Data mining for evaluating the rebounds-associated emissions due to energy-related consumer behavioural shifts in Switzerland

Rhythima Shinde*, Sidi Peng¹, Saloni Vijay¹, Stefanie Hellweg¹, Andreas Froemelt¹

¹ Institute of Environmental Engineering, D-BAUG, ETH Zurich
* Ecological Systems Design, John-von-Neumann-Weg 9, 8093 Zürich, Switzerland
* Corresponding author e-mail: shinde@ifu.baug.ethz.ch

Abstract. Energy-related household consumption lead to a substantial share of the total household GHG emissions (direct and indirect). The policies or technologies, which try to mitigate these emissions often end up ‘rebounding’ i.e. the savings of energy (bills) caused by these measures, induce further expenses in other (e.g. travel) or same (e.g. electricity) categories, leading to (partial) offsetting of the emissions saving. This research introduces application of a data-driven bottom up method to evaluate these rebound emissions based on the household consumption (expenses) and properties. Two scenarios of energy-savings measures are evaluated here: (1) switching to energy-efficient devices, and (2) switching to renewable energy. The results are discussed for households with varying income, region of residence and household size. The results show that higher income and bigger households have higher total rebounds for both scenarios, while Zurich has lowest compared to all other Swiss regions.

1. Introduction
Households heavily contribute around 65% of the total greenhouse emissions i.e. 22 Gigatonnes carbon dioxide equivalent (CO₂-eq) including direct impacts and embodied in household purchases [1], and these emissions continue to grow with the rate of around 1.5% per year [2]. The residential energy sector contributes to about 30% of energy demand and 20% of global greenhouse gas emissions [3]. Swiss households, on an average, contribute with emissions of 9-12 tonnes of CO₂-eq per year per capita, and their residential energy itself contributes to about 26% of total household-induced carbon footprint [4], [5]. Multiple policies and technologies have been introduced to incentivize building owners and households to reduce these emissions, e.g. subsidies for renewable energy, and improved insulations [6], [7]. However, not just are these policies difficult for households to adapt to [8], but these policies often underestimate the effects of re-spending saved costs induced by these measures. This means that cost savings in one area might induce additional consumption in the same or other consumption categories, ultimately leading to a so-called ‘rebound effect’ (relative additional/ reduced emissions than expected). This re-spending of saved costs induces again environmental impacts that decreases the overall reduction potential of the initial measure. For example, cost savings from more efficient insulation and lighting may be spent on overseas holiday [9], [10].

Rebound effects due to the energy efficient devices have been formally studied using econometric methods, survey data-analysis and combining the household expenditures with the Input-Output Analysis [11]–[13]. Though these methods have rightly evaluated the relations among
income/prices and expenses, these are top-down methods, which overlook the other household characteristics e.g. the household size and demographics, in determining their final rebounds. Studies have not just clearly shown the relation of the household properties to the household consumptions and emissions [14], [15], but also shown the relation to the rebound-related emissions[9], [16]. Thus, this study introduces application of a bottom-up data-mining based novel method to understand these energy-savings related rebounds and emissions, where each household’s socio-demographic property is incorporated.

2. Method

The study used the Household Budget Survey (HBS) data from year 2006 to 2017 provided by Swiss Federal Statistical Office (FSO) [18], which is released every three years. The data consists of 32’255 household (after outlier removal; originally there are 38’975 households) and 21 variables (only relevant ones from originally - 600+ variables selected) including household properties and expenses. Here, the dataset was split into a training set (80%) and testing set (20%). The target dataset (the data on which prediction is done) is a subset of the HBS – 50% of randomly selected households.

