Are Solar Panels Commodities? Evidence of Quality Differences and Asymmetric Information from California

Johannes Mauritzen
BI Norwegian Business School
johannes.mauritzen@bi.no
http://jmaurit.github.io

March 23, 2018

Abstract

Solar panels should not be considered commodities. Considerable quality differences, as measured directly by degradation of production over time, are found between manufacturers. This has implications for pricing and competition in the market for solar panel systems. I test two implications from the the theory of asymmetric information of quality and find: 1.) Solar systems with high-information third party owners display higher quality than those owned by low-information hosts. 2.) Furthermore, prices of solar panels are more correlated to quality for solar panels that are owned by high-information owners. I use linear mixed effects models estimated by maximum likelihood and multilevel models estimated by Bayesian Markov Chain Monte-Carlo.

1 Introduction

Solar panels have been referred to colloquially as being commodities. This is in reference to the high degree of competition, especially from Chinese manufacturers, that has dramatically driven down prices of solar panels in the last decade. This has made solar power price competitive without subsidies in many sunny locations. But it has also meant upheaval in the panel manufacturing industry with cheaper Chinese produced panels replacing panels produced in Europe, North America and Japan. Most recently, this led to the imposition of trade tariffs on Chinese produced solar panels by the US government.

More formally, a commodity is generally defined by a certain degree of fungibility. That is, the good from a certain producer will be largely interchangeable with the good of another producer. Brent crude oil produced by Statoil from the Norwegian North Sea and that produced by BP from the UK North Sea is sold at the same price, as established by an exchange. In a similar manner, the electricity produced by a solar panel at a certain location and time can also be considered a commodity.

An important differentiating characteristic between solar panels is the quality of these solar panels. This can be measured as both the failure rate of these panels as well as the degree of degradation of the panels over time. If there were significant differences in quality between manufacturers, then this has implications for the structure and development of the solar panel market. Manufacturers with superior quality could differentiate themselves and extract a premium in the market. A reputation for quality among established manufacturers could also act as a barrier to entry, thereby reducing long-term pricing pressure.

Furthermore, if significant quality differences between manufacturers exists, then the characteristics of solar panel investments suggest the possibility of asymmetric information of quality in the market.

The market for rooftop solar panels can be expected to be particularly vulnerable to issues of asymmetric information on quality. Solar panels can be characterised as an "experience" good, where an investor/consumer needs to learn about the quality through use. In particular, poor quality panels will tend to show a higher degradation of output over time than high quality panels. Even then, solar panel owners may find it difficult to measure the degradation as it can happen gradually, over many years.

More so, solar panel systems are expected to last at least 20 years, thus for all practical purposes, their purchase can be considered a one-shot investment. This eliminates repeat buying as a mechanism for ensuring quality. In the literature, warranties are often suggested as a strong signal of quality. However, warranties may be a relatively weak assurance of quality in the market for solar panels as both contractors and manufacturers are relatively new, tend to be heavily indebted and some have recently shown a tendency to go bankrupt.

Given the difficulty of judging quality and lack of market mechanisms to signal quality, the established economic theory on the subject would suggest that the market may tend to provide low-quality panels at relatively undifferentiated prices (Tirole, 1988).

Over the period from 2010 to 2016, many of the larger solar contractors moved to leasing solar power systems to homeowners. The introduction of a leasing model could potentially get over this information asymmetry problem by shifting ownership to the large contractors that install and finance the solar systems (Mauritzen, 2017). These contractors can then in turn take steps such as testing panels and visiting manufacturing sites to ensure quality of suppliers. In other words, a descriptive understanding of the market suggests that it is possible to divide the market into two types of solar panel owners: high-information owners who lease the solar panel systems to individual home owners, businesses and other organisations, and low-information owners who buy and own the solar panel systems directly.

From this informal model of information asymmetry in the market for solar panels, a testable implication arises. The quality of panels, as measured by degradation of output over time, would be predicted to be higher in the systems owned by the high-information owners (third-party-owner) who can discern between low and high quality solar panels. Conversely, the basic theory of asymmetric information suggests that lower quality panels will tend to push out higher quality panels in the presence of low-information buyers who can not discern quality, and therefor make purchasing decision solely on price.

A related prediction from the theory of asymmetric information of quality is that the cost of solar panel systems should have a higher degree of correlation with quality when the owner is a high-information type. Because a high-information owner can, by definition, differentiate between high and low quality panels.

In this paper, I use a data set of California solar power systems that includes monthly production data from over 3000 solar panel systems. Where many studies of asymmetric information of quality must use proxies for quality, I measure it directly as the average degradation of production over time.

In summary, I have three nested hypothesis that I wish to test:

1. Solar panels are not pure commodities. Significant quality differences can be detected between manufacturers of panels.

2. High information owners (third-party-owners) will purchase higher quality panels than low-information owners (host-owners)

3. Price and quality will be more highly correlated in solar panel systems owned by high information types.

The data has a natural hierarchy of between 36 and 60 monthly observations within the 3000-plus solar panel systems, which in turn can be grouped according to relevant characteristics such as whether they are host-owned or third-party owned. Because these groupings display considerable heterogeneity in sample size, mean values and variance, it is natural to model them directly. For relatively parsimonious models with one or two hierarchies and few covariates, I can use linear mixed effects models (LMEM) estimated by maximum likelihood. However, for richer models with all relevant hierarchies, maximum likelihood becomes impractical.

As a solution, I use a Bayesian multilevel[1] model estimated using Markov Chain Monte Carlo (MCMC) simulation techniques. The main benefit of using a Bayesian multilevel model is that I can fit a complex model with many hundreds of parameters, while avoiding problems of over-fitting and issues of multiple comparison in inference by relying on regularisation of parameters through the use of higher level parameters.

The issue of how information asymmetries of quality affects market outcomes, first introduced by Akerlof (1970) is one of the main topics in the modern industrial organisation literature. Ch. 2 of Tirole (1988) and accompanying citation list provides a good overview of the theoretical foundations of this topic.

[1] Sometimes also called hierarchical

Early theoretical work of particular relevance to this paper includes Chan and Leland (1982), whom model the situation where information on quality is costly rather than directly unavailable and where quality is endogenous to the seller – in other words the seller can choose the level of quality of its product. Under these assumptions, the authors predict several equilibrium – essentially markets for both high and low quality products depending on the extent to which the buyer is informed.

