Solar Power Savings

A Time Series Analysis

Kaylee Hodgson and Michael Christiansen

Brigham Young University

Table of contents

Introduction

Background

- Solar energy sources have become an excellent, environmentally friendly alternative to fossil fuels, especially with the increasing demand for energy in a modern and highly industrialized society.
- Solar energy is particularly appealing for homeowners who can, in the long run, save money by using solar energy instead of making monthly payments to a power company that uses fossil fuels.

Motivation

1. Sales Tactics
    - Solar companies look for sales tactics to convince customers to buy solar panels for their homes.
    - The upfront cost for solar panels is large, which makes it difficult to convince customers that solar panels could end up saving them money.
2. Customer Savings
    - Customers would benefit from being able to more accurately budget for the cost of power

Questions of Interest

1. How much per month do customers with solar panels save on average compared to those without?
2. How much time on average will it take for a customer who has bought solar panels to recoup the initial cost (in this case $8,000)?

Goal of Analysis

1. Inference: Determine how much the customer will save on average by switching to solar
2. Prediction: Predict what bills are going to be in the future to determine how long it will take to recoup the $8,000

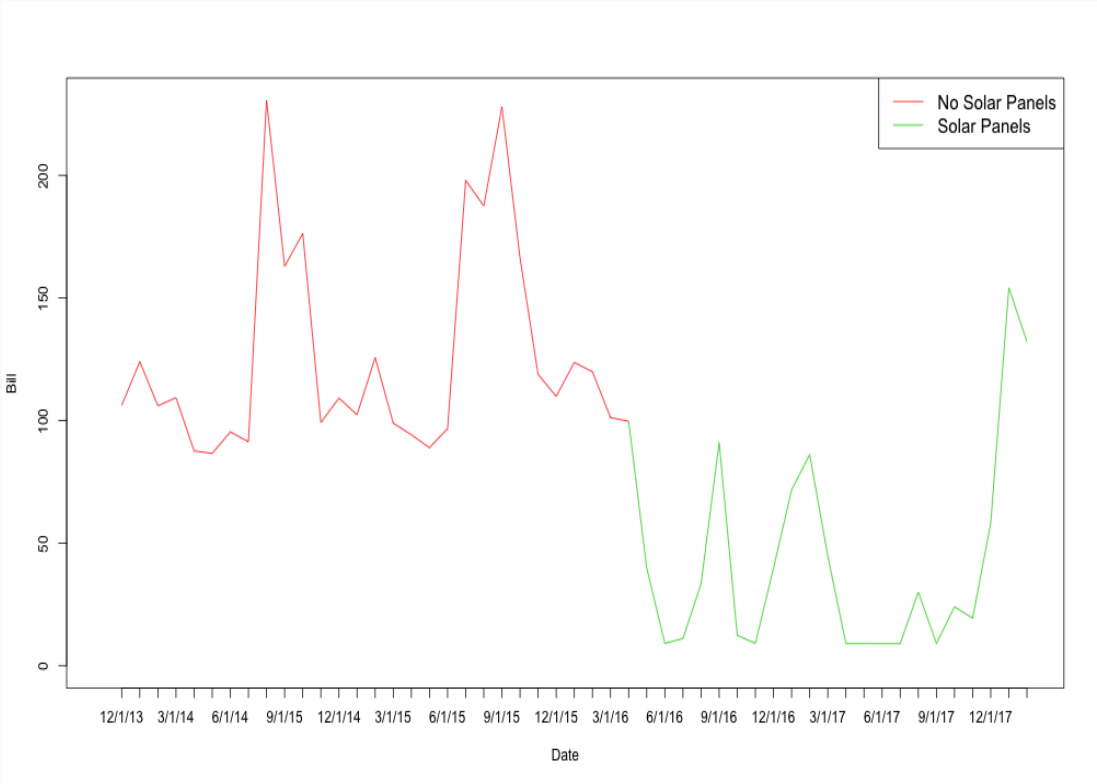
Exploring the Data

The Data

- This dataset has 51 sequential observations of the monthly power bills for a single customer, along with an indicator for whether the customer had solar panels that month or not.
- The first 29 months the customer did not have solar panels. The last 22 months, they did.

