ETH zürich

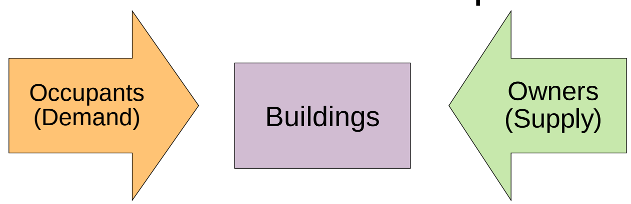
Data mining for evaluating impacts of rebounds in the housing sector of Switzerland

LCA Discussion Forum 74

Rhythima Shinde, Andreas Froemelt, Stefanie Hellweg

Motivation

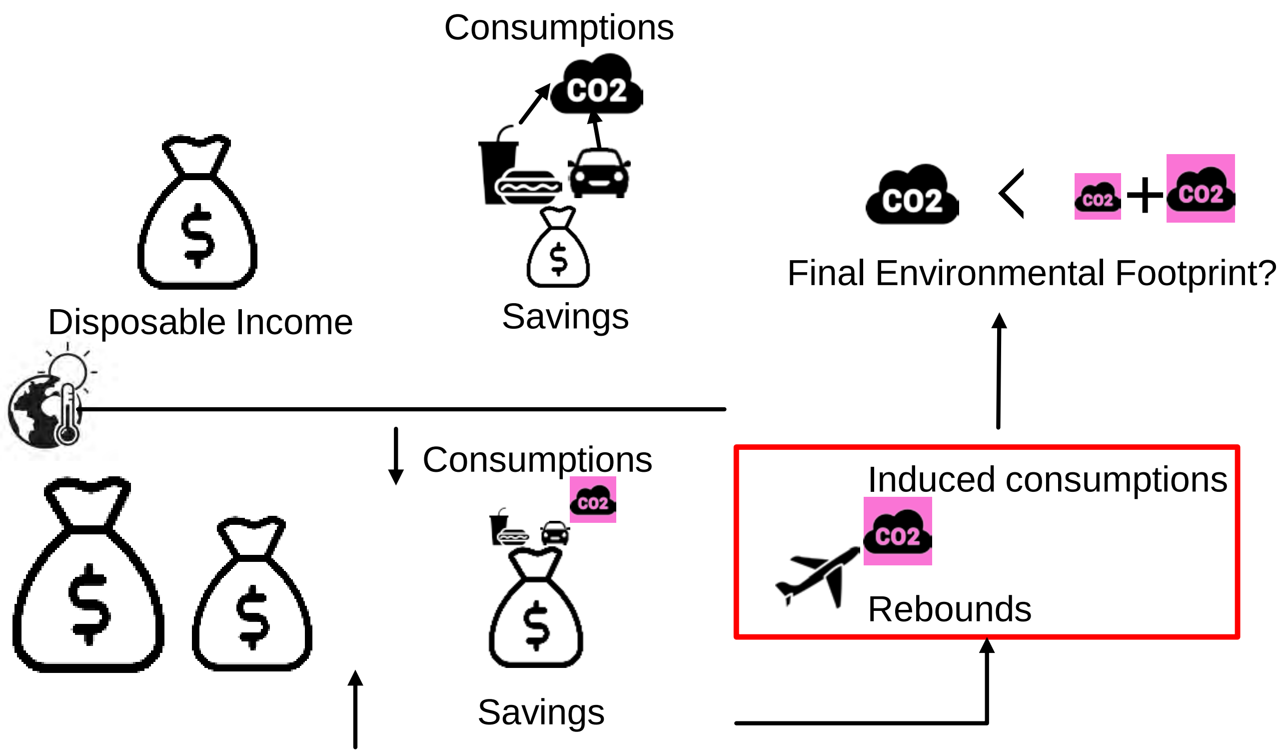
- Shrinking Housing Environmental Footprint:
  - Reducing the emissions of households of Switzerland due to housing market, household behaviors and the material footprint



- 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,

Allgemeine Baugenossenschaft Zürich (abz.ch), Société Coopérative d'Habitation Lausanne (schl.ch), Swiss Mobiliar (mobiliar.ch)

Motivation: Rebounds



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

= income – rent – compulsory fees (taxes + basic healthcare costs)

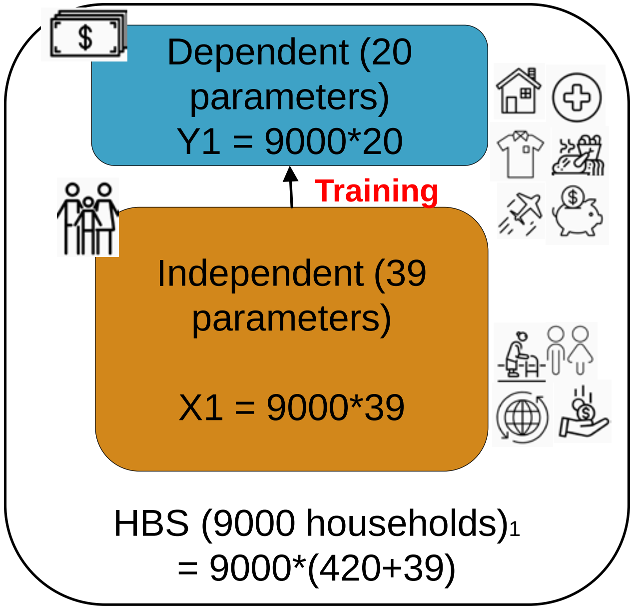
- Induced consumptions / Rebound

= 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

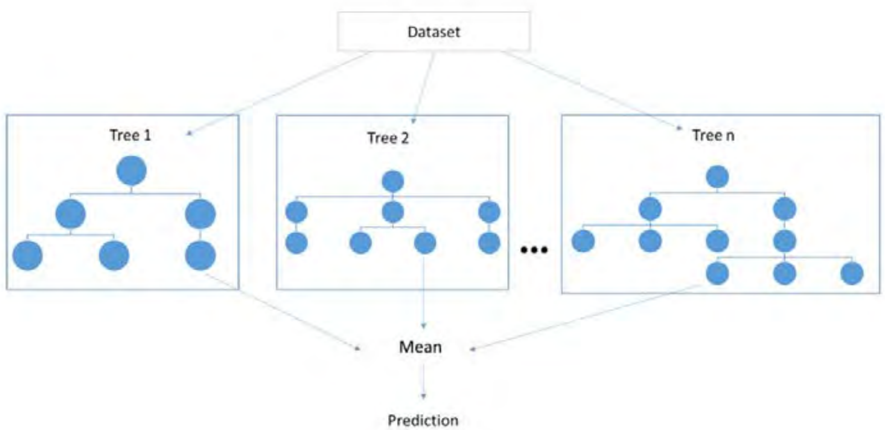


Data sources: Federal Statistical Office, Switzerland (2011). Froemelt et al. (2018)

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[1,2]
  - Advantage: Allows for higher dimensionality , handles missing values
  - Randomly choses decision trees based on given input features/ independent variables