Reputation, or lack thereof, plays a particularly important role in this market. Most of both the solar installers and panel manufacturers are relatively new firms, especially the increasingly dominant Chinese solar panel producers. Shapiro (1982) explores the role of reputation, and finds that because reputation is effective only with a lag, the "steady-state" quality level chosen by a seller should be lower than in the scenario with perfect information.

An expanding empirical literatures exist on asymmetric information of quality, especially in the marketing, finance and accounting fields. In accounting, the issue of firm quality and voluntary disclosure in regulatory filings as a signal has received broad attention by empirical researchers. Healy and Palepu (2001) review the early literature. Empirical studies in finance have, for example, looked at issues of adverse-selection and quality in Initial Public Offerings (Michaely and Shaw, 1994) and small business lending (Petersen and Rajan, 1994) among many other topics.

The literature in the marketing field may be more relevant to this paper, as it is particularly concerned with the effects of information asymmetry on purchases of consumer durables. The effects of signalling on quality, usually in the form of advertising, is particularly important in the field. Kirmani and Rao (2000) provide an overview of this literature.

The use of leasing, discussed in this article, can be seen as a form of costly signalling about quality, however leasing plays more than just an informational role since it also transfers much of the risk of poor quality to the contractor. Traditionally, energy generation has been an activity undertaken by large, specialised firms with specialised knowledge and resources. The role of information asymmetry in investment and purchase decisions has likely been seen as a secondary issue, and to my knowledge, no literature exists on the role of such informational issues in the industry.

However, the growth of distributed energy technologies, like solar panels, has added new focus on informational and behavioural issues related to energy investments by homeowners and small businesses. In turn, a growing literature is forming around solar power investment behaviour of homeowners and small businesses.

For example Dastrup et al. (2012) argue that solar panels can not be considered a pure investment good, but are also bundled as a type of green conspicuous consumption - showing evidence for a “solar price premium” in homes with solar panels. Bollinger and Gillingham (2012) study the the role of peer effects in solar power adoption. They find evidence that the adoption of solar panels by homeowners in a certain zip-code will increase the probability that other households in that zip-code will install solar panels.

Despite the growth of the literature on investments in distributed energy, to my knowledge this is the first paper to identify and directly test for quality differences and related role of information asymmetry in the expanding market. The article has implications for an understanding of the market for distributed energy technologies, subsidy policies and regulation of the market.

2 California Solar Initiative, Data

The data used in this article is an intersection of two datasets from the California Solar Initiative (CSI). The California Solar Initiative is a state-wide program that gives incentives for installing grid-connected solar panels systems. The incentives are based on performance. For smaller systems, typical of most residential and small business installations, this incentive could be given in the form of an upfront payout based on the capacity and the expected performance of the system.

Larger systems were required to accept a performance based incentive based on actual production over 5 years (60 months). From the beginning of the program in 2007 this was defined as those over 100kW, but was lowered to 50kW from January 2008 and 30kW from January 2010. Solar panel installations of all sizes have the option of getting an incentive based on actual performance.

CSI provides data on all grid-connected solar panel systems installed in California since January 2007. In addition, CSI provides another data set of monthly production data from the solar panel systems that received production incentives. The data is openly available on the website of CSI. [2]. A cleaned and merged data set that I use in the following analysis can be found on my website at `jmaurit.github.io#buy_or_lease`.

The installation data can be matched with the production data for those installations receiving a production subsidy. I include only those installations that have been producing for at least 3 years (36 monthly observations), where the maximum number of observations in this data is 5 years (60 months) of production data.

I removed systems from the dataset that may have had reporting errors. for example I removed systems where production was reported to be higher than what would be theoretically possible from a solar power system with a given nameplate capacity. I also removed systems that reported 0 production in a period. This could also be an indication of quality if zero production indicates a malfunction in the system. However it is not possible to identify which zero observations reflect malfunctions versus reporting issues or other non-quality issues.

Table 2 gives an overview of the variables in the dataset that are used in the analysis. Notice that variables can vary by monthly observation or by solar power system. I will take account of this natural data hierarchy in my modelling. There are a total of 171,027 monthly observations used in the analysis. These observations come from 3,145 unique solar power systems. From these solar panel systems, panels from a total of 107 unique manufacturers are used.

[2] `http://www.californiasolarstatistics.ca.gov/current_data_files/`

Variable	Label	Level	Values	Mean Value
Production (kWh)	prod_kwh	Observation	Real	35,908
Months of operation	months_operation	Observation	Integer	–
Month of year	month	Observeration	Categ.[1]	–
Cost per kw	cost_per_kw	System	Real	7,026
Capacity rating (kW)	csi_rating	System	Real	239
First year of production	first_prod_year	System	Integer	2011
Third party owned	lease	System	Categ.[2]	35% leased
Owner sector	own_sector_id	System	Categ.[3]	–
# Observations		171,027		
# Solar panel systems		3,145		
# Module manufacturers		107		
Min. months of data		36		
Max. months of data		60		

[1] January - December. [2] Leased 1, Owned 0.[3] Residential (497), Commercial (1759), Governmental (777), Non-profit (112).

Table 1:

Production is measured in kilowatt hours (kWh). In the analysis, production data is first log transformed and then descaled, as shown in equation 1. For production data for each system, s, the log production data is demeaned and then divided by 2 times the standard deviation of the production data for that solar power system. Figure 1 gives a visualisation of the transformed production data. The red line shows the average degradation over time of all the systems. The strong seasonality of production is also clearly visible.

Other continuous variables are transformed in a similar manner, with observation level variables descaled at the system level, and system-level variables descaled across systems. These transformations avoid confusion that may come when interpreting coefficients on variables with different units and scales. There are also technical reasons for the transformations relating to the MCMC routine. Conversion in the MCMC algorithm is aided when all variables are on the same scale. Setting appropriate weakly informative prior distributions on the parameters is also simplified when continuous variables are free from units.

$$prod\_scaled_{s[i]} = \frac{log(prod\_kwh_{s[i]}) - mean(log(prod\_kwh_{s[i]}))}{2std(log(prod\_kwh_{s[i]}))} \qquad (1)$$

