

Factors of First-Year Financial Savings in Solar Photovoltaic Systems: Data from Open Data Toronto and SolarOT Map*

My subtitle if needed

Yuanchen Miao

December 3, 2024

This paper examines the factors affect first-year financial savings for solar photovoltaic (PV) systems using data from Open Data Toronto and the SolarOT Map application. A multiple linear regression model identifies how variables such as payback period, roof size, system cost, electricity generation, and greenhouse gas (GHG) reduction influence initial bill savings. The analysis finds that payback period, roof size, and system cost are significant predictors of first-year savings, while electricity generation and GHG reduction are not. These findings clarify the financial and physical conditions that maximize early solar benefits that can help improve system design, inform policy, and accelerate the adoption of renewable energy technologies.

Table of contents

1	Introduction	2
2	Data	3
2.1	Overview	3
2.2	Outcome variables	4
2.3	Predictor variables	4
3	Model	7
3.1	Model set-up	8

*Code and data are available at: <https://github.com/NevaeH-9/SolarTO>.

4 Result	8
5 Discussion	9
5.1 First discussion point	9
5.2 Second discussion point	9
5.3 Third discussion point	9
5.4 Weaknesses and next steps	9
Appendix	10
A Additional data details	10
B Model details	10
B.1 Posterior predictive check	10
B.2 Diagnostics	10
References	11

1 Introduction

This study investigates the relationships between several key variables influencing the adoption of solar photovoltaic (PV) systems, utilizing data sourced from **Open Data Toronto** (City of Toronto 2024) and the SolarOT Map application. A multiple linear regression model was developed to analyze the impact of predictors such as payback period, roof size800k, system cost, annual electricity generation, and annual ghg reduction on the response variable, first year bill savings. The research aims to identify the most significant factors affecting early financial benefits from solar PV installations, providing some ideas to people on deciding whether they choose to have a solar PV system.

The primary estimand of this study is the average effect of predictors payback period, roof size, system cost, annual electricity generation, and annual ghg reduction on the expected first year bill savings. The model seeks to quantify how changes in these predictors influence the immediate financial benefits homeowners can expect during the first year of solar PV operation, adjusting for the unique characteristics of individual installations.

The multiple linear regression model reveals that payback period, roof size, and system cost are statistically significant predictors of first year bill savings. In contrast, annual electricity generation and annual ghg reduction do not show a significant relationship with the response variable. This indicates that, while system cost and roof area is very important in determining initial financial benefits, the overall electricity generation capacity and greenhouse gas reductions may have less direct influence on first year bill savings.

Understanding the factors that affects the early financial benefits is important for people who decides to have a solar PV system, as many potential users prioritize immediate cost savings

when making investment decisions. By highlighting the significance of variables like payback period, roof size800k, and system cost, this study provides idea for policymakers, manufacturers, and installers aiming to design cost-effective solar PV systems. Furthermore, the findings underscore the importance of tailoring incentives and educational efforts to emphasize the factors that most directly impact early financial returns, potentially accelerating the transition to sustainable energy solutions.

The rest of paper are structured as follows: Section 2 introduced the programming language that is used in this paper and packages are used to generate models, graphs and tables, and tidy the paper. The predictors are also introduced in the section. Section 3 is introducing the multiple linear regression model, the coefficients of each predictor. The result of the model is discussed in Section 4.

2 Data

2.1 Overview

The statistical programming language **R** (R Core Team 2023) is used as a base of this paper. Packages **lubridate** (Grolemund and Wickham 2011), **dplyr** (Wickham et al. 2023), **tidyverse** (Wickham et al. 2019) and **opendatatoronto** (Gelfand 2022) are used to download, clean the data and generate the multiple linear regression model. Package **ggplot2** (Wickham 2016) has been used to make figures and labeling the graph. Packages **kableExtra** (Zhu 2024) and **broom** (Bolker and Robinson 2024) have been used to generate tables for the paper. Package **modelsummary** (Arel-Bundock 2022) has been used to summary the model. The data was downloaded from Open Data Toronto that refreshed daily and collected using SolarTO Map. With the cleaning process, observations with NA values and address that is not found by the software are removed from the data set. Multiple linear regression model are being used to find the relationship between first year bill savings with other factors. ## Measurement