The Data

Figure 1: Customer's Monthly Power Bill



The data is correlated in time, so a simple linear regression model, which assumes independence of the observations, will give inaccurate uncertainty measurements.

Methods

Correlated Generalized Least Squares Model

Generalized Least Squares does not assume a distribution

$$Y \sim N(X\beta, \sigma^2 R) \Rightarrow Y \sim N(X\beta, \Sigma), \tag{1}$$

where R depends on the covariance structure We estimate the MLE's of $\hat{\beta}$ and the $\hat{\sigma}^2$ as follows:

$$\hat{\beta} = (X'R^{-1}X)^{-1}X'R^{-1}Y \tag{2}$$

$$\hat{\sigma}^2 = \frac{1}{N}(Y - X\beta)'R^{-1}(Y - X\beta) \tag{3}$$

Covariance Structures

- Lag-1 Autoregressive Process: adjusts the correlation value according to the distance of the observations, decreasing the correlation as the time between two observations increases
- Moving Average Process: only accounts for correlation between consecutive time points
- Exponential Correlation: can be used for unequally spaced time periods

Because our data is equally spaced and we suspect that correlation decreases as time between observations increases, we choose the Lag-1 Autoregressive Process.

Lag 1 Autoregressive

Covariance Structure

$$R = \sigma^2 \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & 1 & \phi & \dots & \phi^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{T-1} & \dots & \dots & \dots & 1 \end{bmatrix} \tag{4}$$

Model Variables

- In our model, we decided to include fixed effects for whether a customer is using solar power, and interaction terms between power type and the seasons of Summer and Winter
- These interactions were included due to seasonal variability in the efficiency of different power types.
- Our covariance structure is based on time between observations

Gaussian Process Regression Assumptions

- Data is multivariate normal - residuals should be normally distributed
- Constant variance

To check the assumptions of the model, we used Cholesky Decomposition to decorrelate the residuals