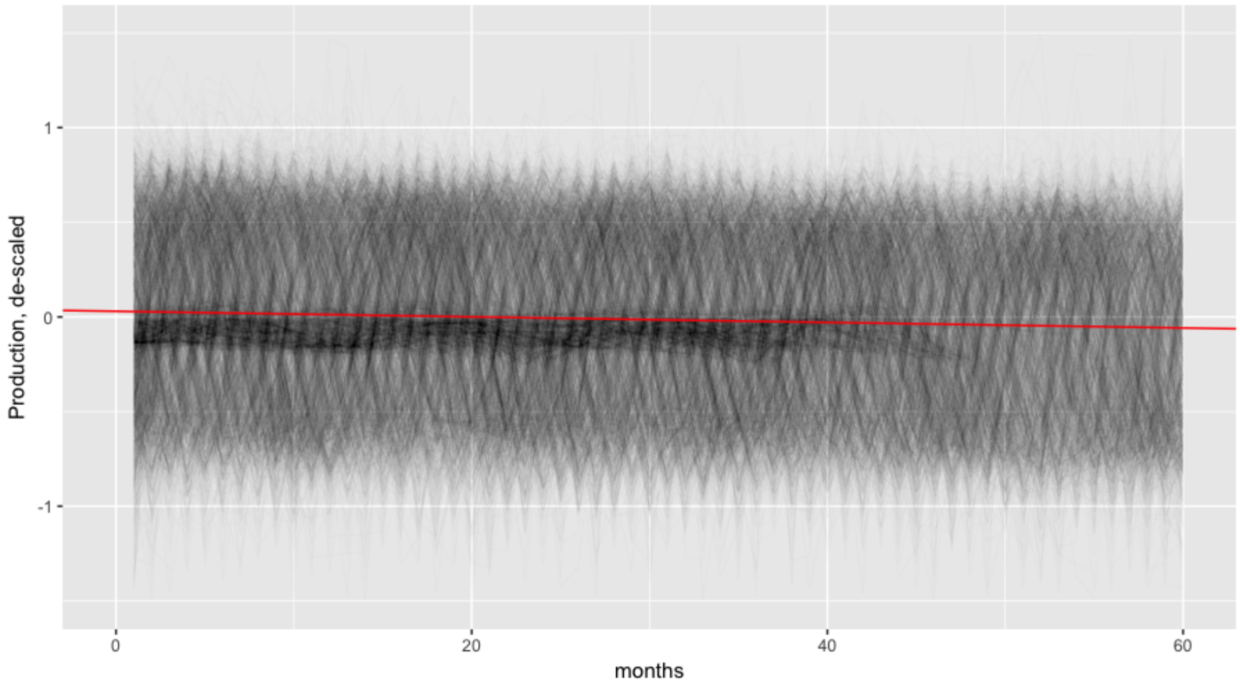

Figure 1: The figure shows the de-scaled production over time from the sample of solar panel systems. The red line represents the average degradation over time for all panels.

Month of year dummy variables are used to control for the seasonality in the data. The capacity of the solar panel system is the actual maximum capacity of a solar power system as reported to and audited by the CSI. This can be somewhat lower than the total nameplate capacity of the solar panels, and can depend on not only the technical capacity of the panels, but also on supporting equipment, such as the inverter, and factors such as the angle of the roof and orientation of the solar panel system.

The cost per kilowatt ($/kW) variable is the reported total cost of the solar panel system divided by the CSI capacity rating. For solar panel systems that are sold outright, the total cost is simply the transaction price of the entire system. For systems with third party owners, the total cost is an estimate reported by the system owner based on the sum of component costs, installation costs and a mark-up. I will discuss how this may effect the results.

In the analysis I distinguish between solar panels that are host-owned and those that are third-party owned, which I refer to as being leased. What I refer to as leased can however take two main forms. The most common is a Power Purchase Agreement (PPA) where the host agrees to buy the electricity from the solar panel system for an agreed upon price, (which may increase at an agreed upon rate over time). Alternatively, a host may pay a flat monthly fee for the electricity produced by the solar panel.

Finally, among the 3000 plus solar power systems, 497 had residential owners, 1759 had commercial owners, 777 had governmental owners and 112 had non-profit owners. A portion of the commercial owners operate as third-party-owners, with hosts in the other sectors. Alternatively, I could include host sector in my analysis: 1241 solar panel systems had a commercial host, 501 had a residential host, 1199 had a governmental host and 204 had a non-profit host. However, In a study of asymmetric information, it makes most sense to focus on the sector of the owner.

3 Linear Mixed Effects Models Estimated by Maximum Likelihood

3.1 Testing for differences between module manufacturer

Following the notation of Bates et al., the conditional distribution of the response variables, y in a multilevel model can be written compactly as in equation 3. Here β is a p-dimensional coefficient vector (or "fixed effects") of the *ntimesp* matrix of (p-1) explanatory variables $\mathbf{X}$ plus intercept term. $\mathbf{Z}$ is an $n \times q$ model matrix of the variables whose groups are to be modelled directly with corresponding vector of coefficients, or "random effects" $\mathbf{B}$ that are fixed at values $\mathbf{b}$. [3] Here q is a dependent on the number of terms, k, that are allowed to vary by group and the number of groups, l for each term: $q = \sum_i^k q_i = \sum_i^k l_i p_i$. $\mathbf{W}$ is a diagonal matrix of known prior weights. B is distributed as multivariable normal with zero mean and where Σ is variance-covariance matrix.

[3] The use of "random" refers to the fact that these groups are assumed to be specific to the sample. Many applications of mixed effects model thus treat the varying group coefficients as group-level error-terms, thus the "random" label. Still, interpreting coefficients as group varying or as group error terms is mathematically equivalent.

$$(y|\mathbf{B} = \mathbf{b}) \sim \mathcal{N}(\mathbf{X}\beta + \mathbf{Zb}, \sigma^2\mathbf{W^{-1}}) \tag{2}$$

$$B \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \tag{3}$$

Models of this form can be solved efficiently using a penalised maximum likelihood routine. For further details of both the model and routine, I refer to Bates et al..

To test whether I can detect quality differences between module manufacturers, I need to first estimate a restricted comparison model. I write the restricted comparison model explicitly in equation 4.

In this model, the transformed production variable, $prod\_scaled_i$ for observation i is distributed normally with the mean modelled as a linear function of an intercept term a_s, a slope term, b_s on the transformed months of operation variable $months\_operation\_scaled$, and a set of dummy variables representing the month of year, **month**. Notice that the intercept and slope are indexed by s, indicating that they are allowed to vary by the individual system. Thus in this model formulation I take into account the groupings by system, explicitly modelling varying relative initial production values (intercepts) and degradation over time (slope). Moving up a level, the estimated a and b parameters are in turn modelled as multivariate normal.

$$\begin{gathered} prod\_scaled_i \sim N(\mathbf{month} + a_s + b_s months\_operation\_scaled_i, \sigma^2) \\ \begin{pmatrix} a \\ b \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_b \\ \rho\sigma_a\sigma_b & \sigma_b^2 \end{pmatrix}\right) \end{gathered} \tag{4}$$

The inclusion of the matrix of month dummies **month** controls for seasonal variation across all systems. This is a simplified way to model seasonality, but because the data is itself at a monthly frequency, and the seasonality is not of direct importance, this should suffice. As long as un-modeled variation in seasonality is not correlated with the long-run slope of production, the point estimates for the slope should be unbiased.

