Online Appendix

Appendix B.1: Statistical Matching of Income Data from EU-SILC & Microcensus

A major limitation of the Austrian Microcensus (MC) survey is its lack of income information. In the interviews, no information about income is gathered from the respondents. Income data of employees (unselbstständig Beschäftigte) is added later on from wage tax registers. However, no income data is available for other groups like the unemployed, retired people, self-employed, students, etc. and also information about possible additional income components (e.g. earnings from rent and lease, social transfers, etc.) is missing. To address this lack of income data, it was attempted to impute the missing income information from the EU-SILC survey dataset by applying two statistical matching techniques (Nearest Neighbour Distance Hot Deck Imputation and a Random Forest Algorithm), following the approaches by Baud and Wegscheider-Pichler (2019), Puchner (2015), and Wegscheider-Pichler (2014).

Statistical matching is a model-based approach to integrate information from two (or more) data sources (usually sample surveys), which both refer to the same target population and share a set of common variables as well as one or more variables that are only available in one of the datasets (D'Orazio et al. 2006; Eurostat 2014). It aims to create a synthetic dataset in which all variables are jointly available.

In the study at hand, the goal was to impute the personal monthly net income, which is available in the EU-SILC dataset (the donor data source), in the MC dataset based on a set of matching variables. Matching variables need to fulfil two essential criteria: 1) they are available in both data sources, and 2) they exhibit a significant correlation with the variable that shall be imputed. The choice of matching variables in this study is based on the work of Wegscheider-Pichler (2014). The following matching variables were selected: gender, age, occupational status, level of job, highest educational attainment, household size, country of birth, nationality and income from employment. After harmonisation of the specification of the variables in the two datasets, OLS regression was performed to test the association with the income variable to be imputed. Table A1 shows that all matching variables are highly significantly ($p < 0.01$) correlated with the personal monthly net income, and the model predicts 42% of the variance in personal monthly net income ($R2 = 0.42$), which proves that the selected set of matching variables is a good choice for the matching procedure.

Two different statistical matching techniques were applied:

- *Nearest Neighbour Distance Hot Deck* imputation (NND) using the *StatMatch*-package in R (D'Orazio 2020). This matching algorithm selects the *nearest neighbour* from the donor dataset to each unit of the recipient dataset based on a distance function of the matching variables and matches the records from both datasets. Therefore, the imputed values in the recipient dataset are actually *observed* values from the donor dataset. Due to the mixed nature of the matching variables (categorical and continuous variables) in the study at hand, the *Gower*-distance (D'Orazio et al. 2006, p.216) was selected as the distance function of choice.

- a *Random Forest* (RF) algorithm with an additional *predictive mean matching* step as implemented in the *missRanger*-package in R (Mayer 2021). The RF algorithm is a machine learning imputation algorithm that creates *synthetic* values based on multiple iterations of a decision tree, which aims to calculate the best fitting value based on the matching variables.
  The *predictive mean matching* step in the *missRanger*-package improves the RF by ensuring a more realistic distribution of the imputed data and avoiding unrealistic values (e.g. values with many fractional digits, though the variable contained only integer values in the donor dataset.

Results of the statistical matching procedure are presented in Figure A1 and A2, which compare the distribution of personal monthly net income in the donor dataset (EU-SILC) to the distribution of personal monthly net income in the recipient dataset (MC) and therefore allow for graphical evaluation of the quality of the statistical matching approaches. Both techniques deliver valid results that represent the overall income distribution in the EU-SILC well. However, both approaches exhibit significant discrepancies on the lower end of the income bandwidth. This is problematic, because lower-income groups are of particular interest in analysing configurations of *double burden* regarding the perceived exposure to adverse environmental conditions. Therefore, the results need to be treated with great caution. As the income distribution generated by the NND approach provides a slightly better fit to the income distribution in the EU-SILC dataset, the income data generated by the NND technique is used to analyse the association with perceived environmental burdens and the income data gained by the RF algorithm is discarded.

Tables A2 and A3 show the results of the logistic regression models, including the imputed personal monthly net income as an explaining variable of environmental quality or exposure, respectively. The personal monthly net income does not exhibit significant influence in any of the models. Considering the findings from Baud and Wegscheider-Pichler (2019) and the fact that employee income is significantly associated with the perception of overall environmental quality and the exposure to air pollution, the results do not seem plausible and rather suggest that the statistical matching procedure did not deliver results of the necessary quality for further detailed analysis.