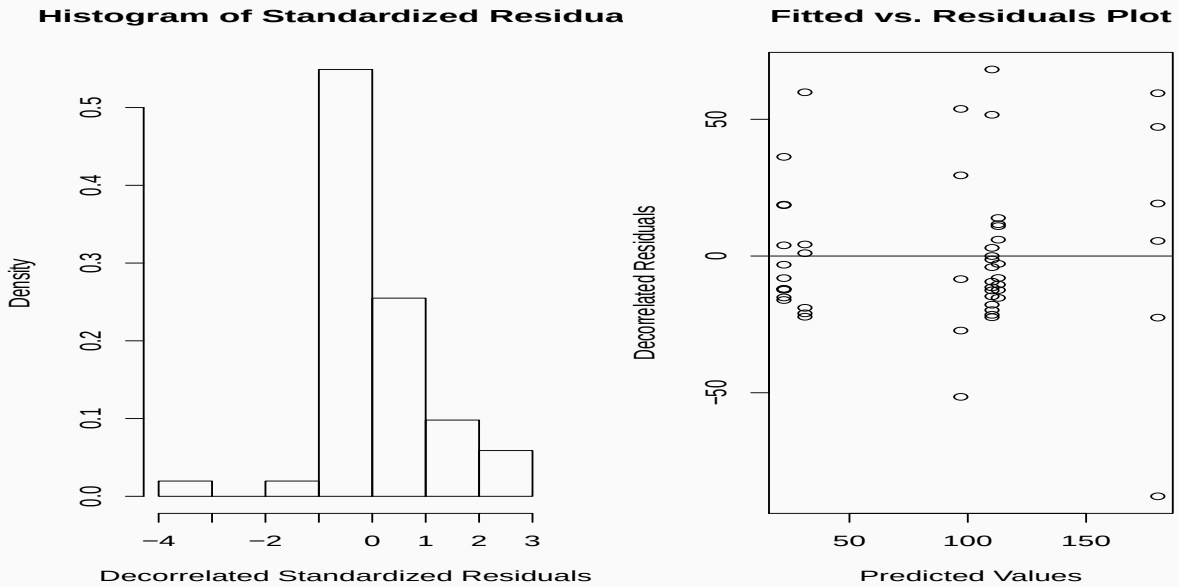


Figure 2: Model Assumptions Check

Model Fit

Results from Cross-Validation Study

- Mean Bias = -0.557
- RPMSE = 34.38
- Mean Prediction Interval Coverage = 0.842
- Mean Prediction Interval Width = 104.098