The data set is downloaded from **Open Data Toronto** (City of Toronto 2024), a website that post data collected in Toronto City can be used and re-published as everyone wish freely that the website is aiming to make data innovated by anyone to draw insights and use evidence to inform the design of civic services. The raw data is collected by using SolarTO Map, a Geographic Information Systems analysis of Light Detection and Ranging data. The software takes into account geographical latitude and the sun’s daily position throughout the year. By using SolarTO Map, people are able to find out that the data of the roof area they could put a device and how much solar energy it could generate and how much money they are saving by using the solar energy device by selecting the area they are looking for. First several rows of cleaned data is in Table 1

Table 1: First Several Rows of Cleaned SolarTO Map Data provided by Open Data Toronto

structureid	annual_electricity_generation_kwh	annual_electricity_bill_savings_dollars	system_size_kw	payback_period_years	annual_ghg_reduction_kg	annual_ghg_reduction_tons	system_cost_dollars
38 Westleigh Cres Structure 1	5627	881	4.893044	10	394	44	14700
36 Westleigh Cres Structure 1	6012	941	5.227826	9	421	51	14100
34 Westleigh Cres Structure 1	4061	636	3.531304	10	284	32	10600
34 Westleigh Cres Structure 2	1480	232	1.286957	10	104	12	3900

2.2 Outcome variables

The outcome variable in this paper is first year bill savings that it represents the estimated electricity bill savings over the first year of operation of the solar PV system in dollars. This is calculated by multiplying the average utility rate with the amount of electricity produced by solar panel but not consumed by users. First year bill savings is also considered as response variable in the multiple linear regression model.

2.3 Predictor variables

Annual Electricity Generation(annual_electricity_generation_k): Estimated annual electricity production of the proposed rooftop solar system measured in kWh. To estimate Annual Electricity Generation the following assumptions are used: 15% panel efficiency and 86% performance ratio.

System Size(system_size): The size of the solar photovoltaic (PV) system that can fit on the rooftop, measured in kilowatts (kW). It is assumed that 1 kW of solar will generate 1,150 kWh/year, based on average solar radiation in Toronto. The System Size is calculated by dividing the Annual Electricity Generation (kwh) by 1,150 kWh.

Payback Period(payback_period): The payback period refers to the number of years it will take to recover the cost of the solar PV system through the savings generated by the solar PV system. The payback period is based on annual electricity bill savings minus the upfront cost of the solar system.

Annual GHG Reduction(annual_ghg_reduction_kg): Greenhouse gas reduction is based on the reduced consumption from the grid and it's associated GHGs each year, measured in kilograms of carbon dioxide equivalent (kg of CO₂e). It is assumed that solar will offset 0.07 kg (000.07 tonnes) of CO₂e (carbon dioxide equivalent) for each kWh of solar electricity produced.

***Roof Size(roof_size800k)**:** Total roof area that receives at least 800 kWh of (kilowatt hours) of solar radiation per square metre. Based on industry best practices, the SolarTO Map uses certain criteria to identify a suitable rooftop, one of which is that the rooftop receive at least receives at least 800 kWh of solar radiation per square meter.

System Cost(system_cost): The upfront cost of the solar PV system in dollars. The System Cost is based on the System Size and the average solar installation rates in Toronto. A estimated \$/Watt rate ranging from \$3/watt to \$2,2/watt (dependent on system size) is applied to calculate cost.