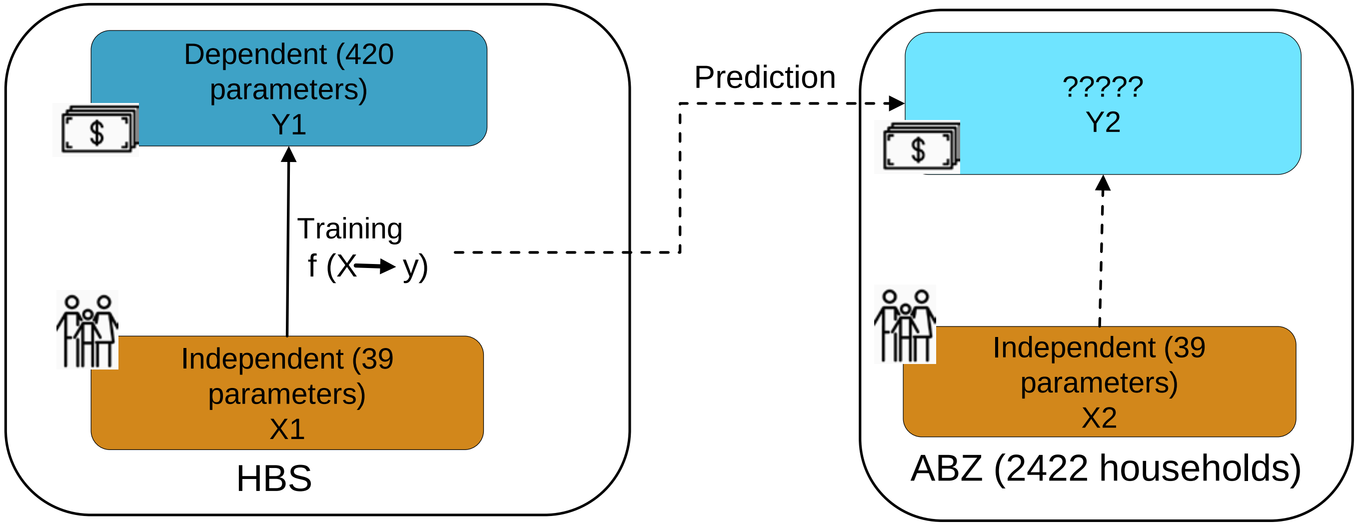


1. *Linusson, Henrik. "Multi-output random forests." (2013).*
2. *Segal, Mark, and Yuanyuan Xiao. "Multivariate random forests." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1.1 (2011): 80-87.*
3. *Image source: https://towardsdatascience.com/random-forests-and-decision-trees-from-scratch-in-python-3e4fa5ae4249*

Model: Methodology ('Prediction')

Based on the supervised learning/ training:

- Prediction of consumption profiles of households in Zurich cooperative (ABZ[1], 2009-2011)



Data HBS vs ABZ

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

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

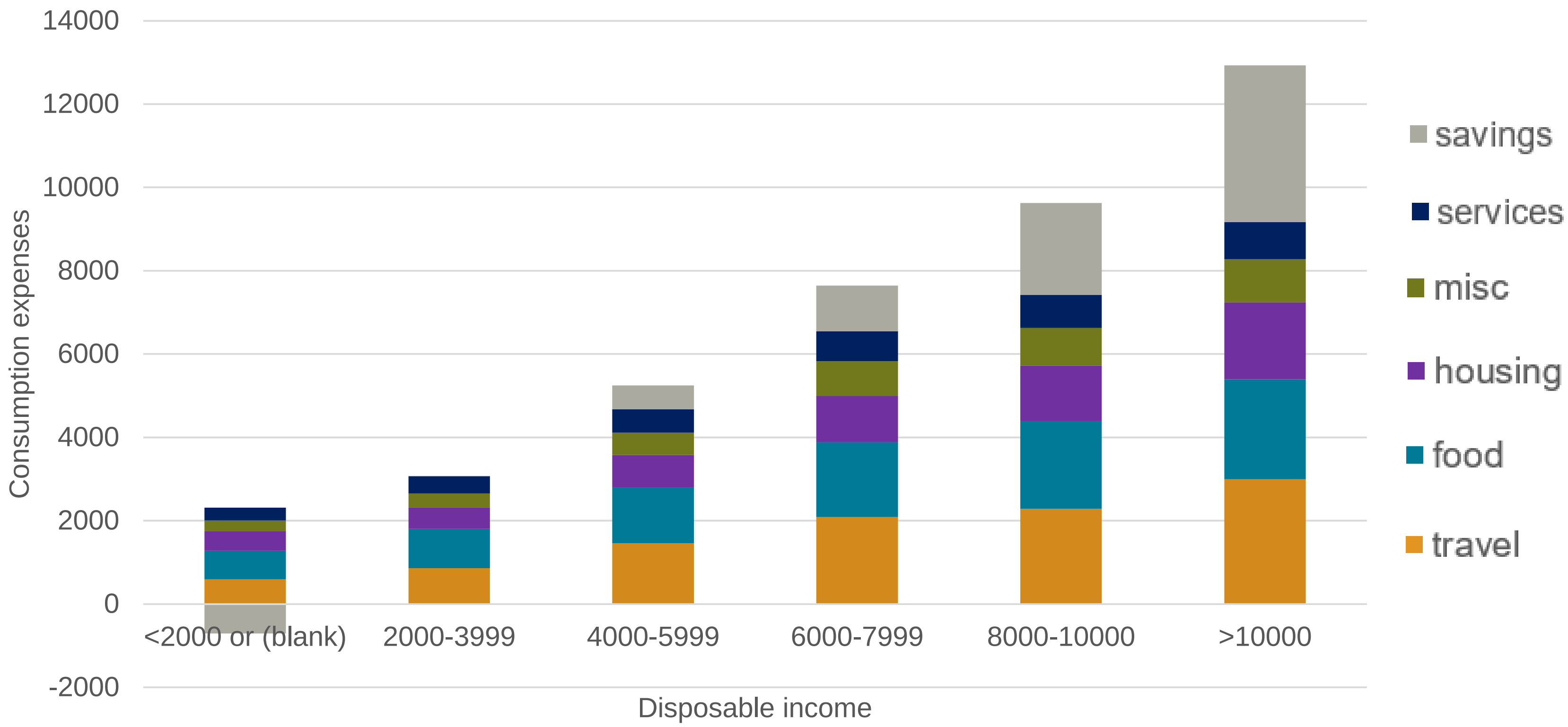
Monthly clusters for de-seasonalising

- Consumption expenses clustered together based on seasons (Statistical tests)