Percentage of the Variance Explained Using the Decorrelated Residuals

- R-squared = 0.77

Results

Data with Model Fitted

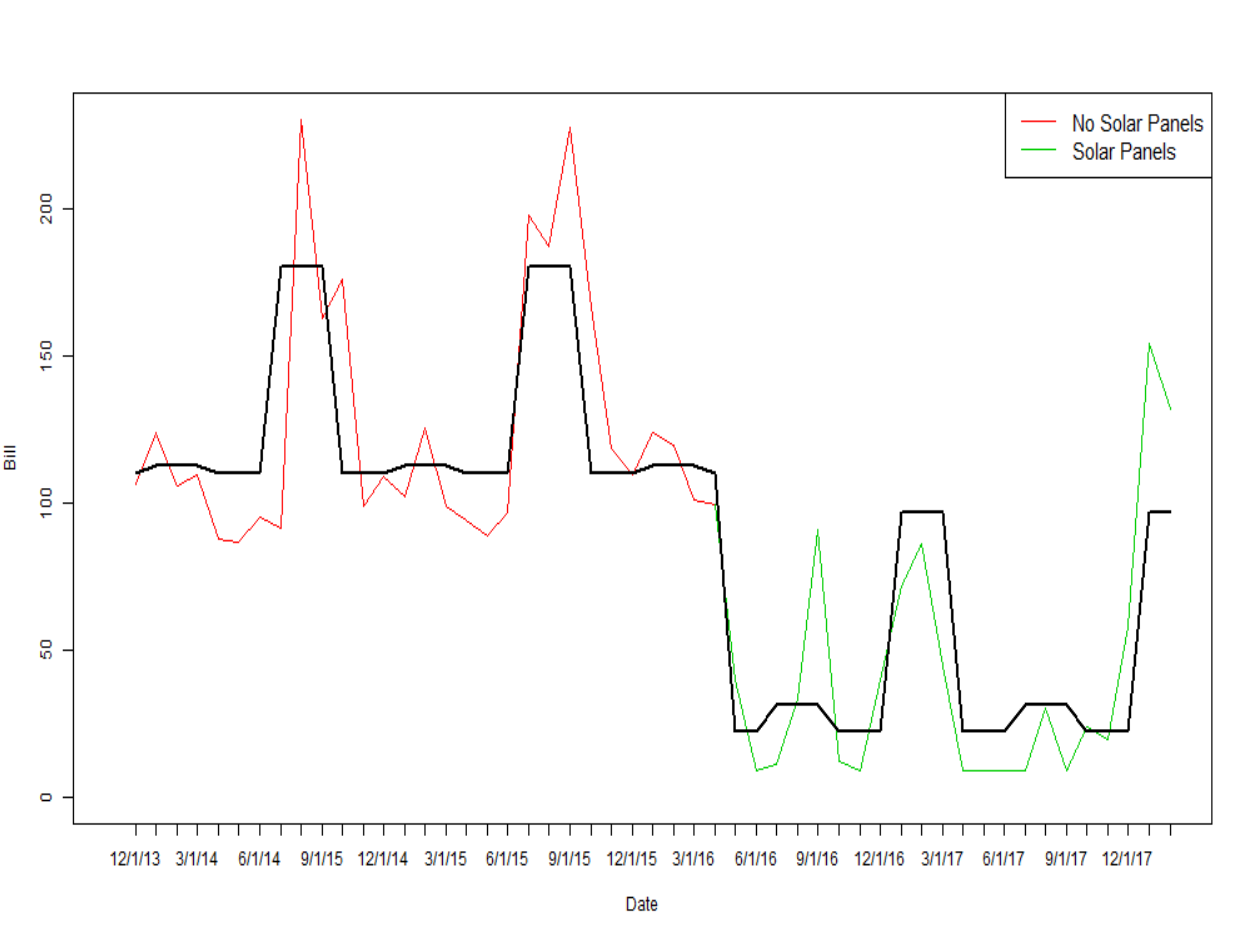


Figure 3: Model Fitted to Power Bill Data

Model Coefficients

Table 1: AR(1) Model Coefficient Estimates

Coefficient	Estimate	95% CI	p-value
(Intercept)	110.241309	(93.25474,127.22788)	0.0000
SolarY	-87.910388	(-113.49516,-62.32561)	0.0000
SolarN:Winter	2.642468	(-23.84270,29.12763)	0.8416
SolarY:Winter	74.717153	(41.06610,108.36821)	0.0001
SolarN:Summer	70.056139	(39.57363, 100.53865)	0.0000
SolarY:Summer	8.863714	(-22.76523,40.49266)	0.5753

Table 2: AR(1) Model Variance Component Estimates

	Estimate	95% CI
σ	28.08791	(23.07942,34.18329)
ϕ	0.1006577	(-0.2019074,0.3856839)

Inference

A Customer saves $1022.28 on average per year from switching to Solar, approximately $85.19 per month

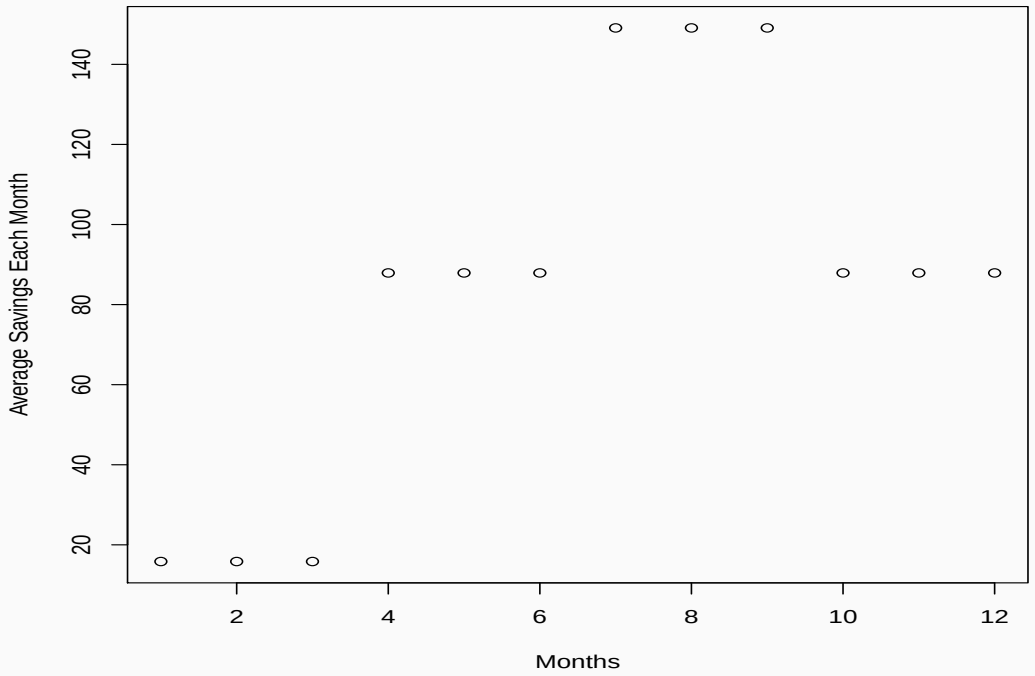


Figure 4: Average Savings Each Month of the Year from Switching to Solar

Prediction

It will take a customer on average just over 8 years (96.4 months) to recoup the $8,000, which a 95% prediction interval of 91 months to 104 months.