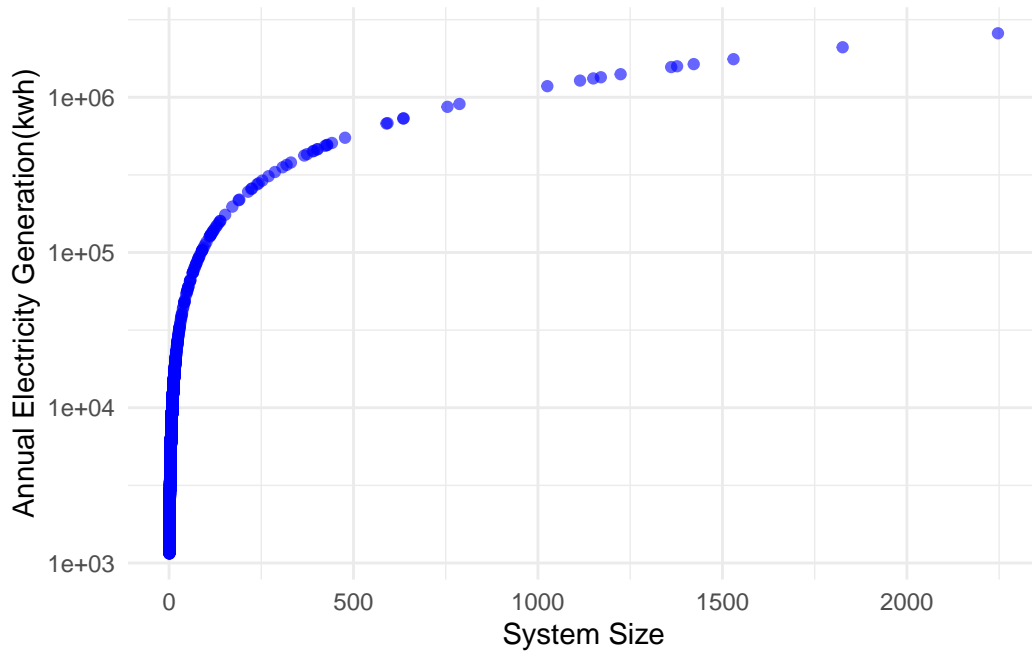


Figure 1: Distribution of System Size and Annual Electricity Generation

The roof size refers to the total area capable of receiving at least 800 kWh of solar radiation per square meter. To maximize the efficiency of the solar photovoltaic (PV) system, it is assumed that the entire roof area will be utilized for system installation. It is further assumed that as the roof size increases, the system size will also increase proportionally. As demonstrated in Figure 1 and Figure 2, there is a clear positive relationship between system size, annual electricity generation, and annual greenhouse gas (GHG) reduction. Specifically, larger PV systems generate more electricity than smaller systems, leading to a reduction in electricity