Preparing the stakeholder data

- Matching the household properties as per the Household budget survey

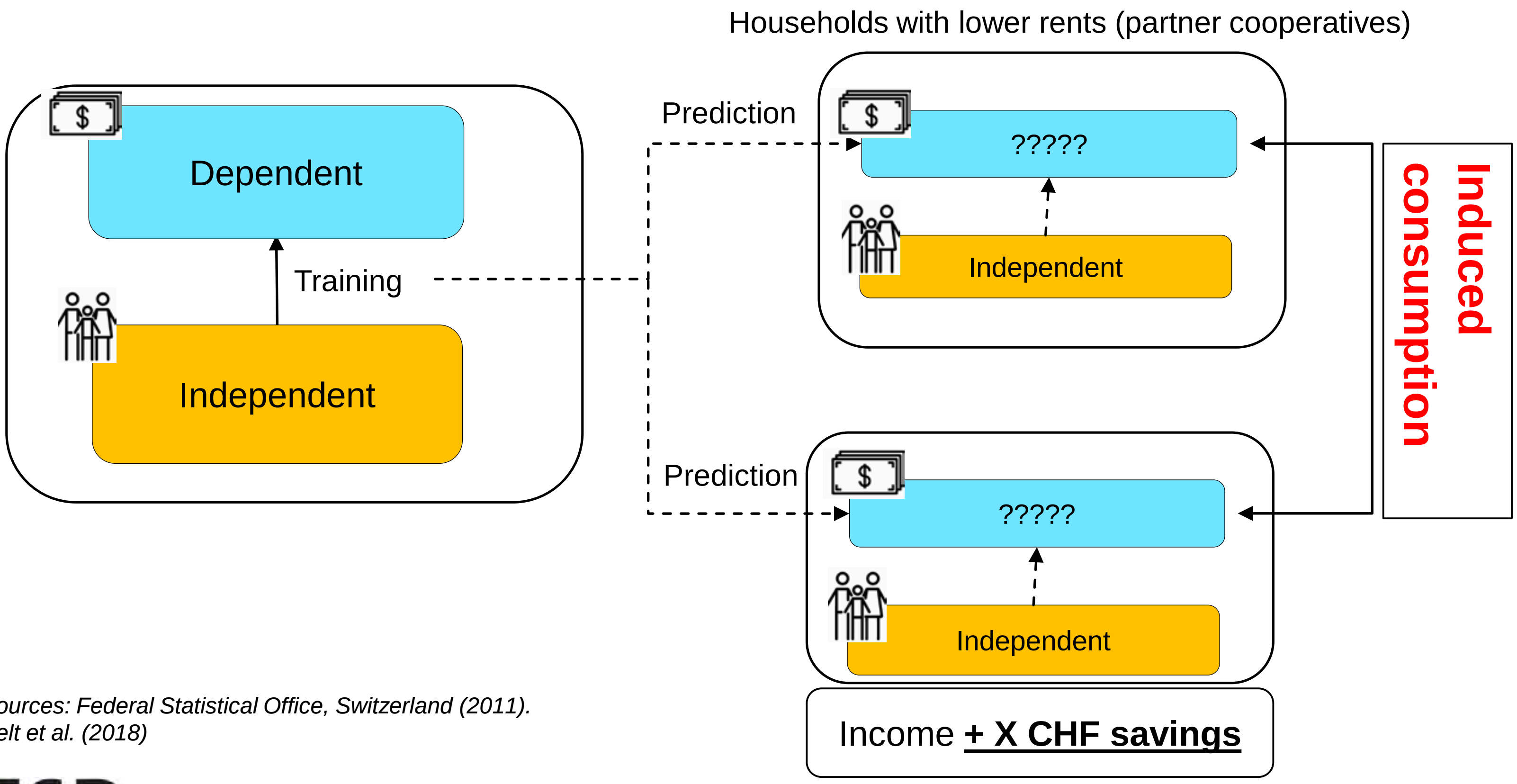
Prediction of consumption expenses



R2 -> coefficient of determination -> 0.52 to 0.97

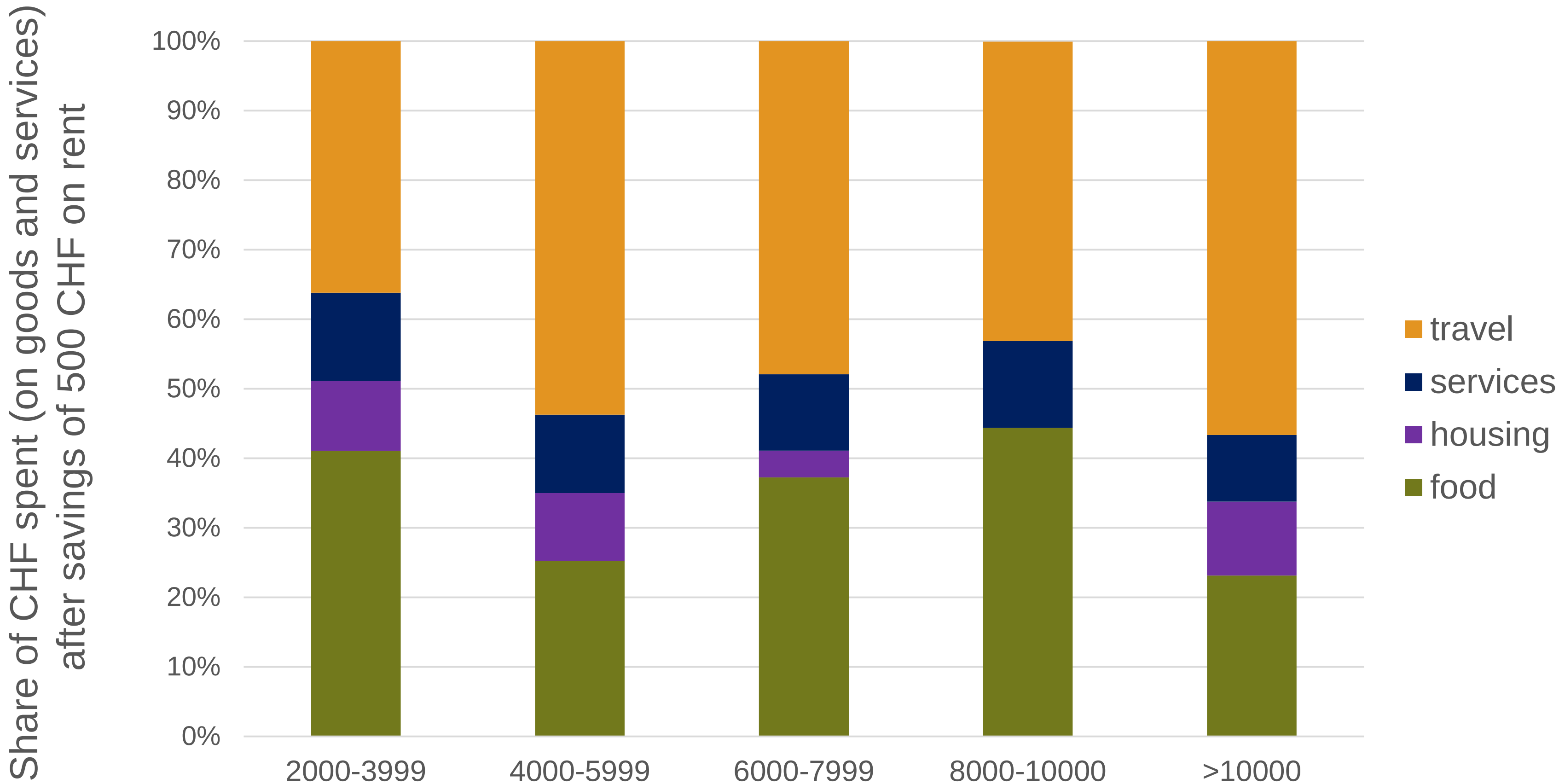
Method: Calculation of rebounds

- Random Forest (Prediction of consumptions)