A random forest regression method was used to train the dependency of the expenses (Y) on the household properties (X), as shown in Table 1. In training, the key hyperparameters of the random forest, i.e., number of trees, minimum number of samples required for further splitting and maximum number of features considered, were tuned manually by checking the sensitivity of R-squared score of ten-fold cross validation to the parameters. The threshold values after which the R-squared score (which gives indication of how well the model predicts the real values) does not increase significantly were chosen. Independent variables (X) includes households' economic and socio-demographic properties (e.g., disposable income the number of females or males, etc.), private properties (e.g., used cars, dishwashers, etc.), regions, years, and months when the consumption happened, while the dependent variables (Y) includes expenses in the main consumption categories and saving (e.g., food and non-alcoholic beverages, air travel, etc.). All monetary values were inflation adjusted based on price level in 2015 [19]. Since the correlation matrix of expenses data revealed that the dependent variables are weakly or uncorrelated with each other, the study used the “MultiOutputRegressor” which trains an individual sub model for each category in Y.

| Independent parameters (X)                                      | Dependent parameters (Y)                             |
|----------------------------------------------------------------|-----------------------------------------------------|
| Households with a woman as a reference person                  | Food and non-alcoholic beverages                     |
| Major region: lake geneva, zh, central ch, northwestern ch, ticino, eastern ch, espace mittelland | Alcoholic beverages and tobacco                      |
| # Foreign or swiss persons in household                         | Restaurants and hotels                               |
| # Divorced, married, or unmarried persons in household          | Clothing and footwear                                |
| # Widowed persons in household                                 | Miscellaneous goods and services                     |
| # Students/trainees/apprentices in household                    | Housing, water, electricity, gas, and other fuels    |
| # Employed or self-employed persons in the household           | Furnishings, household equipment and routine household maintenance |
| # Pensioners in the household                                   | Communication                                       |
| # Other persons in household (wrt employment)                  | Education                                           |
| # Members in each age group                                    | Health                                              |
| # Persons per household                                        | Purchase & operation of vehicles                     |
| # Year and month                                                | Transportation services                              |
| # Electrical housing appliances of the household                | Passenger transport by air                           |
| # Vehicles of the households                                   | Recreation and culture                               |
| Infrequent income                                               | Package holidays                                    |
| Disposable income                                              | Saved amount (computed)                             |
Note that, induced expenses (CHF) show the absolute values, while the rebound expenses (%) depict the ratio of induced expenses and the expected monetary energy savings due to the following energy saving scenarios. Two scenarios were developed based on the literature on energy savings – (1) Electricity expenses will decrease by 17.5% by switching to energy-efficient devices [17]; (2) Electricity expense will drop by 67.48% by switching to renewable energy sources (e.g., solar energy) [18]. In scenario 2, only rebound effects caused by electricity expenses savings that takes place after the pay-back period were considered. To calculate the final rebound emissions, (1) the expenses \((X_1)\) were predicted directly using the Random Forest Regressor and the independent variables. (2) Then, after adjusting the income of the households with the new savings (17.50% or 67.48% of the electricity expenses for respective scenarios), the expenses \((X_2)\) were predicted again, and finally (3) these two expenses were subtracted from each other to obtain the induced expenses \((X_2-X_1)\), as the effects of the saved income. For model performance evaluation, besides R-squared value obtained from model testing, loss percent, as in Eq.1, \((L)\) helps compare the predicted total expenses \((T)\) with the actual savings \((S)\). Also, the total predicted expenditure was compared with the income to validate the model.

Based on feature importance scores of independent variables, the result was analysed as per income group. In addition, for policy recommendations to the building stakeholders, further analysis was performed on regions, and household size.

To calculate rebound emissions \((E)\), we used the emission factors based on a consumption LCA that was performed by Froemelt et.al [15]. Rebound emission \((E)\) (and expense, \(R\)) are the relative emissions (and expenses) compared to the expected emissions, \(R_e\), caused by the expected saving \((S)\), calculated as in Eq. 2 & Eq. 3, respectively. In other words, Rebounds are the ratio of absolute induced expense/emissions to the expected savings in the tested scenario.

\[
L = \frac{(S - T)}{S} \quad \text{Eq.1}
\]
\[
E = \frac{(X_2 - X_1) \times GI}{R_e} \quad \text{Eq.2}
\]
\[
R = \frac{(X_2 - X_1)}{S} \quad \text{Eq.3}
\]

