Data mining for evaluating impacts of rebounds in the housing sector of Switzerland

LCA Discussion Forum 74

Rhythima Shinde, Andreas Froemelt, Stefanie Hellweg
Motivation

- Shrinking Housing Environmental Footprint:
  - Reducing the emissions of households of Switzerland due to housing market, household behaviors and the material footprint

- Project Partners: 2 cooperatives and 1 insurance firm (>10,000 apartments)
  - Cooperatives tries to provide affordable and environmentally sustainable housing, but savings in rent may lead to increase in other consumptions -> induced consumptions
  - Similarly, savings in housing operational expenses (heating costs) due to energy remediation may lead to a rebound effect,
Motivation: Rebounds

Disposable Income

Savings

Consumptions

CO₂

CO₂ + CO₂

Final Environmental Footprint?

Consumptions

CO₂

Induced consumptions

Rebounds

Savings
Aim

- Quantify the environmental impact due to the savings in housing expenses
  - How can rebound expenses be calculated due to the savings?
  - What are the associated environmental impact due to the savings?
Terminologies

- Disposable income
  \[= \text{income} - \text{rent} - \text{compulsory fees} \quad (\text{taxes} + \text{basic healthcare costs})\]

- Induced consumptions / Rebound
  \[= \text{difference in consumption with change in disposable income}\]
Model: Methodology (‘Training database’) 

Household budget survey

- Independent: Household properties (e.g. age, region distribution) and disposable income
- Dependent: expenses for 41 aggregated (350) consumption categories
- Supervised Machine Learning Approach

Model: Methodology (Training)

Supervised Machine Learning Approach:

- Learning how dependent parameters are determined by the independent ones (Training)
- Choice of ‘best’ model by comparing root mean square errors: MO Random Forest\textsuperscript{1,2}
  - Advantage: Allows for higher dimensionality, handles missing values
  - Randomly choses decision trees based on given input features/ independent variables

Model: Methodology (‘Prediction’)

Based on the supervised learning/ training:

- Prediction of consumption profiles of households in Zurich cooperative (ABZ₁, 2009-2011)

![Diagram of the model showing dependent and independent variables, training set, and prediction process.]
## Data HBS vs ABZ

<table>
<thead>
<tr>
<th></th>
<th>HBS (Swiss average)</th>
<th>ABZ (Zurich cooperative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median disposable income</td>
<td>4136 CHF</td>
<td>2900 CHF</td>
</tr>
<tr>
<td>Avg. occupancy per household</td>
<td>2.38</td>
<td>1.78</td>
</tr>
<tr>
<td>Avg. employed people per household</td>
<td>1.15</td>
<td>0.33</td>
</tr>
<tr>
<td>Percentage of retired/pensioners</td>
<td>35%</td>
<td>59%</td>
</tr>
<tr>
<td>Percentage of students</td>
<td>9.1%</td>
<td>1.9%</td>
</tr>
<tr>
<td>Percentage of international (non-Swiss)</td>
<td>13%</td>
<td>30.5%</td>
</tr>
</tbody>
</table>

Data Adaptations:
1. ABZ: Income as per professions and minimum income category provided
2. Outliers cleaned for very high or very low savings (this possibly represents large (+/-) savings during a given time of year, e.g. debt/investments. Also these outliers are peculiar groups which are not representative of ABZ tenants, and thus removed.)
Data and preprocessing

- Household Budget Survey
  - Includes all the household properties and the consumption expenses and quantities

- Monthly clusters for de-seasonalising
  - Consumption expenses clustered together based on seasons (Statistical tests)

- Preparing the stakeholder data
  - Matching the household properties as per the Household budget survey
Prediction of consumption expenses

R2 -> coefficient of determination -> 0.52 to 0.97
Method: Calculation of rebounds

- Random Forest (Prediction of consumptions)

Results on induced consumptions

Share of CHF spent (on goods and services) after savings of 500 CHF on rent

- 2000-3999
- 4000-5999
- 6000-7999
- 8000-10000
- >10000

Categories:
- travel
- services
- housing
- food
Results on induced consumptions: Food
Results on induced consumptions: Travel

- Travel
- Services
- Housing
- Food

Package holidays:
- Air
- Railway

Travel services:
- Housing
- Food

Shinde, Rhythima | 30.06.2020 | 15
Consumption LCA

- Following the study of Froemelt et al. 2018, every consumption category was approximated as process model
- The life cycle inventory data were extracted from three databases: ecoinvent v3.6, Agribalyse v1.3, and EXIOBASE v3.4
- For food and lubricants, quantities were used instead of expenses to convert to the relevant associated impact

Adaptations
- As all the consumption categories could not be predicted (reducing accuracy of the model with more outputs), aggregated consumption categories were reduced down to sub-categories as average % expense share of the household income group
- Upgraded environmental databases
Consumption LCA

- “On” determines if the unit process is active (whether it shall be included or not)
- “Activity” holds the key to find the activity in the respective database via brightway2. This also shows if the unit process originates from ecoinvent, Agribalyse or EXIOBASE;
- “DB Act” shows a human readable name of the unit process;
- “CFL Act” indicates a conversion factor for individual unit processes.
- “ConversionDem2FU” in order to convert the functional unit of the process model into the units of the demand.