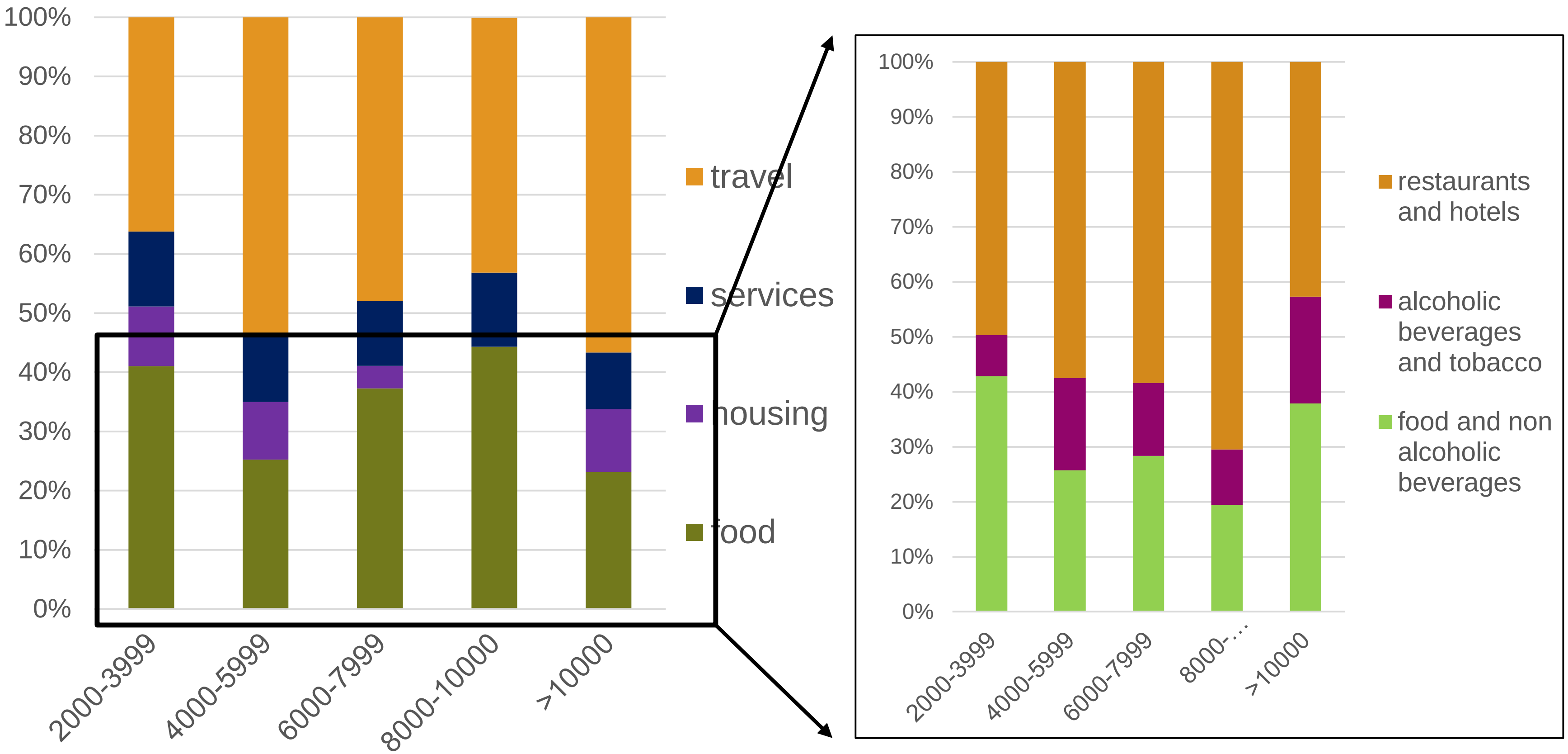


Data sources: Federal Statistical Office, Switzerland (2011). Froemelt et al. (2018)

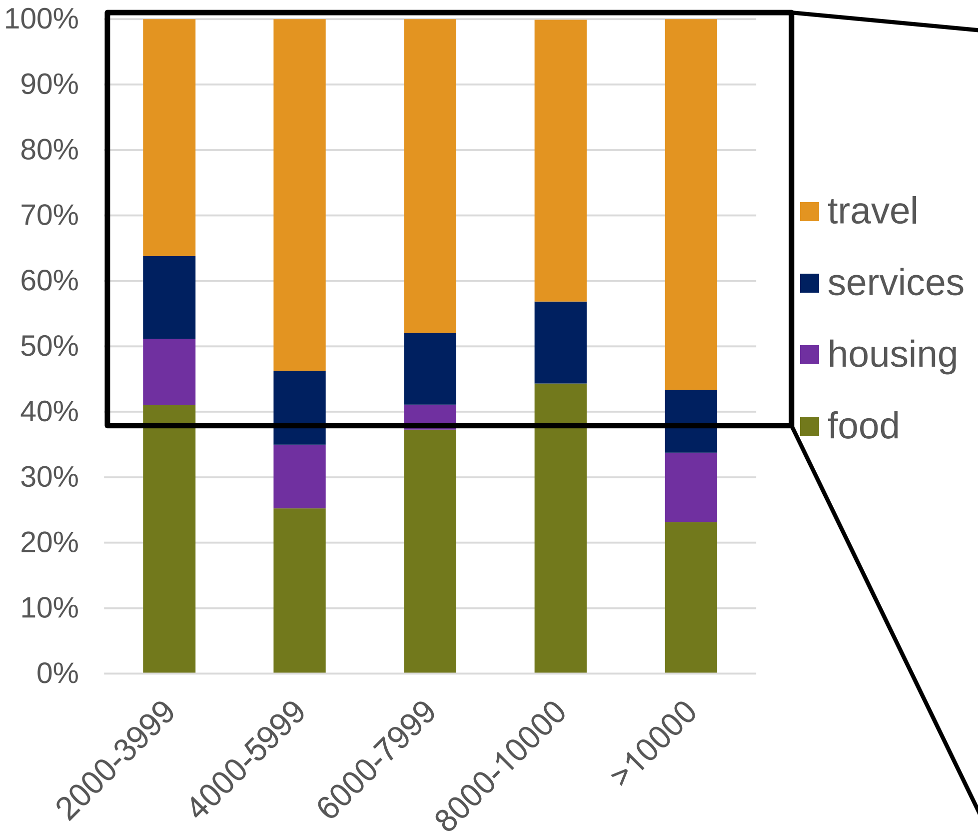
Results on induced consumptions

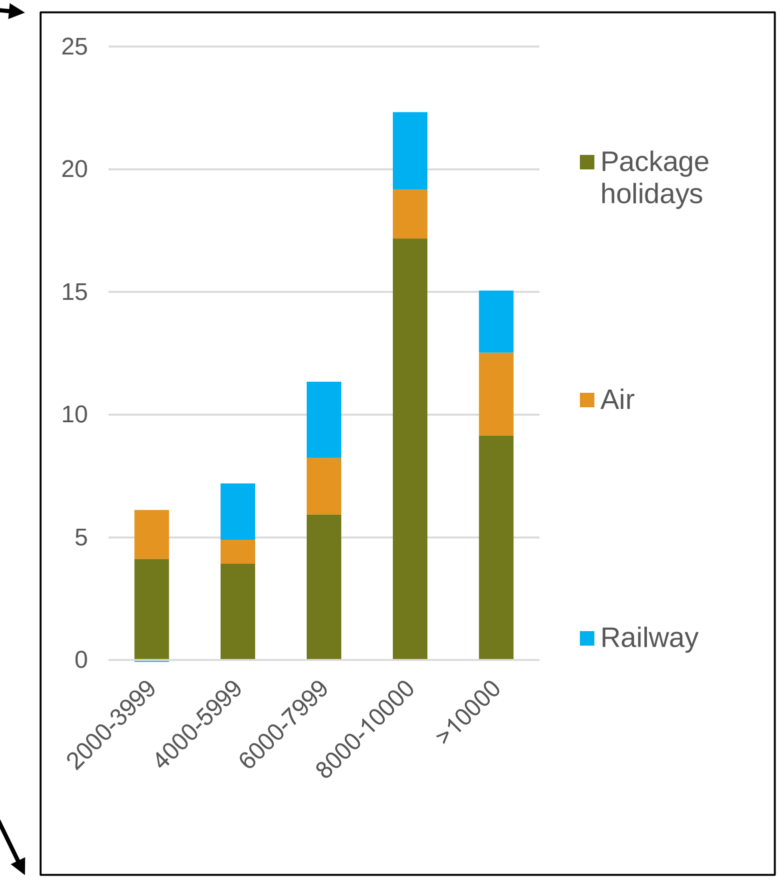


Results on induced consumptions: Food



Results on induced consumptions: Travel





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.