3. Results and Discussion

The average R-squared score of selected consumption category is 0.38. For the electricity sector, it is 0.26 and 0.25 for Scenario 1 and 2, respectively, which is because electricity expenses after saving happens were used to train and test model while savings are different in two scenarios. For other categories, R-squared scores are same for two scenarios and they lie between 0.29-0.58. As mentioned in [20], values beyond 0.25 are acceptable R-squared values for such a prediction model. The slopes between the curves of true total expenses vs disposable income is 0.97, while the one between the predicted total expenses vs disposable income is 0.87, which is comparatively close to the true slope. The loss percent of electricity saving for Switzerland is 17% and 28% for the two scenarios, i.e. the total predicted expenses are 0.17 and 0.28 times less than the actual savings. All these checks confirm that the model results are acceptable.

The total rebound emissions in case of Scenario 1 and 2 are 12% and 11%, which means that these energy efficiency measure only provides 88% or 89% of the initially thought reduction potential. The total induced emission is 10 and 34 kg CO₂-eq/year/household in Scenario 1 and 2, respectively. A general trend can be seen from Figure 1 that higher income groups have higher induced emissions for both scenarios. Households with a monthly income of 2000-4000 CHF spend most on rent and utility. Other income groups spend most on restaurants & hotels.

As the household income increases, induced expenses increase in all categories, except food. Also, the percentage of total induced expense on food is lower for higher income group. For electricity, most income groups tend to show negative induced expenses. Additionally, induced emissions increase with income, and major contributing categories to higher induced emissions include restaurants & hotels, transport, and recreation & culture.

An increase in household size (with an increase in average disposable income of household) leads to increased induced expenses and emissions for both the scenarios. Considering the share of total expenses for individual categories, larger households re-spend more on transport and less on
rent and utility than smaller households. Figure 2 shows that the induced expenses of households living in Zurich is the lowest in both the scenarios. However, the average disposable income of Zurich is the highest. The highest induced expenses can be seen for Lake Geneva in Scenario 1 and Ticino in Scenario 2. Zurich re-spend the least on food. Restaurants & hotels and rent & utility are among the categories having highest share of total induced expenses in all the regions. Like the induced expenses, induced emission is the lowest for Zurich. However, the induced emissions are the highest for Central Switzerland (for travel and restaurant/hotels) in both the cases.

Table 2: Rebound expenses and emissions for the three cases corresponding to Scenario 1 and 2

| Case 1: Income level [CHF/month/household] | Case 2: Household size [persons/household] |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Rebound expenses [%] | Scenario 1 | -3% | 34% | 32% | 37% | 20% | 26% | 34% | 33% |
| Rebound expenses [%] | Scenario 2 | 3% | 30% | 28% | 30% | 20% | 23% | 28% | 30% |
| Rebound emissions [%] | Scenario 1 | -4% | 15% | 14% | 17% | 7% | 11% | 15% | 16% |
| Rebound emissions [%] | Scenario 2 | 0% | 14% | 13% | 15% | 8% | 0% | 13% | 15% |

Case 3: Regions

| Lake Geneva | Espace Mittelland | Northwestern Switzerland | Zurich | Eastern Switzerland | Centra Switzerland | Ticino | Switzerland |
|-----|-----|-----|-----|-----|-----|-----|-----|
| Rebound expenses [%] | Scenario 1 | 32% | 26% | 28% | 26% | 27% | 32% | 29% |
| Rebound emissions [%] | Scenario 1 | 13% | 12% | 13% | 12% | 11% | 15% | 12% |
| Rebound expenses [%] | Scenario 2 | 11% | 11% | 11% | 11% | 11% | 13% | 12% |
| Rebound emissions [%] | Scenario 2 | 11% | 11% | 11% | 11% | 11% | 11% | 11% |

Figure 1: Induced expenses (left) and induced emissions (right) for different income groups

Figure 2: Induced expenses (left) and induced emissions (right) for different regions
A noticeable general trend in this study is that the higher income groups have higher induced expenses and emissions, as well as rebound expenses and emissions. This trend for absolute rebounds seems plausible. Since 93% of households spend 1%-5% of disposable income on electricity, higher income usually indicates higher electricity expenses and more electricity savings, which leads to higher induced expenses and savings. Another general trend is that households with higher income spend less on food items. Households with larger household size and higher income may still re-spend more on food due to more people.