To test whether quality varies among module manufacturers, I simply add panel manufacturer as an extra hierarchy to the model. In this model, the observation level regression is identical to the restricted comparison model as in equation 4. The system-level intercepts, a, and slope, b, terms are also modelled as a multivariate normal. But the mean values of the multivariate normal distribution, μ_a and μ_b, are now indexed by m, as shown in equation 5, representing that they are allowed to vary by module manufacturer.

$$\begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix} &\sim N \begin{pmatrix} \begin{pmatrix} \mu_{a[m]} \\ \mu_{b[m]} \end{pmatrix}, \begin{pmatrix} \sigma^2_{a[m]} & \rho\sigma_{a[m]}\sigma_{b[m]} \\ \rho\sigma_a\sigma_b & \sigma^2_b \end{pmatrix} \end{pmatrix} \\ \mu_a &\sim N(\theta_a, \sigma^2_{\mu_a}) \\ \mu_b &\sim N(\theta_b, \sigma^2_{\mu_b}) \end{aligned} \tag{5}$$

If there is no significant variation between the module manufacturer groupings, this model collapses into the restricted comparison model, which is the null hypothesis. Figure 2 gives a visual summary of the system-level slope terms ordered from lowest to highest. The lines represent approximate 95% confidence intervals around the point estimate. One anomaly we can notice is that a handful of the systems have slopes that are estimated to be positive. This is clearly problematic if we wish to interpret the slope coefficient as degradation over time. I address this in the full Bayesian model without resorting to excluding data.

3 gives a summary of the estimated manufacturer-level mean slopes. Again the lines represent 95% confidence intervals around the point estimates. Visually, we can see substantial variation in quality between the manufacturer groups. We can more formally test the hypothesis that manufacturer groupings improve the explanatory power of the model by comparing to the restricted model with an LM test. The results are shown in table 3.1. The information criteria measures (AIC, BIC) are both lower, indicating a better predictive power of the model with manufacturer groupings. The chis-square statistic also indicates that the model with manufacturer groupings substantially improves model performance.

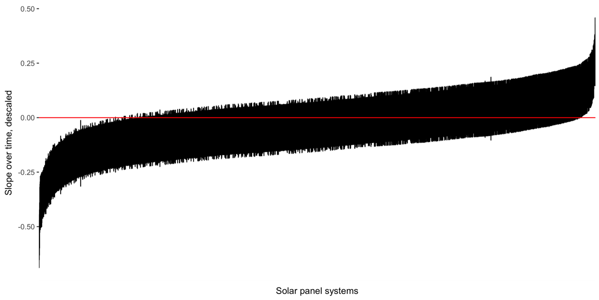

Figure 2: Approximate 95% confidence intervals around maximum likelihood point estimates of system-level slope terms, ordered from lowest to highest.

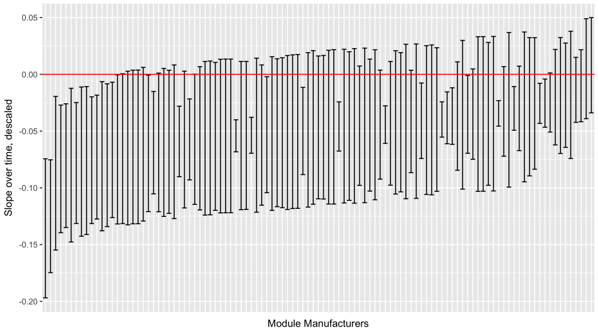

Figure 3: Approximate 95% confidence intervals around maximum likelihood estimators of manufacturer-level slope terms μ_b, ordered from lowest to highest.

Model	DF	AIC	BIC	logLik	deviance	Chisq	Chi df	$Pr(>Chisq)$
Null	17	8805	8976	-4385	8771	–	–	–
Grouped Manuf.	20	8430	8631	-4195	8390	381	3	2.6e-82

Table 2: A likelihood ratio test indicates that the addition of manufacterer as a explanatory variable significantly improves the predictive power of the model

The first result, of which the evidence appears to be quite strong is that solar panel systems do appear to vary substantially in quality between module manufacturers. Given that module quality is an important differentiating characteristic of panels, it is then difficult to formally define solar panels as commodities.

3.2 testing for differences in quality between leased and host-owned solar power systems

Provided the evidence that solar panel systems are not commodities and that substantial differences in quality do exist between manufacturers, we can now move on to the testable implication that comes from a combination of standard economic theory of asymmetric information and a descriptive understanding of the market that suggests the market for solar panels could plausibly be subject to issues of asymmetric information of quality.

From our informal model of asymmetric information where we see host owners of solar systems as being low-information types, while owners of leased panels as being high information types, we would predict that on average leased panels will display higher quality over time. To test this directly with the data, I only need to do a slight modification of the model in the previous subsection. Instead of grouping the system-level coefficients by manufacturer, I now group into whether the solar power system was leased or not, as indexed by l. We can then compare the average slopes between these groups. For convenience the model is presented in full in equations 6.

$$\begin{aligned}
prod\_scaled_i &\sim N(\mathbf{month} + a_s + b_s months\_operation\_scaled_i, \sigma^2) \\
\begin{pmatrix} a \\ b \end{pmatrix} &\sim N\begin{pmatrix}\begin{pmatrix} \mu_{a[l]} \\ \mu_{b[l]} \end{pmatrix}, \begin{pmatrix} \sigma^2_{a[l]} & \rho\sigma_{a[l]}\sigma_{b[l]} \\ \rho\sigma_a\sigma_b & \sigma^2_b \end{pmatrix}\end{pmatrix} \\
\mu_a &\sim N(\theta_a, \sigma^2_{\mu_a}) \\
\mu_b &\sim N(\theta_b, \sigma^2_{\mu_b})
\end{aligned} \tag{6}$$

A visual summary of the results are shown in figure 4. Again, the vertical lines represent approximate 95% confidence intervals over the point estimates for the system-level slope terms, b_s. But here the the slope estimates are grouped into host-owned (left) and leased (right). The estimated group means are shown as horizontal lines. At the scale of the figure, the effect appears small, but the average slopes are estimated to be significantly different from each other at the 95% confidence interval. The difference is also economically significant. Extrapolating out from the difference in the point estimates between leased and host-owned suggests that on average a leased system can expect to be producing 11% more power after 10 years compared to an otherwise identical solar panel systems.