Translated name	*Other cereal products*
Variable code	a511108
Quantity code	m511108
ConversionDem2FU	1
On 1	0
Activity 1	('ecoinvent 3.3 cutoff', '8c7d59a4d38803
DB Act 1	'market for wheat grain' (kilogram, GLO,
CFL Act 1	1.18
Amount Act 1	0.4625
On 2	0
Activity 2	('ecoinvent 3.3 cutoff', '8c7d59a4d38803
DB Act 2	'market for wheat grain' (kilogram, GLO,
CFL Act 2	0.83
Amount Act 2	0.02
On 3	1
Activity 3	('ecoinvent 3.3 cutoff', '8a792c09ad0aed
DB Act 3	'market for sugar, from sugar beet' (kilog

Froemelt et al. (2018)

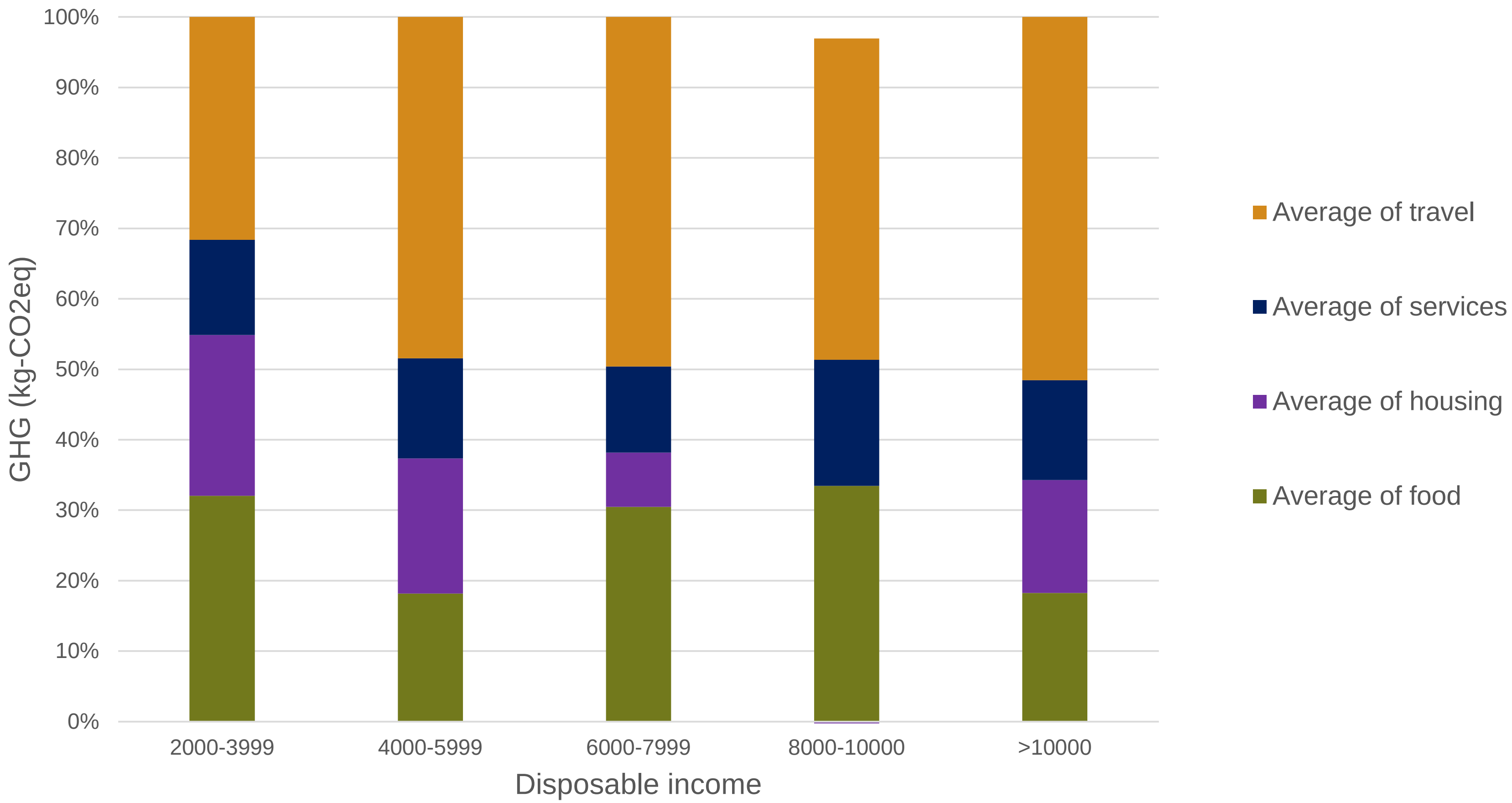
Consumption LCA

A process for every consumption category

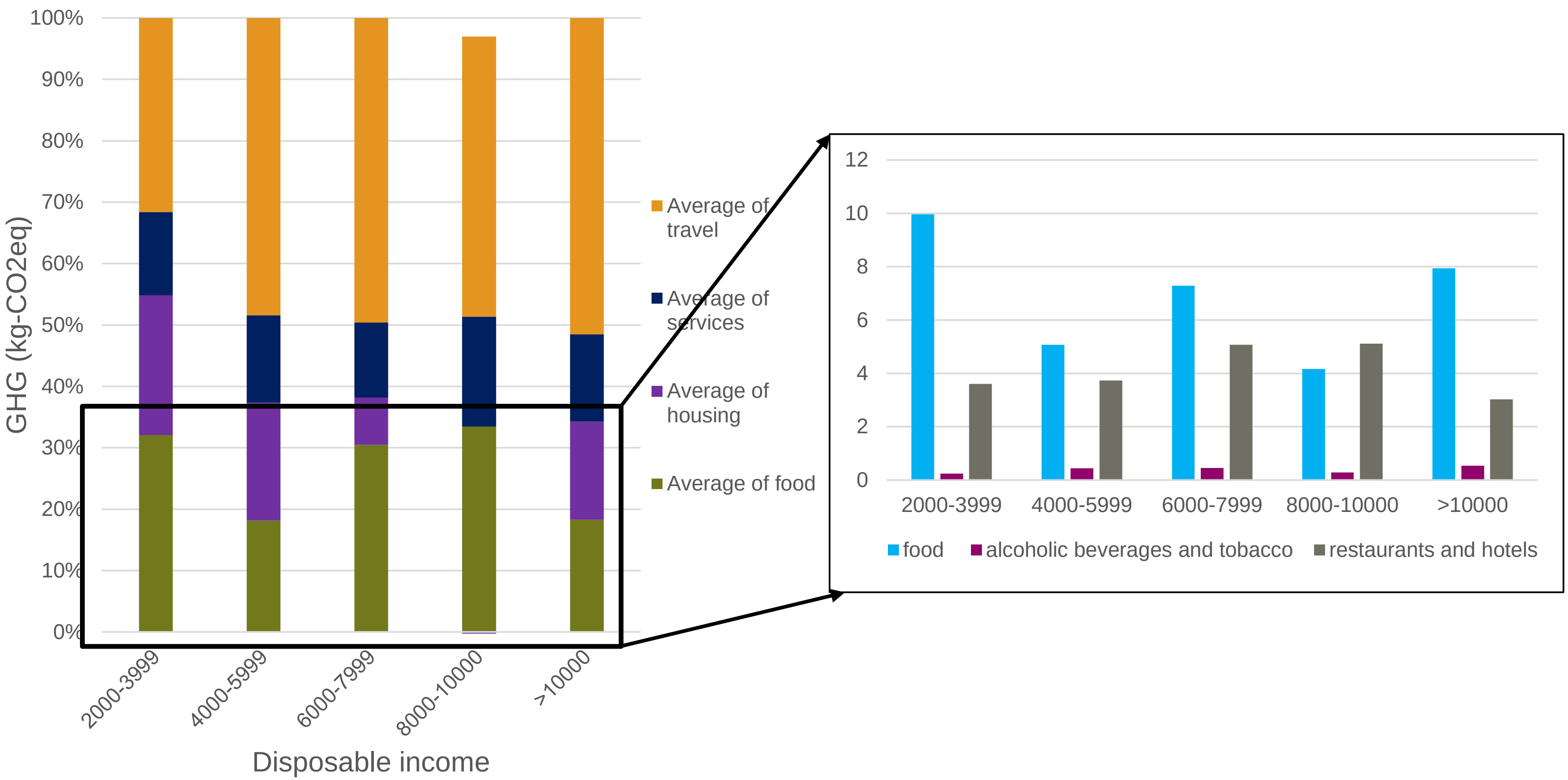
		m: quantity reported	LCA-Modeling		Detailed descri… on how to read the LCA-Modeling can be found in the Supporting Information document										
Translated name	Variable code	Quantity code	ConversionDem2FU	On 1	Activity 1	DB Act 1	CFL Act 1	Amount Act 1	On 2	Activity 2	DB Act 2	CFL Act 2	Amount Act 2	On 3	Act
Desktop computers	cg_nodesktoppcs		1	1	('ecoinvent 3.3 cu	'market for compu	0.25	1	1	('ecoinvent 3.3 cu	'market for displa	0.166666667	0.5	1	('ecoinv
Portable computers	cg_nolaptops		1	1	('ecoinvent 3.3 cu	'market for compu	0.25	1	1	('ecoinvent 3.3 cu	'market for pointin	0.25	1		
Printers (incl. multifunctional printers)	cg_noprinters		1	1	('ecoinvent 3.3 cu	'market for printer	0.25	0.5	1	('ecoinvent 3.3 cu	'market for printer	0.25	0.5		
Rice	a511101	m511101	1	0	('ecoinvent 3.3 cu	'market for rice' (k	1.44	1	0					0	
Pasta products	a511102	m511102	1	0	('ecoinvent 3.3 cu	'market for wheat	1.01	0.9	1	('Agribalyse 1.2',	'Egg, national ave	1	0.1	0	
Bread	a511103	m511103	1	0	('ecoinvent 3.3 cu	'market for wheat	0.88	1	1	('ecoinvent 3.3 cu	'market for sunflo	2.22	0.02	0	
Bakery products (sweet and salty)	a511104		0.070921986	0	('ecoinvent 3.3 cu	'market for wheat	1.01	0.483333333	1	('ecoinvent 3.3 cu	'market for sugar,	1	0.283333333	1	('heia',
Sandwich	a511105		0.037806122	0	('ecoinvent 3.3 cu	'market for wheat	1.01	0.5	1	('Agribalyse 1.2',	'Egg, national ave	1	0.2	1	('ecoinv
Wheat flour	a511106	m511106	1	0	('ecoinvent 3.3 cu	'market for wheat	1.01	1	1	('ecoinvent 3.3 cu	'wheat production	1.01	0.125356545	1	('ecoinv
Other flours and meals, starches, semolina, flakes and grains	a511107	m511107	1	0	('ecoinvent 3.3 cu	'market for maize	1	0.104440203	1	('ecoinvent 3.3 cu	'market for rye gr	1	0.053493927	1	('ecoinv
Other cereal products	a511108	m511108	1	0	('ecoinvent 3.3 cu	'market for wheat	1.18	0.4625	0	('ecoinvent 3.3 cu	'market for wheat	0.83	0.02	1	('ecoinv
Beef	a511201	m511201	1	1	('ecoinvent 3.3 cu	'cattle for slaughte	1.41	1	0					0	
Veal	a511202	m511202	1	1	('ecoinvent 3.3 cu	'cattle for slaughte	1.41	1	0					0	
Pork, fresh or frozen	a511203	m511203	1	1	('ecoinvent 3.3 cu	'market for swine	1.37	1	0					0	
Horse meat	a511204	m511204	1	1	('ecoinvent 3.3 cu	'cattle for slaughte	1.41	1	0					0	
Sheep and Goat meat	a511205	m511205	1	1	('ecoinvent 3.3 cu	'sheep for slaught	1.2	1	0					0	
Poultry, fresh or frozen	a511206	m511206	1	1	('ecoinvent 3.3 cu	'market for chicke	1.09	1	0					0	
Hare, game and rabbit meat	a511207	m511207	1	1	('ecoinvent 3.3 cu	'cattle for slaughte	1.41	1	0					0	