As shown in Case 1 in table 2, under similar re-spending structure, the higher income group with higher induced expenses causes higher induced emissions. Central Switzerland was estimated to have lower induced expenses but higher induced emissions than Ticino, due to households in the former region having higher income than the latter region. Zurich with smaller households also tend to show lower induced expenses.

There are certain limitations of the method, leading to high uncertainties, wide standard deviations, and high loss percent. One of the reasons is the high variation in the HBS dataset. Thus, it is best to use the model results to depict the income groups and not as recommendations for individual households. Additionally, the model does not allow putting a constraint over the total predicted expenses, which leads to either over- or under-estimation of the total expenses, as compared to the total expected savings (loss percent measures this). To correct this loss percent, a regression model with better constraint-based approach could be used [22].

Studies looking at the rebound emissions associated with replacing all existing lighting with LEDs or replacing existing incandescent bulbs with compact fluorescent bulbs (CFLs), quantified rebound emissions for an average UK household at about 15%, which is similar with the estimated rebound of an average Swiss household in the study [9]. Other studies which look into all type of energy efficiency improvements e.g. in industry, replacing the heating systems, etc. have estimated the rebounds to be below 30% for industrialised countries [23]. This shows that our results in the range of 15-30% are valid and comparable to other results.

4. Conclusion
Rebounds are highly discussed in top-down econometric studies when it comes to studies about the actual environmental benefits of the energy savings. This study proposes application of a novel bottom-up data-driven method to understand the rebounds for energy efficiency and renewable energy measures. Applying it to the two scenarios of energy savings, efficient devices and renewable energy transition, shows that there is a significant rebound for the higher income groups, especially in recreation & culture and restaurant & hotels. Household size also affects these rebounds – with bigger households having a higher rebound. Finally, Zurich tends to show lower rebound emissions, while Central Switzerland shows high rebound emissions. The method used in this study has certain uncertainties due to its nature of capturing the wide household range and their properties, but it still gives an overview into the distinction between different households and their distinct rebound tendencies.

In this study, the results of rebound effects are discussed as per income groups, regions, and size of the households. This helps develop various policies for the major regions of Switzerland and the building stakeholders. Also, the income is chosen as the main determinant for the discussion of the result due to its high feature importance obtained in the current study i.e. income is one of the most important independent parameters defining the final predictions of the rebound expenses. It is important to keep in mind that an accurate uncertainty estimate or analysis is needed to address the high variability of the dataset here. Finally, it should be noted that the scope of the method developed here could be implemented beyond the savings in the energy category, to other categories in housing e.g. evaluating effects of savings in rent.

Acknowledgments
This research is carried out with the grants from the NRP73 – Sustainable Economy from SNSF Swiss National Science Foundation).
References

[1] D. Ivanova et al., “Environmental Impact Assessment of Household Consumption,” J. Ind. Ecol., vol. 20, no. 3, pp. 526–536, 2016.

[2] Y. Liu, “Random forest algorithm in big data environment,” Comput. Model. New Technol., vol. 18, no. 12A, pp. 147–151, 2014.

[3] P. Nejat, F. Jomehzadeh, M. M. Taheri, M. Gohari, and M. Z. Muhd, “A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries),” Renew. Sustain. Energy Rev., vol. 43, pp. 843–862, 2015.

[4] S. M. L. M. Jungbluth N., Nathani C., “Environmental impacts of Swiss consumption and production,” Foen, vol. 54, no. 2, p. 173, 2011.