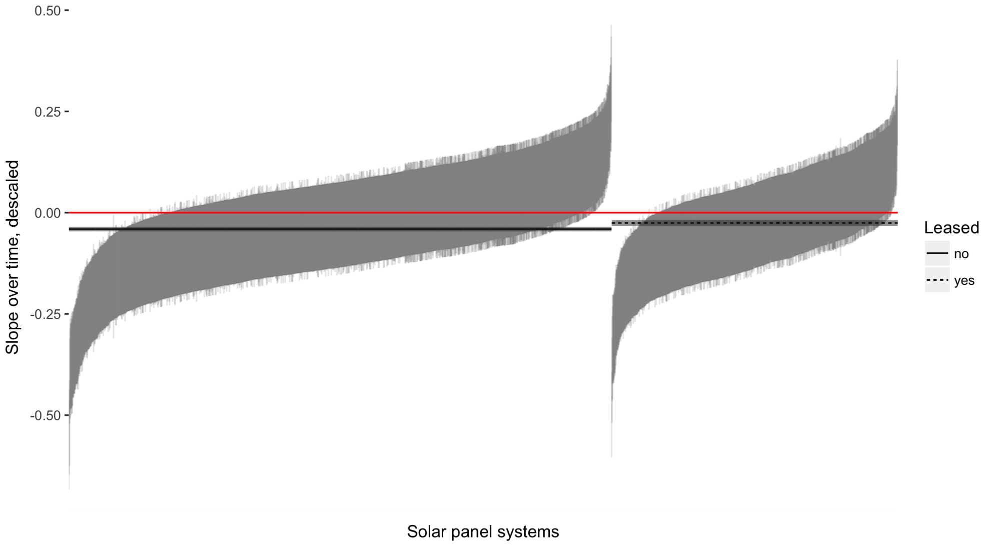

Figure 4: A visual summary of the estimation results. The vertical lines again represent the estimated system-level slope coefficients.

4 Full Bayesian multilevel model estimated by Markov Chain Monte Carlo

Until now, I have used relatively parsimonious models with few fixed-effects covariates and only one or maximum two hierarchies (”random effects”). Ideally, I would like to estimate a full probability model with all relevant and available covariates and hierarchies. First and foremost so that the parameters of interests to the testable implications can be estimated from one model. But also, Barr et al. (2013) shows, using Monte-Carlo simulations, that mixed effects models that have the maximum random effects structure justified by the model and data, provide the best out-of-sample prediction. Estimating mixed effects model by maximum likelihood is efficient and straight-forward. However, convergence of the maximum likelihood routine is not guaranteed under richer models with more hierarchies and larger number of parameters to be estimated.

Specific to my model, I wish to estimate a model where the variable for cost per-kW is included as an interaction that varies by system and module manufacturer. I also wish to have a model where both system, module manufacturer and leasing hierarchies are estimated simultaneously. I am not able to obtain convergence of the maximum likelihood routines under this rich model.

As a solution, I build a fully Bayesian multilevel model[4] estimated by Markov Chain Monte Carlo (MCMC) simulation techniques.

The main advantage to using the Bayesian techniques here are practical rather than philosophical. Models up to levels of nearly arbitrary richness can be estimated, with computation time and available data being the main constraints. Inference is also well defined and exact.[5]

A Bayesian MCMC approach has some additional attractive properties. Results are presented as estimates of marginal posterior distributions over coefficients of interest, which can be interpreted directly and intuitively as probabilities. Group-level variances are not constrained to be equal and assumptions of normality of the likelihood and posterior distributions can be loosened. For further discussion of the Bayesian frameworks and techniques, I refer to the excellent discussions in Gelman et al. (2013), Kruschke (2014), or McElreath (2015).

The model is presented in equations x through y. The response variable is production data that has been log transformed and scaled as described earlier. This response variable is assumed to have a normal likelihood with mean, $\hat{y}$ and standard deviation, σ. At the observation level, the fitted values, $\hat{y}$, are modelled as in equation **??**. This observation level regression, as before, is a simple linear regression with an intercept α, and a slope β which are both allowed to vary by system. In addition a vector of month dummies, **month** is included in order to control for seasonality.

[4] I purposefully change my nomenclature here and refer to the Bayesian model as a ”multilevel model” to distinguish it from the models estimated by maximum likelihood, which I refer to as a linear mixed effects model. This distinction is however mostly artificial. Both models resemble each other closely in structure. ”Multilevel model” is however a broader term. There is no restriction to linearity, for example. Models with non-linear or non-parametric components can also be incorporated and estimated with MCMC.

[5] Exact in the sense that given a correctly specified probability model, then an MCMC routine that is given infinite time to sample, will in theory converge to the correct posterior probability distribution. Maximum likelihood is approximate in the sense that inference is based on asymptotic theory and assumptions of normality.

Equations 7 describe the model at the system level. The intercept terms, a_s are given a common higher level normal distribution with mean μ^a and standard deviation $sigma^a$. But because the intercept term is not of direct economic importance, it is not modelled further.

The slope terms, b_s are again modelled directly. I decompose the variation into three mean terms that vary by lease, μ_l^{lease}, host sector μ_c^{sect}, and manufacturer μ_m^m. In addition three continuous variables are included in the system-level equation: the year of first production, $first\_prod\_year_s$, log capacity, $log\_capacity_s$ and the log system cost per kilowatt, log_cost_s with corresponding coefficients β^{fy}, β^{size}, and β^{cost}.

$$\begin{aligned}
&prod\_scaled \sim N(\hat{y}, \sigma) \\
&\qquad\qquad \hat{y} = a_s + b_s months\_operation_i + \mathbf{month} \\
&\qquad\qquad a_s \sim N(\mu^a, \sigma^a) \\
&\qquad\qquad b_s = \mu_l^{lease} + \mu_s^{sect} + \mu_m^{manuf} + \beta^{fy} first\_prod\_year_s + \beta^{size} log\_csi\_rating_s + \beta^{cost} log\_cost_s
\end{aligned} \tag{7}$$

The μ terms are all in term modelled as coming from a Cauchy distribution with a location meta-parameter μ^{meta} and scale parameter σ^{mu}, which are also estimated by the model. These meta-parameters are in turn given Cauchy prior distributions with location 0 and scale 5.