Other eatable meat products incl. offal, fresh and frozen	a511208	m511208	1	1	('ecoinvent 3.3 cu	'cattle for slaughte	1.41	0.133333333	1	('heia', 'vegetable	'vegetablenes' (ki	1.54	0.133333333	0	('ecoinv
Sausages, cold meat and pies	a511209	m511209	1	1	('ecoinvent 3.3 cu	'market for swine	1.37	0.9	0	('ecoinvent 3.3 cu	'market for wheat	0.88	0.1	0	
Ham, bacon and other cured or smoked pork	a511210	m511210	1	1	('ecoinvent 3.3 cu	'market for swine	1.3	1	0					0	
Poultry, grilled or smoked	a511211	m511211	1	1	('ecoinvent 3.3 cu	'market for chicke	1.09	0.7	0	('ecoinvent 3.3 cu	'market for wheat	0.88	0.2	1	('ecoinv
Other boiled, dried, cured or smoked meat	a511212	m511212	1	1	('ecoinvent 3.3 cu	'market for swine	1.25	0.5	1	('ecoinvent 3.3 cu	'cattle for slaughte	2.17	0.5	0	
Tinned meat and other meat-based preparations	a511213	m511213	1	1	('ecoinvent 3.3 cu	'cattle for slaughte	1.41	1	0					0	
Fish	a5113	m5113	1	1	('Agribalyse 1.2',	'Sea bass or sea	1.18	0.601250827	1	('Agribalyse 1.2',	'Large trout, 2-4k	1.18	0.199374587	1	('Agriba
Whole milk	a511401	m511401	1	1	('ecoinvent 3.3 cu	'market for cow m	1	1	0					0	
Skimmed and low-fat milk	a511402	m511402	1	1	('ecoinvent 3.3 cu	'market for skimm	1	1	0					0	
Hard and semi-hard cheese	a511403	m511403	1	1	('ecoinvent 3.3 cu	'market for chees	1	1	0					0	
Fresh, soft and melted cheese	a511404	m511404	1	1	('ecoinvent 3.3 cu	'market for chees	1	1	0					0	
Cream	a511405	m511405	1	1	('ecoinvent 3.3 cu	'market for cream	1	1	0					0	
Curd	a511406	m511406	1	1	('ecoinvent 3.3 cu	'market for chees	1	1	0					0	
Yoghurt	a511407	m511407	1	1	('ecoinvent 3.3 cu	'market for yogurt	1	1	0					0	
Milk-based beverages and other similar milk-based products	a511408	m511408	1	1	('ecoinvent 3.3 cu	'market for cow m	0.5	0.5	1	('ecoinvent 3.3 cu	'market for sugar,	1	0.1	1	('heia',
Fresh eggs	a511409		0.162735849	1	('Agribalyse 1.2',	'Egg, national ave	1	1	0					0	
Processed eggs	a511410		0.162735849	1	('Agribalyse 1.2',	'Egg, national ave	1	1	0					0	
Butter	a511501	m511501	1	1	('ecoinvent 3.3 cu	'market for butter,	1	1	0					0	
Margarine	a511502	m511502	1	0	('ecoinvent 3.3 cu	'market for rape s	1	1	0					0	
Other vegetable fats	a511503	m511503	1	0	('ecoinvent 3.3 cu	'market for rape s	1	1	0					0	
Olive oil	a511504	m511504	1	0	('ecoinvent 3.3 cu	'market for olive' (	4.8	1	0					0	
Other vegetable oils and edible animal fats	a511505	m511505	1	1	('ecoinvent 3.3 cu	'market for vegeta	1	1	0					0	
Lemons	a511601	m511601	1	0	('ecoinvent 3.3 cu	'market for lemon'	1	1							
Oranges and other citrus fruits	a511602	m511602	1	0	('ecoinvent 3.3 cu	'market for orang	1	0.89214826	1	('ecoinvent 3.3 cu	'market for manda	1	0.10785174		
Banana	a511603	m511603	1	0	('ecoinvent 3.3 cu	'market for banan	1	1							
Apples	a511604	m511604	1	1	('ecoinvent 3.3 cu	'market for apple'	1	1							
Pears and quinces	a511605	m511605	1	1	('ecoinvent 3.3 cu	'market for pear' (	1	1							
Stone fruit	a511606	m511606	1	0	('ecoinvent 3.3 cu	'market for peach	1	0.241340861	1	('ecoinvent 3.3 cu	'market for aprico	1	0.166963722	1	('ecoinv
Berries	a511607	m511607	1	1	('ecoinvent 3.3 cu	'market for strawb	1	1							