[5] A. Froemelt, R. Buffat, and S. Hellweg, “Machine learning based modeling of households: A regionalized bottom-up approach to investigate consumption-induced environmental impacts,” J. Ind. Ecol., vol. 24, no. 3, pp. 639–652, 2020.

[6] M. Aune, “Energy comes home,” Energy Policy, vol. 35, no. 11, pp. 5457–5465, 2007.

[7] D. Štreimikiene, “Residential energy consumption trends, main drivers and policies in Lithuania,” Renew. Sustain. Energy Rev., vol. 35, pp. 285–293, 2014.

[8] P. Du, L. Q. Zheng, B. C. Xie, and A. Mahalingam, “Barriers to the adoption of energy-saving technologies in the building sector: A survey study of Jing-jin-tang, China,” Energy Policy, vol. 75, no. 2014, pp. 206–216, 2014.

[9] M. Chitnis, S. Sorrell, A. Druckman, S. K. Firth, and T. Jackson, “Who rebounds most? Estimating direct and indirect rebound effects for different UK socioeconomic groups,” Ecol. Econ., vol. 106, pp. 12–32, 2014.

[10] J. D. Khazzoom, “Economic implications of mandated efficiency in standards for household appliances,” Energy J., vol. 1, no. 4, pp. 21–40, 1980.

[11] J. D. Khazzoom, An econometric model integrating conservation measures in the residential demand for electricity. United States: JAI Press, 1986.

[12] M. Chitnis, S. Sorrell, A. Druckman, S. K. Firth, and T. Jackson, “Turning lights into flights: Estimating direct and indirect rebound effects for UK households,” Energy Policy, vol. 55, pp. 234–250, 2013.

[13] K. Takase, Y. Kondo, and A. Washizu, “An analysis of sustainable consumption by the waste input-output model,” J. Ind. Ecol., vol. 9, no. 1–2, pp. 201–219, 2005.

[14] A. Froemelt, D. J. Dürenmatt, and S. Hellweg, “Using Data Mining to Assess Environmental Impacts of Household Consumption Behaviors,” Environ. Sci. Technol., vol. 52, no. 15, pp. 8467–8478, 2018.

[15] A. Froemelt and T. Wiedmann, “A two-stage clustering approach to investigate lifestyle carbon footprints in two Australian cities,” Environ. Res. Lett., vol. 15, no. 10, 2020.

[16] E. Aydin, N. Kok, and D. Brounen, “Energy efficiency and household behavior: the rebound effect in the residential sector,” RAND J. Econ., vol. 48, no. 3, pp. 749–782, 2017.

[17] N. Boogen, “Estimating the potential for electricity savings in households,” Energy Econ., vol. 63, pp. 288–300, 2018.

[18] European commission, “Best practices on Renewable Energy Self-consumption.”, 2015. [Online]. Available: https://ec.europa.eu/energy/sites/ener/files/documents/1_EN_autre_document_travail_service_part1_v6.pdf.

[19] A. Druckman and T. Jackson, “Understanding Households as Drivers of Carbon Emissions,” in Taking Stock of Industrial Ecology, R. Clift and A. Druckman, Eds. Cham: Springer International Publishing, 2016, pp. 181–203.

[20] R. Nau, “What’s a good value for r-squared?,” 2020. [Online]. Available: https://people.duke.edu/~rnau/rsqrd.htm.

[21] M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” 32nd Int. Conf. Mach. Learn. ICML 2015, vol. 1, pp. 97–105, 2015.

[22] P. Rana, C. Berry, P. Ghosh, and S. S. Fong, “Recent advances on constraint-based models by integrating machine learning,” Curr. Opin. Biotechnol., vol. 64, pp. 85–91, 2020.

[23] J. Freire-González, “Methods to empirically estimate direct and indirect rebound effect of energy-saving technological changes in households,” Ecol. Model., vol. 223, no. 1, pp. 32–40, 2011.