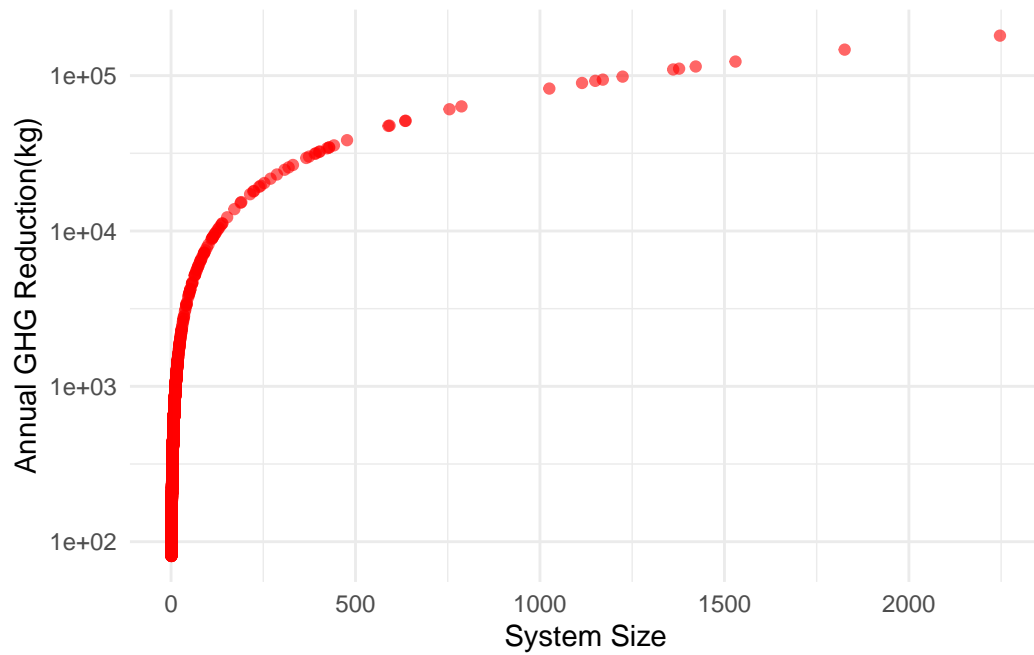


Figure 2: Distribution of System Size and Annual GHG reduction

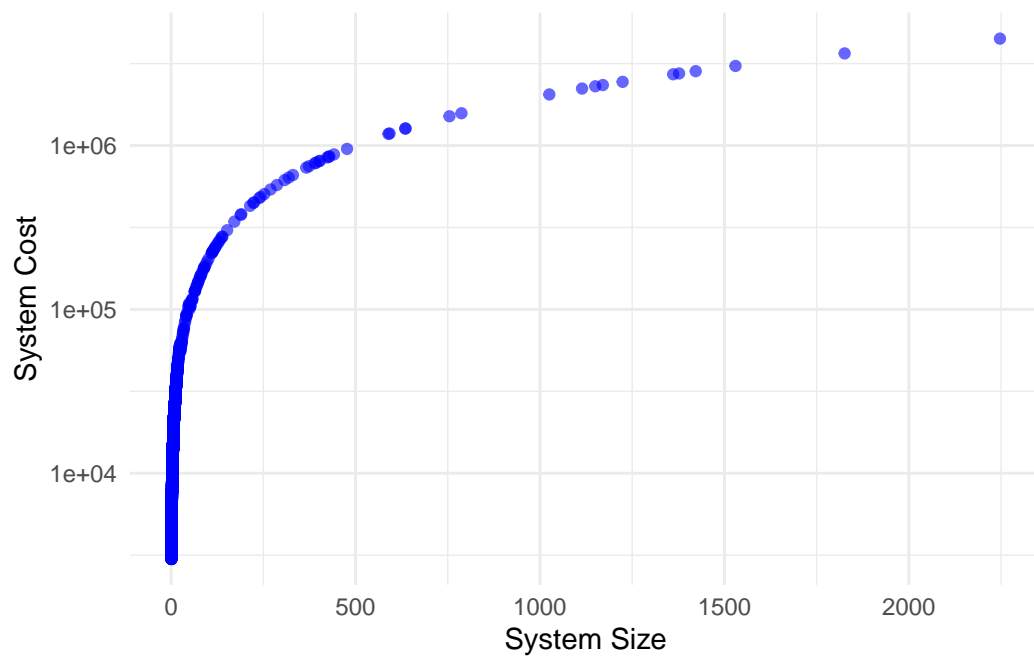


Figure 3: Distribution of System Size and System Cost

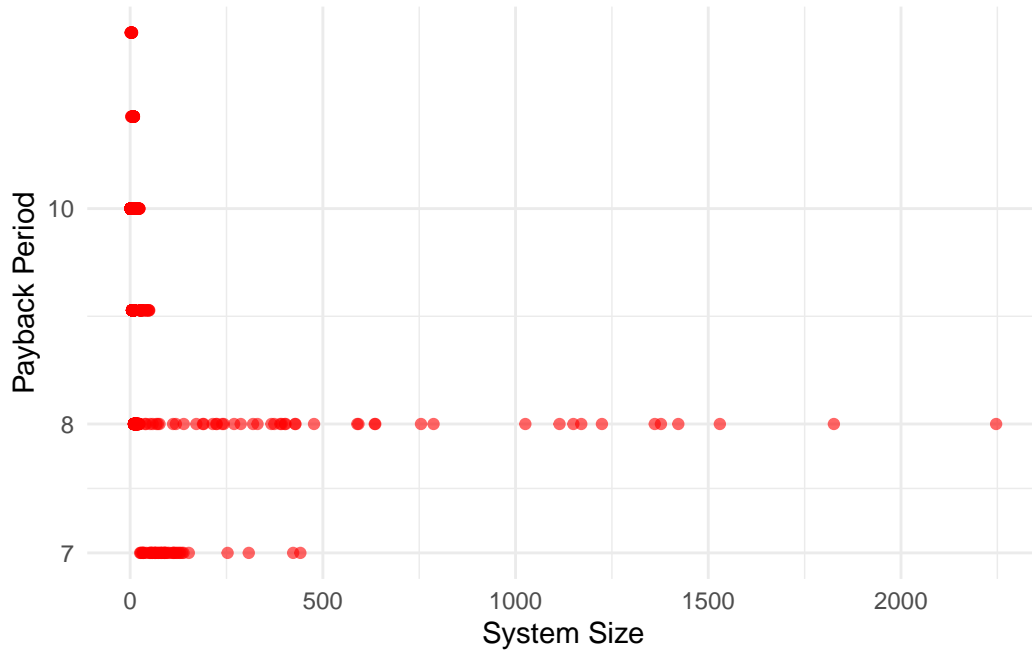


Figure 4: Distribution of System Size and Payback Period

consumption from alternative sources. Consequently, this results in a greater annual GHG reduction for larger systems.

In Figure 3 and Figure 4, it is observed that the payback period is not directly correlated with system size. While the cost of larger systems is higher, these systems also generate more electricity, leading to greater savings on electricity bills. However, the number of years required to recover the initial investment through bill savings remains relatively consistent, regardless of system size.

3 Model

The multiple linear model is generated to investigate the relationships between annual billing savings and predictor system cost, roof size and payback period. These predictors are selected for significance in influencing the change of the annual billing savings and the other predictors are reduced by insignificance.

Here we briefly describe the multiple linear regression model used to investigate the relationship. Background details and diagnostics are included in Appendix B.

Table 2: Model Summary

```
# A tibble: 6 x 5
```

	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	(Intercept)	3913.	108.	36.3	3.66e-252
2	payback_period	-367.	10.4	-35.4	9.42e-241
3	roof_size800k	3.07	0.495	6.19	6.41e- 10
4	system_cost	-0.0545	0.00420	-13.0	8.48e- 38
5	annual_electricity_generation_k	1.97	1.59	1.24	2.15e- 1
6	annual_ghg_reduction_kg	-25.4	22.7	-1.12	2.64e- 1

3.1 Model set-up

Annual Billing Savings = $\beta_0 + \beta_1 \cdot \text{system size} + \beta_2 \cdot \text{payback period} + \beta_3 \cdot \text{system cost} + \beta_4 \cdot \text{annual electricity generation}$

- β_0 is the intercept of the multiple linear regression model, which represents the value of annual billing savings when other predictors are zero