The $Cauchy(0, 5)$ prior is what is referred to as a weakly-informative prior. It provides some information to the MCMC algorithm for where to explore for probability mass, but does not impose any strong prior knowledge that meaningfully effects the posterior distribution. The Cauchy distribution, which is a T-distribution with 1 degree of freedom, has wide tails and thus allows for the occasional large coefficient in a lower level regression, but still provides some shrinkage towards zero (Gelman et al., 2008). The coefficients β^{fy}, β^{size}, β^{cost} and **month** are all also given $Cauchy(0, 5)$ prior distributions.

$$\begin{aligned}
\mu^{lease} &\sim Cauchy(\mu^{meta}, \sigma^{\mu}) \\
\mu^{sect} &\sim Cauchy(\mu^{meta}, \sigma^{\mu}) \\
\mu^{manuf} &\sim Cauchy(\mu^{meta}, \sigma^{\mu}) labeleqn : meta \qquad (8) \\
\mu^{meta} &\sim Cauchy(0, 5) \\
\sigma^{mu} &\sim half - Cauchy(0, 5)
\end{aligned}$$

4.1 Computation and Results from Bayesian multilevel model

The model is estimated using the Stan probabilistic programming language and sampler, which implements Hamiltonian MCMC and a No U-turn sampler (NUTS) (Stan Development Team, 2014). The sampler was run with four chains and 1000 iterations. R-statistics of 1 indicate convergence in the model. 2000 samples are taken from the estimated posterior distribution, and model results for the marginal distributions of the parameters of interest are presented in the form of histograms of the samples.

Recall, the testable implications from the theory of asymmetric information and quality were that 1.) average quality should be higher in the panels owned by the high-information type (leased panels). 2.) Price should be more highly correlated with quality in the panels owned by the high-information type.

The parameters of interest in our model are then the higher level parameters μ_{lease} and β. The first represents the mean value of the system level slope terms, b grouped by whether the systems were leased or host-owned. The latter is a measure of the correlation between the cost of the system and the system slope terms. The β parameter is also allowed to vary by leased or host-owned. It is important to note, that the estimated β paramaters are estimated conditional on the installation date of the solar panels, thus they do not reflect changes in price over time. This could lead to a heavy bias in the results if, as seems plausible, both prices and quality changed over time.

For the testable implications, the inference we are interested is in the contrasts, $\mu_{lease,yes}-$

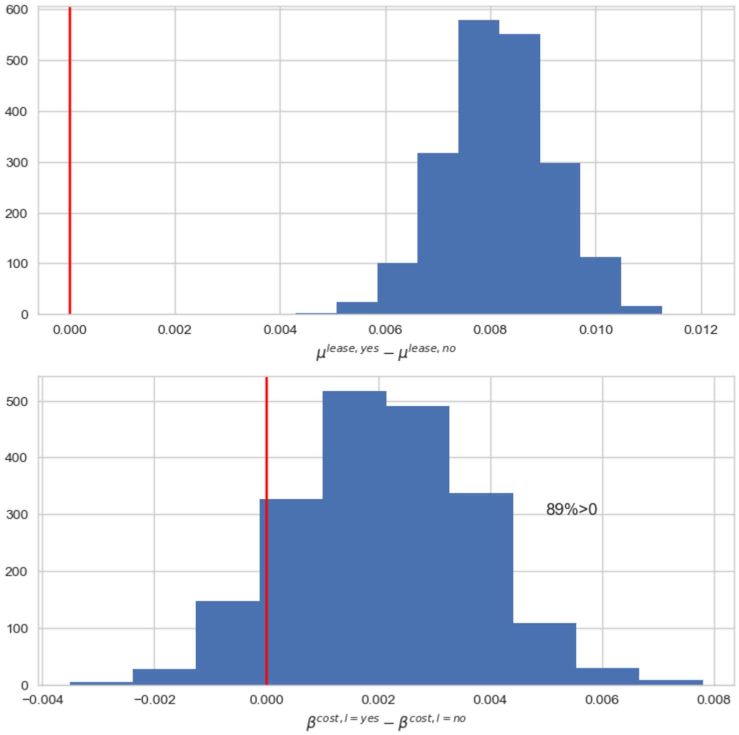

Figure 5: Marginal posterior distributions for the contrasts of interest. The top panel shows that the distributions of the slope parameters, b_s, are estimated to have means that are significantly different from each other between leased and not leased systems as defined as the contrast, $\mu_{lease,yes} - \mu_{lease,no}$. The lower panel shows the marginal posterior distribution of the contrast between the β parameter for leased and not-leased solar systems. The results indicate that there is an approximately 89% probability that costs are more highly correlated with quality in leased systems compared to those that are host-owned.

$\mu_{lease,no}$ and $\beta_{leased} - \beta_{owned}$. Histograms representing the marginal distributions of these contrasts are presented in figure 5. The top panel shows the distribution of the contrast $\mu_{lease,yes} - \mu_{lease,no}$. Essentially all of the probability mass is located in the positive range. This reinforces the finding from the mixed effects model that leased panels due appear to have substantially higher quality, as the theory of asymmetric information of quality would suggest.

The evidence for the second testable implication is less conclusive. The model estimates that there is approximately a 89% probability that there is a higher correlation between quality and cost for the high information type owner of leased panels compared to the low information type owner of host-owned panels.

A direct comparison of prices and costs between these groups is however complicated. The prices of host-owned panels are market prices: simply the purchase price that the owner of the solar panel system paid to the contractor. On the other hand, the price of leased systems is a reported price from the contractors consisting of the component and installation costs plus an estimated mark-up. If these costs were reported correctly, then the theory should still hold. The reported cost should reflect the ability of the high-information owners to judge quality of the panels that they purchase. In practice, the reliability of reported cost is a source of added uncertainty, and should be taken into consideration when interpreting the results.

When that is said, if we assume that misreporting of the true costs will tend to lead to less variation in reported costs, then that should have the affect of attenuating the magnitude of the contrast, $\beta_{cost,l=yes} - \beta_{cost,l=no}$. In that case, the estimated contrast can be seen as being a conservative estimate.

Other higher level parameters are reported in table 4.1. The sector groups can be of interest in a discussion of the role of asymmetric information. Figure 6 shows the distributions of the contrasts relative to the residential host sector. The top panel represents the contrast with commercial sector hosts, the middle with government sector host and the bottom panel with non-profit sector host. All three contrasts indicate relatively lower quality for installations for these sectors compared to the residential sector.

