [24] Ferguson, C. J., Lees, B., MacArthur, E., and Irgens, C., 1998. "An application of data mining for product design". In IEE Colloquium on Knowledge Discovery and Data Mining (1998/434), IET, pp. 5/1–5/5.
- [25] Kusiak, A., and Salustri, F., 2007. "Computational intelligence in product design engineering: Review and trends". *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, **37**(5), Sept., pp. 766–778.

- [26] Honda, T., Chen, H., Chan, K., Yang, M., et al., 2011. “Propagating uncertainty in solar panel performance for life cycle modeling in early stage design”. In 2011 AAAI Spring Symposium Series.
- [27] Rucks, C. T., and Whalen, J. M., 1983. “Solar-energy users in arkansas: their identifying characteristics”. *Public Utilities Fortnighly*, **111**(9), Apr., pp. 36–38.
- [28] Wander, J., 2006. “Stimulating the diffusion of photovoltaic systems: A behavioural perspective”. *Energy Policy*, **34**(14), Sept., pp. 1935–1943.
- [29] Faiers, A., and Neame, C., 2006. “Consumer attitudes towards domestic solar power systems”. *Energy Policy*, **34**(14), Sept., pp. 1797–1806.
- [30] Jetter, A., and Schweinfart, W., 2011. “Building scenarios with fuzzy cognitive maps: An exploratory study of solar energy”. *Futures*, **43**(1), Feb., pp. 52–66.
- [31] Griffin, A., and Page, A. L., 1993. “An interim report on measuring product development success and failure”. *Journal of Product Innovation Management*, **10**(4), Sept., pp. 291–308.
- [32] Griffin, A., and Page, A. L., 1996. “PDMA success measurement project: Recommended measures for product development success and failure”. *Journal of Product Innovation Management*, **13**(6), Nov., pp. 478–496.
- [33] CSI, 2011. California solar initiative: Current CSI data. [http://www.californiasolarstatistics.org/current\\_data\\_files/](http://www.californiasolarstatistics.org/current_data_files/), Sept.
- [34] SEIA, 2011. U.S. solar market insight. Executive summary, SEIA/GTM Research.
- [35] Cortes, C., and Vapnik, V., 1995. “Support-vector networks”. *Machine Learning*, **20**(3), Sept., pp. 273–297.
- [36] Obrien, R. M., 2007. “A caution regarding rules of thumb for variance inflation factors”. *Quality & Quantity*, **41**(5), Mar., pp. 673–690.
- [37] Rojas, R., 1996. *Neural networks: a systematic introduction*. Springer.
- [38] Breiman, L., 2001. “Random forests”. *Machine Learning*, **45**(1), Oct., pp. 5–32.
- [39] Friedman, J., 2001. “Greedy function approximation: a gradient boosting machine”. *Annals of Statistics*, p. 11891232.
- [40] Friedman, J., 2002. “Stochastic gradient boosting”. *Computational Statistics & Data Analysis*, **38**(4), p. 367378.
- [41] Beale, M., and Demuth, H., 1998. “Neural network toolbox”. *For Use with MATLAB, Users Guide, The MathWorks, Natick*.

- [42] Liaw, A., and Wiener, M., 2002. "Classification and regression by randomForest". *Resampling Methods in R: The boot Package*, 2(3), Dec., pp. 18–22.
- [43] Ridgeway, G., 2007. "Generalized boosted models: A guide to the gbm package". *Update*, 1, p. 1.
- [44] Laroche, M., Bergeron, J., and Barbaro-Forleo, G., 2001. "Targeting consumers who are willing to pay more for environmentally friendly products". *Journal of Consumer Marketing*, 18(6), Jan., pp. 503–520.
- [45] Gaskell, G. D., O'muircheartaigh, C. A., and Wright, D. B., 1994. "Survey questions about the frequency of vaguely defined events: The effects of response alternatives". *Public Opinion Quarterly*, 58(2), June, pp. 241–254.
- [46] Bennett, J., and Blamey, R., 2001. *The Choice Modelling Approach to Environmental Valuation*. Edward Elgar Publishing.
- [47] Courage, C., and Baxter, K., 2005. *Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques*. Gulf Professional Publishing.
- [48] Price, L. L., Feick, L. F., and Higie, R. A., 1989. "Preference heterogeneity and coorientation as determinants of perceived informational influence". *Journal of Business Research*, 19(3), Nov., pp. 227–242.
- [49] Dietrich, E., 2002. "Combining revealed and stated data to examine housing decisions using discrete choice analysis". *Journal of Urban Economics*, 51(1), Jan., pp. 143–169.
- [50] Brownstone, D., Bunch, D., and Train, K., 2000. "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles". *Transportation Research Part B: Methodological*, 34(5), p. 315338.
- [51] Hensher, D., and Bradley, M., 1993. "Using stated response choice data to enrich revealed preference discrete choice models". *Marketing Letters*, 4(2), p. 139151.