Froemelt et al. (2018)
Consumption LCA

A process for every consumption category

Froemelt et al. (2018)
Results on environmental footprint of rebounds
Results on environmental footprint: Food
Results on environmental footprint: Travel

Disposal income

GHG (kg CO2-eq)

- Average of travel
- Average of services
- Average of housing
- Average of food

GHG (kg CO2-eq) vs Disposable income

- Railway
- Road
- Air
- Vehicles

Average of travel

Average of services

Average of housing

Average of food

0 5 10 15 20 25
0-3999 4000-5999 6000-7999 8000-10000 >10000

Disposal income

GHG (kg CO2-eq)
Results on consumption rebounds

- Lower income group (<4000 CHF) have
  - High housing direct rebounds: energy and appliances
  - Food rebounds (dairy and meat products)

- Middle to slightly high income group (4000-8000) have
  - Traveling/recreation (services like hotels) rebounds
  - Increasing restaurant/hotel rebounds

- Highest income groups (>8000) have
  - High traveling rebounds especially air travel, but also personal
  - Savings start to dominate again after 10,000 CHF income
Outlook/ limitations

- Need to include trend of households from last 10 years (currently only trained on 2009-2011 HBS data)
- Multi-output regression models have lower coefficient of determination compared to single output model (preprocessing of data can make/break model)
- This model can be extended to any consumption rebound study, provided Household budget survey is available
Further steps

- Shrinking housing environmental footprint

- Household consumption and rebounds
  (Data Mining & Life Cycle Assessment)

- Building material and energy consumptions
  (Material Flow Analysis)

- Interaction in owners and occupants’ footprint
  (Agent Based Modeling)

Overall Housing Environmental Footprint for Switzerland
Overarching questions

- Useful instruments for combining environmental, economic and societal aims

(Explorative) Data analysis/ Data mining:
This study allows to look into the economic aspect of the consumptions, affordable housing and the consequences of this on the environmental footprint
Overarching questions

- Useful instruments for combining environmental, economic and societal aims

Agent Based Modeling

- Complete consumption footprint of household
- Tenant choices in selecting an apartment
Overarching questions

- How can decision-makers use life cycle based approaches to boost sustainable decisions?

  - Case-in point: Sustainable measures by building owners/cooperatives which induce saving of rent (e.g. energy savings, smaller houses) might have worse-off effects

  - Multi-stakeholder decision making (and risk/ opportunities spillover) -> upcoming slide
Overarching questions

- Which life cycle based approaches are best suited to reveal opportunities and risks for sustainability within the different economic sectors?
Overarching questions

- Which life cycle based approaches are best suited to reveal opportunities and risks for sustainability within the different economic sectors?

- This study allows to consider effects of one consumption industry on another and vice versa, and as it can extended to multiple sectors, the risks and opportunities of rebounds (spill-overs) can be clearly calculated.
Questions?

Thank you for your attention
Why Random Forest

<table>
<thead>
<tr>
<th>Need of model</th>
<th>RF</th>
<th>Linear</th>
<th>SVM</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>High dimensionality</td>
<td>Yes</td>
<td>Depends</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Handles missing value / outliers</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Learns non-linear complex relations</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Prediction possible</td>
<td>Yes</td>
<td>Yes</td>
<td>Depends</td>
<td>Yes</td>
</tr>
<tr>
<td>Handle data volatility</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
HABE->Clustering months

- **Step 1:** Box plots for HBS – visual aid (ascending order here)
- **Step 2:** ANOVA and post-hoc test (turkey-hsd)

### ANOVA tests

```
# ANOVA tests
stats.f_oneway(data_plot['food'][data_plot['month_name']=='January'],
data_plot['food'][data_plot['month_name']=='February'],
data_plot['food'][data_plot['month_name']=='March'])
```

```
F_onewayResult(statistic=4.263161581355489, pvalue=0.014183994064246183)
```

```
# Import pairwise comparison
from statsmodels.stats.multicomp import pairwise_tukeyhsd, MultiComparison
x=pairwise_tukeyhsd(data_plot['food'], data_plot['month_name'])
print(x)
```

**Multiple Comparison of Means - Tukey HSD, PHER=0.05**

<table>
<thead>
<tr>
<th>group1</th>
<th>group2</th>
<th>meandiff</th>
<th>lower</th>
<th>upper</th>
<th>reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>August</td>
<td>-12.6054</td>
<td>-144.3312</td>
<td>119.9604</td>
<td>False</td>
</tr>
<tr>
<td>April</td>
<td>March</td>
<td>-27.4215</td>
<td>-159.0065</td>
<td>105.0814</td>
<td>False</td>
</tr>
<tr>
<td>April</td>
<td>May</td>
<td>-32.7545</td>
<td>-162.928</td>
<td>97.419</td>
<td>False</td>
</tr>
<tr>
<td>April</td>
<td>November</td>
<td>-61.3852</td>
<td>-191.6423</td>
<td>69.93</td>
<td>False</td>
</tr>
<tr>
<td>April</td>
<td>October</td>
<td>-44.8692</td>
<td>-176.3565</td>
<td>88.198</td>
<td>False</td>
</tr>
<tr>
<td>April</td>
<td>September</td>
<td>-56.6317</td>
<td>159.9433</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>August</td>
<td>December</td>
<td>91.2929</td>
<td>-42.031</td>
<td>224.4307</td>
<td>False</td>
</tr>
<tr>
<td>August</td>
<td>February</td>
<td>50.4037</td>
<td>156.4247</td>
<td>62.447</td>
<td>False</td>
</tr>
<tr>
<td>August</td>
<td>January</td>
<td>159.4948</td>
<td>292.6479</td>
<td>26.3417</td>
<td>True</td>
</tr>
</tbody>
</table>
Clustering months

- **Step 3: Combine all categories on the statistical tests**

<table>
<thead>
<tr>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Feb</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mar</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Apr</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>May</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jun</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jul</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aug</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sep</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oct</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nov</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dec</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **Step 4: Verifying with plots and means**

  July – August Oct-Nov

  Jan-Feb-Mar
  Apr-May-Jun
  Dec
  Sep

www.esd.ifu.ethz.ch/