While it is not clear why this should be, an explanation consistent with the theory of asymmetric information of quality, is that there exists an agency problem. A homeowner will, in some form, have a direct financial incentive to acquire solar panels of good quality. On the other hand, A manager or employee of a firm, government agency or non-profit may not have any direct financial incentive in ensuring that a solar panel system is of high quality.

Summary statistics for the parameters on the manufacturer group means as well as the lower-level intercept and slope paramaters a_s and b_s are too numerous to present in a table. Visual summaries are presented in figures 7, 8 and 9. Figure 7 shows box plots of the

	mean	std	min	2.5%	25%	50%	75%	97.5%	max
μ^{meta}	0.004’	0.022’	-0.018’	-0.010’	-0.007’	-0.005’	-0.002’	0.069’	0.084’
β^{meta}	0.003’	0.838’	-8.120’	-1.707’	-0.074’	-0.007’	0.075’	1.738’	6.787’
$\mu^{meta,mon}$	0.005’	0.036’	-0.138’	-0.065’	-0.019’	0.006’	0.030’	0.075’	0.146’
μ^{b_s}	-0.000’	0.015’	-0.043’	-0.028’	-0.012’	-0.002’	0.011’	0.031’	0.044’
μ^{jan}	-0.164’	0.013’	-0.188’	-0.186’	-0.173’	-0.163’	-0.154’	-0.144’	-0.142’
μ^{feb}	-0.121’	0.013’	-0.146’	-0.143’	-0.131’	-0.120’	-0.111’	-0.102’	-0.100’
μ^{mar}	-0.054’	0.013’	-0.078’	-0.076’	-0.064’	-0.053’	-0.044’	-0.035’	-0.033’
μ^{apr}	0.040’	0.013’	0.017’	0.018’	0.031’	0.042’	0.051’	0.060’	0.062’
μ^{may}	0.087’	0.013’	0.063’	0.064’	0.077’	0.088’	0.097’	0.106’	0.108’
μ^{jun}	0.111’	0.013’	0.087’	0.089’	0.102’	0.113’	0.121’	0.131’	0.133’
μ^{jul}	0.108’	0.013’	0.084’	0.086’	0.099’	0.110’	0.118’	0.128’	0.130’
μ^{aug}	0.101’	0.013’	0.077’	0.079’	0.092’	0.103’	0.112’	0.121’	0.123’
μ^{sep}	0.082’	0.013’	0.058’	0.060’	0.072’	0.083’	0.092’	0.102’	0.103’
μ^{oct}	0.026’	0.013’	0.002’	0.004’	0.016’	0.027’	0.036’	0.045’	0.047’
μ^{nov}	-0.031’	0.013’	-0.056’	-0.053’	-0.041’	-0.030’	-0.021’	-0.011’	-0.010’
μ^{dec}	-0.123’	0.013’	-0.147’	-0.146’	-0.133’	-0.122’	-0.113’	-0.104’	-0.101’
$\mu^{lease,no}$	-0.030’	0.045’	-0.193’	-0.159’	-0.015’	-0.011’	-0.008’	-0.002’	0.015’
$\mu^{lease,yes}$	-0.022’	0.045’	-0.186’	-0.151’	-0.007’	-0.003’	0.000’	0.006’	0.023’
$\mu^{sect,com}$	0.002’	0.022’	-0.024’	-0.015’	-0.010’	-0.007’	-0.004’	0.066’	0.087’
$\mu^{sect,res}$	0.012’	0.022’	-0.014’	-0.005’	-0.000’	0.003’	0.006’	0.075’	0.096’
$\mu^{sect,gov}$	0.006’	0.022’	-0.017’	-0.010’	-0.005’	-0.003’	0.001’	0.071’	0.092’
$\mu^{sect,npr}$	-0.001’	0.022’	-0.026’	-0.018’	-0.012’	-0.010’	-0.006’	0.063’	0.084’
$\beta^{cost,l=no}$	-0.002’	0.002’	-0.007’	-0.005’	-0.003’	-0.002’	-0.001’	0.001’	0.004’
$\beta^{cost,l=yes}$	0.001’	0.001’	-0.002’	-0.001’	0.000’	0.001’	0.001’	0.002’	0.003’
β^{size}	0.002’	0.001’	-0.001’	0.000’	0.002’	0.002’	0.003’	0.004’	0.006’
β^{fy}	-0.018’	0.001’	-0.021’	-0.020’	-0.018’	-0.018’	-0.017’	-0.016’	-0.014’
σ	0.065’	0.000’	0.065’	0.065’	0.065’	0.065’	0.065’	0.066’	0.066’
σ^{b_s}	0.476’	0.006’	0.454’	0.465’	0.472’	0.476’	0.480’	0.488’	0.495’
σ^{mon}	0.114’	0.030’	0.059’	0.074’	0.094’	0.108’	0.128’	0.185’	0.341’
σ^{β}	0.833’	1.773’	0.001’	0.006’	0.047’	0.195’	0.836’	5.675’	19.879’
σ^{mu}	0.006’	0.001’	0.003’	0.004’	0.005’	0.006’	0.006’	0.008’	0.011’

Table 3: Summary statistics of the estimated posterior distributions of higher level variables

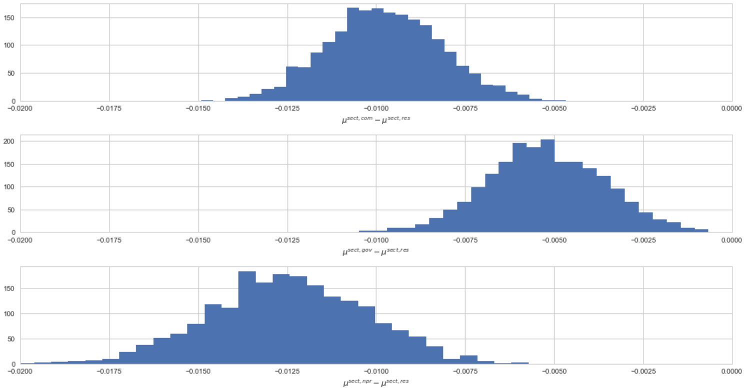

Figure 6: Distribution of the sector contrasts relative to residential sector. Installations for commercial, governmental and non-profit hosts are shown to have on average lower quality then installations in residential

manufacturer distributions, with the box representing 70 percent of probability and the lines representing 95 percent of the probability. Mirroring earlier results, the plot shows most manufacturers with roughly equivalent quality, with handful with meaningfully lower quality. Figures 8 and 9 show 95% intervals of the distribution of the a_s and b_s parameters. The intercept terms are shown to have considerable variation between systems, as would be expected, but are individually estimated precisely.