- β_1 is the coefficient of **system size** that represents the effect of adding system size on annual billing savings
- β_2 is the coefficient of **payback period** that represents the effect of adding a year on payback period on annual billing savings
- β_3 is the coefficient of **system cost** that represents how much of adding a a dollar has effect on system cost on annual billing savings
- β_4 is the coefficient of **annual electricity generation** that represents the change of annual billing savings when one kWh electricity is generated by the solar PV system
- β_5 is the coefficient of **annual ghg reduction** that represents the change of annual billing savings when one kg carbon dioxide is reduced by using the solar PV system

4 Result

As shown in Table 2, payback period, roof size and system cost are very significant to the annual billing savings, as they are having the p-value equal to zero and annual electricity generation and annual ghg reduction are having p-value larger than 0.5 that these two predictors should be removed from the multiple linear regression model since they are not having a lot of influence on the money saved when the solar PV system is placed on the roof. Payback period has large negative estimate -315.06846232, it is representing that with a year of payback period added, the annual billing savings is significantly decreased, the roof size is having the only positive estimate in significant

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

B.2 Diagnostics

References

- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- Bolker, Ben, and David Robinson. 2024. *Broom.mixed: Tidying Methods for Mixed Models*. <https://CRAN.R-project.org/package=broom.mixed>.
- City of Toronto. 2024. “SolarTO Map Dataset.” <https://open.toronto.ca/dataset/solarto/>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with ‘Kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.