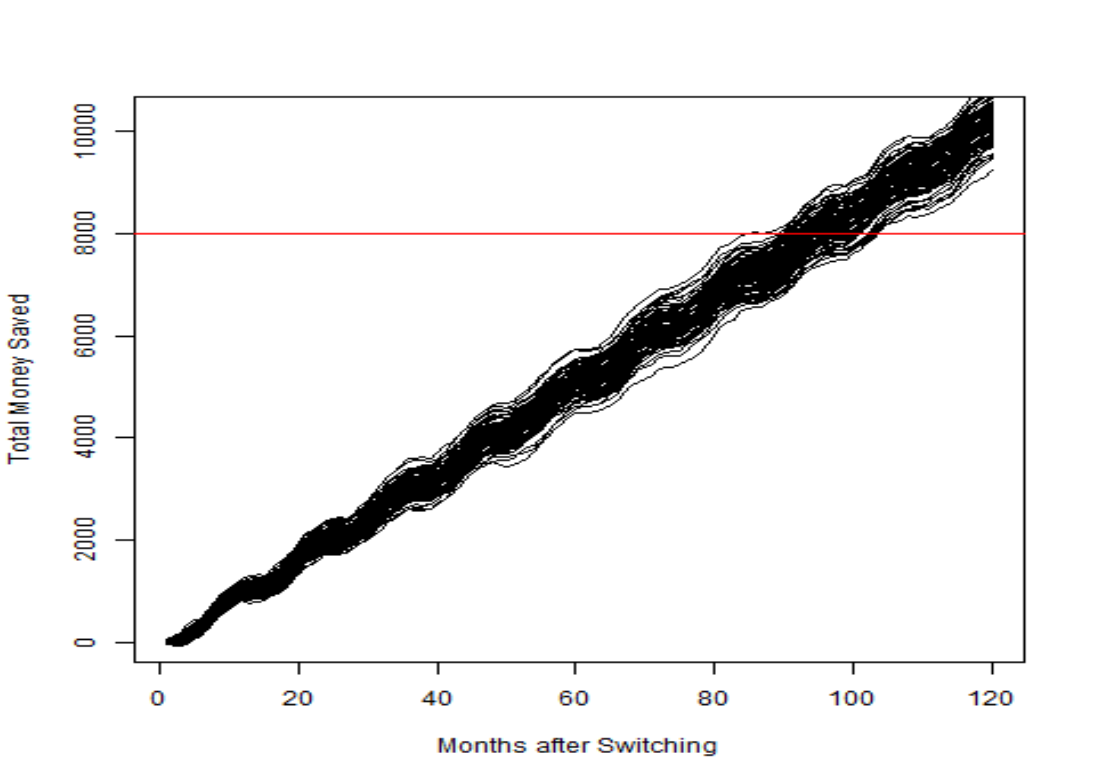


Figure 5: 100 Samples from Simulation Calculating amount saved with Time. Red Line at $8000 corresponds to installation costs of Solar Panels.

Conclusions

- We find how much on average a customer will save by switching to solar, and predict how long it will take to recoup the initial $8,000 cost of the solar panels
- These findings inform both the sales tactics of solar companies and the customer's budgeting for the initial cost of the panels and the subsequent monthly payments.

Shortcomings and Future Research

Shortcomings of the Model:

- Doesn't account for unequal time periods
- Model doesn't fully capture extreme data values

Future Research:

- This dataset had only the power bills for one customer living in Provo, UT. The model should be fit using additional customers to determine if these findings are generalizable