The slope parameters, b_s are modelled explicitly and are also allowed to have varying variance parameters. This is evident in the plot. A similar pattern to the higher-level manufacturer group means is apparent: modest differences in quality are seen between most systems, with the exception of a lower tail of systems with particularly poor quality.

5 Discussion and Conclusion

References

George A. Akerlof. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, August 1970. doi: 10.2307/1879431. URL http://www.jstor.org/stable/1879431.

Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*,

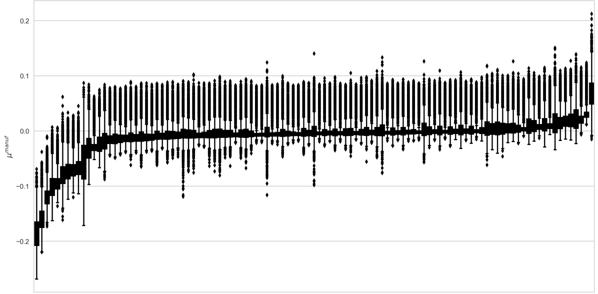

Figure 7: Box plot of the the parameters μ^{manuf}, representing the distributions of the mean parameter for manufacturer groupings.

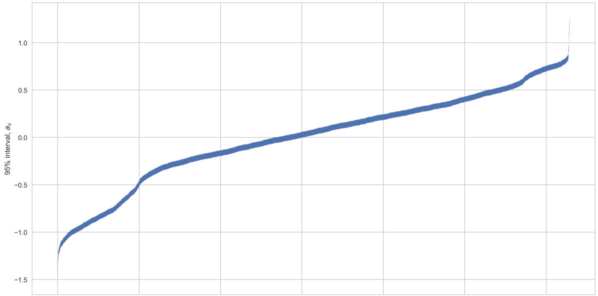

Figure 8: 95% interval of the distributions of the system level intercept parameters, a_s.

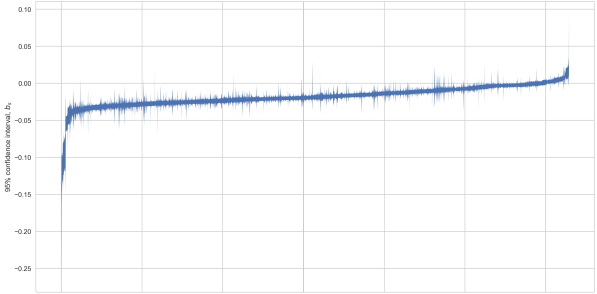

Figure 9: 95% interval of the distributions of the system level slope parameter, b_s

68(3):255–278, April 2013. ISSN 0749-596X. doi: 10.1016/j.jml.2012.11.001. URL http://www.sciencedirect.com/science/article/pii/S0749596X12001180.

Douglas Bates, Martin Mchler, Steven C. Walker, and Benjamin Bolker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. doi: 10.18637/jss.v067.i01. URL https://www.jstatsoft.org/article/view/v067i01/0.

Bryan Bollinger and Kenneth Gillingham. Peer Effects in the Diffusion of Solar Photovoltaic Panels. *Marketing Science*, 31(6):900–912, September 2012. ISSN 0732-2399. doi: 10.1287/mksc.1120.0727. URL http://pubsonline.informs.org/doi/abs/10.1287/mksc.1120.0727.

Yuk-Shee Chan and Hayne Leland. Prices and Qualities in Markets with Costly Information. *The Review of Economic Studies*, 49(4):499–516, 1982. ISSN 0034-6527. doi: 10.2307/2297283. URL http://www.jstor.org/stable/2297283.

Samuel R. Dastrup, Joshua Graff Zivin, Dora L. Costa, and Matthew E. Kahn. Understanding the Solar Home price premium: Electricity generation and Green social status. *European Economic Review*, 56(5):961–973, July 2012. ISSN 0014-2921. doi: 10.1016/j.euroecorev.2012.02.006. URL http://www.sciencedirect.com/science/article/pii/S0014292112000244.

Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360–1383, December 2008. ISSN 1932-6157, 1941-7330. doi: 10.1214/08-AOAS191. URL http://projecteuclid.org/euclid.aoas/1231424214.

Andrew Gelman, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. *Bayesian Data Analysis, Third Edition*. Chapman and Hall/CRC, Boca Raton, 3 edition edition, November 2013. ISBN 978-1-4398-4095-5.

Paul M Healy and Krishna G Palepu. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(13):405–440, September 2001. ISSN 0165-4101. doi: 10.1016/S0165-4101(01)00018-0. URL http://www.sciencedirect.com/science/article/pii/S0165410101000180.

Amna Kirmani and Akshay R. Rao. No Pain, No Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality. *Journal of Marketing*, 64(2):66–79, 2000. ISSN 0022-2429. URL http://www.jstor.org/stable/3203443.

John K. Kruschke. *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*. Academic Press, Boston, 2 edition edition, November 2014. ISBN 978-0-12-405888-0.

Johannes Mauritzen. Cost, Contractors and Scale: An Empirical Analysis of the California Solar Market. *The Energy Journal*, Volume 38(Number 6), 2017. URL https://ideas.repec.org/a/aen/journl/ej38-6-mauritzen.html.

Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC, Boca Raton, 1 edition edition, December 2015. ISBN 978-1-4822-5344-3.

Roni Michaely and Wayne H. Shaw. The Pricing of Initial Public Offerings: Tests of Adverse-Selection and Signaling Theories. *Review of Financial Studies*, 7(2):279–319, April 1994. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/7.2.279. URL http://rfs.oxfordjournals.org/content/7/2/279.

Mitchell A. Petersen and Raghuram G. Rajan. The Benefits of Lending Relationships: Evidence from Small Business Data. *The Journal of Finance*, 49(1):3–37, March 1994. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1994.tb04418.x. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1994.tb04418.x/abstract.

Carl Shapiro. Consumer Information, Product Quality, and Seller Reputation. *The Bell Journal of Economics*, 13(1):20–35, 1982. ISSN 0361-915X. doi: 10.2307/3003427. URL http://www.jstor.org/stable/3003427.

Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.2*. 2014. URL http://mc-stan.org/.

Jean Tirole. *The Theory of Industrial Organization*. The MIT Press, Cambridge, Mass, 1st edition edition, 1988. ISBN 978-0-262-20071-4.