Froemelt et al. (2018)

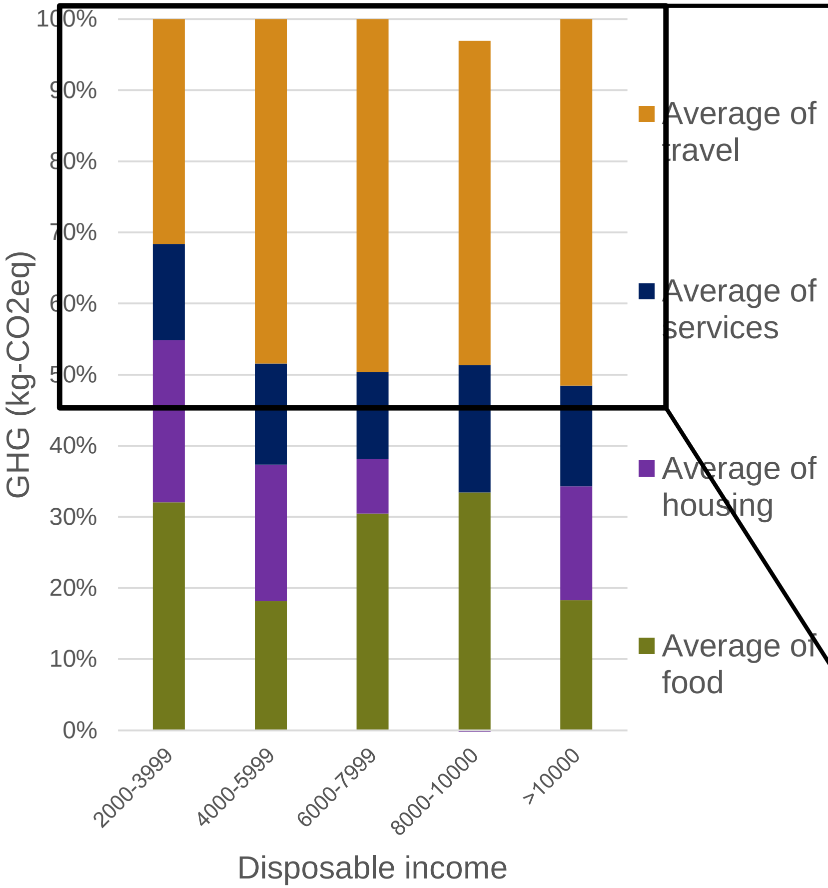
Results on environmental footprint of rebounds

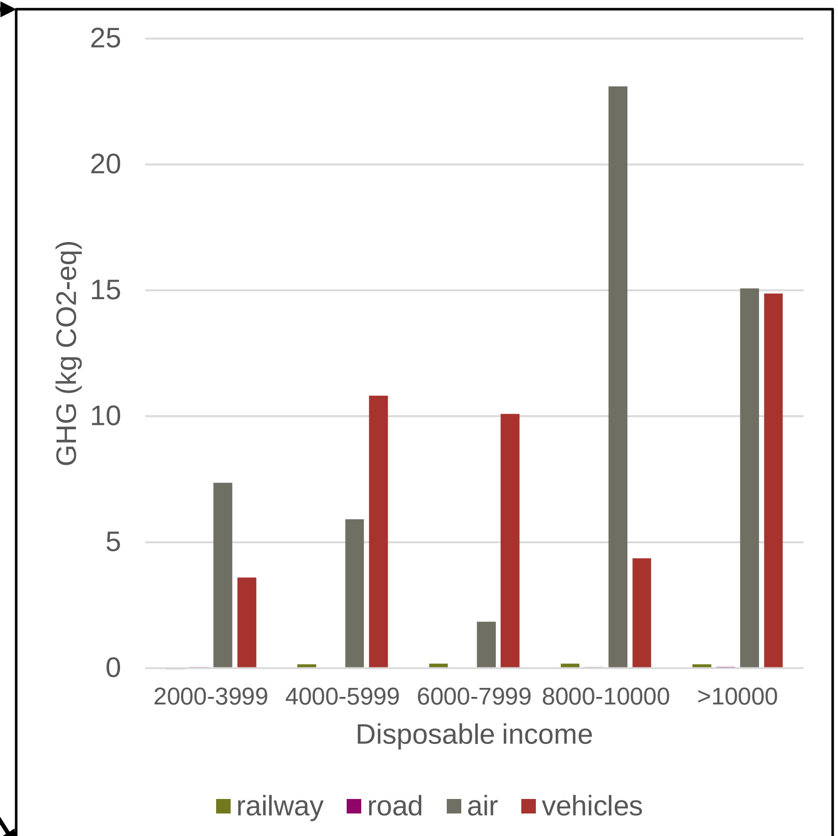


Results on environmental footprint: Food



Results on environmental footprint: Travel





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

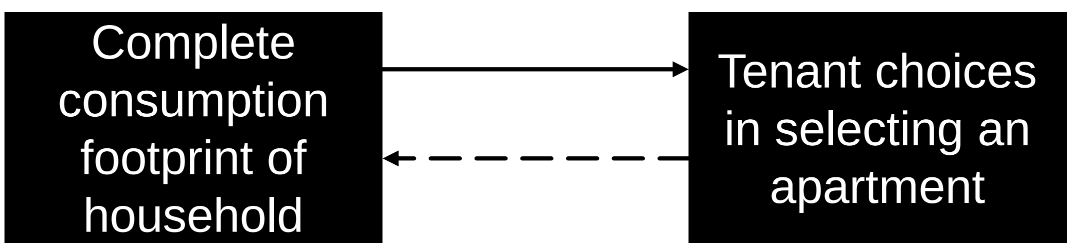
  (Explorative) Data analysis/ Data mining:

  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



Overarching questions

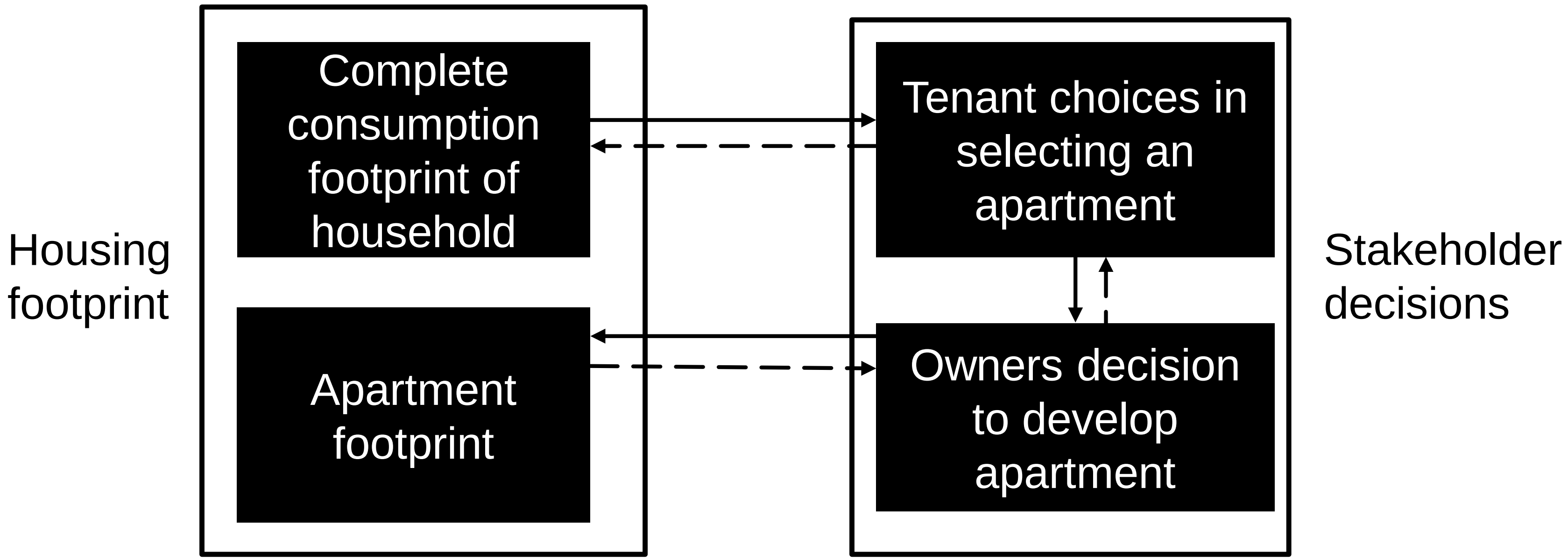
- How can decision-makers use life cycle based approaches to boost sustainable decisions?

  - 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

ETH zürich

Questions?

73 NRP Sustainable Economy National Research Programme

FNSNF Swiss National Science Foundation

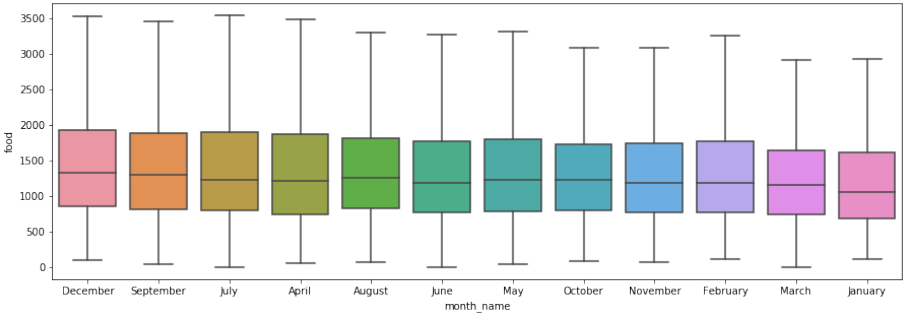
Thank you for your attention

Why Random Forest

Need of model	RF	Linear	SVM	ANN
High dimensionality	Yes	Depends	Yes	Yes
Handles missing value / outliers	Yes	No	Yes	Yes
Learns non-linear complex relations	Yes	No	Yes	Yes
Prediction possible	Yes	Yes	Depends	Yes
Handle data volatility	Yes	No	Yes	Yes

HABE->Clustering months

- Step 1: Box plots for HBS – visual aid (ascending order here)



- Step 2: ANOVA and post-hoc test (turkey-hsd)

```python
## 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)
```

```python
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,FWER=0.05
=======================================================
 group1   group2   meandiff   lower     upper   reject
-------------------------------------------------------
 April    August   -12.6854 -144.3312 118.9604 False
 April   December  78.5175   -52.024  209.0589 False
 April   February  -63.1755 -194.4319  68.0809 False
 April   January  -172.1802 -302.6393 -41.7212  True
 April     July    36.0324    -94.02  166.0848 False
 April     June    -27.4125 -159.9065 105.0814 False
 April    March    -92.2668 -222.8496  38.316  False
 April     May     -32.7545  -162.928   97.419  False
 April   November  -61.3062 -191.6423  69.03   False
 April   October   -44.0692 -176.3365  88.198  False
 April  September  71.6558   -56.6317  199.9433 False
 August  December  91.2029   -42.031   224.4367 False
 August  February  -50.4901 -184.4245  83.4443 False
 August  January  -159.4948 -292.6479 -26.3417  True
```

Clustering months

July –August
~~Oct-Nov~~??

- Step 3: Combine all categories on the statistical tests

total												
	January	February	March	April	May	June	July	August	September	October	Novembe	December
January												
February	0											
March	0	0										
April	1	1	0									
May	1	0	0	0								
June	1	0	0	0	0							
July	1	0	0	0	0	0						
August	1	0	0	1	0	0	0					
September	1	2	2	0	0	1	0	1				
October	0	0	0	0	0	0	0	0	0			
November	0	0	0	0	0	1	0	0	1	0		
December	2	3	3	1	0	2	1	2	1	1	3	

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

Jul –Aug
Oct-Nov

- Step 4: Verifying with plots and means

	month_n	housing_data		
9	Septemb	614.469		
11	Novembe	612.113	0.38342	
1	January	576.317	6.20902	
2	February	567.255	7.68374	
12	Decembe	558.65	9.08422	
5	May	555.985	9.51792	
10	October	546.407	11.0766	10.69
7	July	545.656	11.1989	
8	August	544.817	11.3353	
3	March	541.597	11.8594	
4	April	508.458	17.2524	
6	June	472.65	23.0799	

	month_n	misc_data		
12	Decembe	610.672		
4	April	553.302	9.39463	
9	Septemb	552.119	9.58833	
5	May	540.75	11.45	
10	October	536.806	12.0959	
11	Novembe	496.43	18.7075	6.61
1	January	495.356	18.8835	
7	July	479.135	21.5397	
3	March	473.997	22.381	
6	June	456.579	25.2333	
2	February	449.166	26.4472	
8	August	442.943	27.4663	

	month_n	housing_data	
9	Septemb	614.469	
11	Novembe	612.113	0.38342
1	January	576.317	6.20902
2	February	567.255	7.68374
12	Decembe	558.65	9.08422
5	May	555.985	9.51792
10	October	546.407	11.0766
7	July	545.656	11.1989
8	August	544.817	11.3353
3	March	541.597	11.8594
4	April	508.458	17.2524
6	June	472.65	23.0799

	month_n	misc_data		
12	Decembe	610.672		
4	April	553.302	9.39463	
9	Septemb	552.119	9.58833	
5	May	540.75	11.45	
10	October	536.806	12.0959	
11	Novembe	496.43	18.7075	
1	January	495.356	18.8835	
7	July	479.135	21.5397	
3	March	473.997	22.381	
6	June	456.579	25.2333	
2	February	449.166	26.4472	
8	August	442.943	27.4663	5.93