D'Orazio, M. (2020). StatMatch: Statistical Matching or Data Fusion. R package version 1.4.0. https://CRAN.R-project.org/package=StatMatch

D'Orazio, M., Di Zio, M., & Scanu, M. (2006). Statistical Matching: Theory and Practice. Wiley.

Eurostat. (2014). Statistical Matching Methods (Memobust Handbook on Methodology of Modern Business Statistics). https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en

Mayer, M. (2021). missRanger: Fast Imputation of Missing Values. R package version 2.1.4. https://github.com/mayer79/missRanger

Puchner, V. (2015). Evaluation of Statistical Matching and Selected SAE Methods: Using Micro Census and EU-SILC Data. Springer. https://doi.org/10.1007/978-3-658-08224-6

Wegscheider-Pichler, A. (2014). Umweltbetroffenheit und -verhalten von Personengruppen abhängig von Einkommen und Kaufkraft: Mikrozensus und EU-SILC - Statistical Matching. Statistik Austria.

Table B1: OLS Regression of Matching Variables

	Personal Monthly Net Income
gender (female)	-0.298***
age	0.011***
working (yes)	1.234***
workinghours	0.008***
status (unemployed)	0.099
status (retired)	1.069***
status (student)	-1.171***
status (household)	-3.528***
status (other)	-0.991***
joblevel	-0.043***
education	0.079***
household size	-0.111***
employee income	0.0002***
countryofbirth (EU)	-0.352*** (0.098)
countryofbirth (other)	-0.020
nationality (EU)	-0.361***
nationality (other)	-0.635***
Constant	5.471***
Observations	10,042
R2	0.418
Adjusted R2	0.417
Residual Std. Error	0.017 (df = 10024)
F Statistic	423.261*** (df = 17; 10024)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure B1: Distribution of Personal Monthly Net Income: EU-SILC vs. Imputation in MC

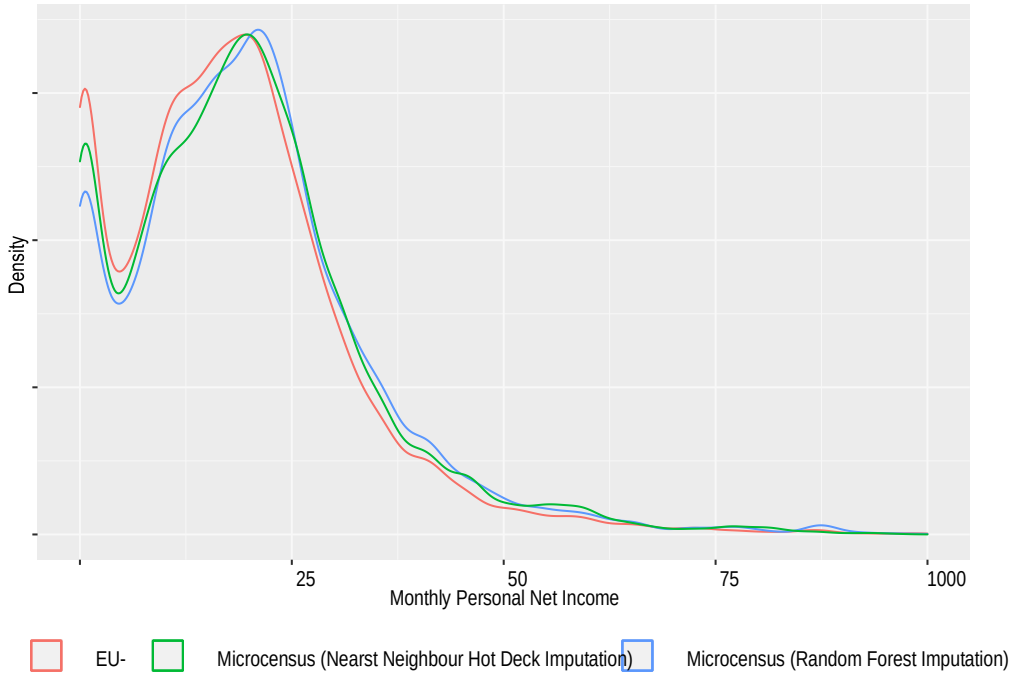

Figure B2: Personal Monthly Net Income: EU-SILC vs. Imputation in MC

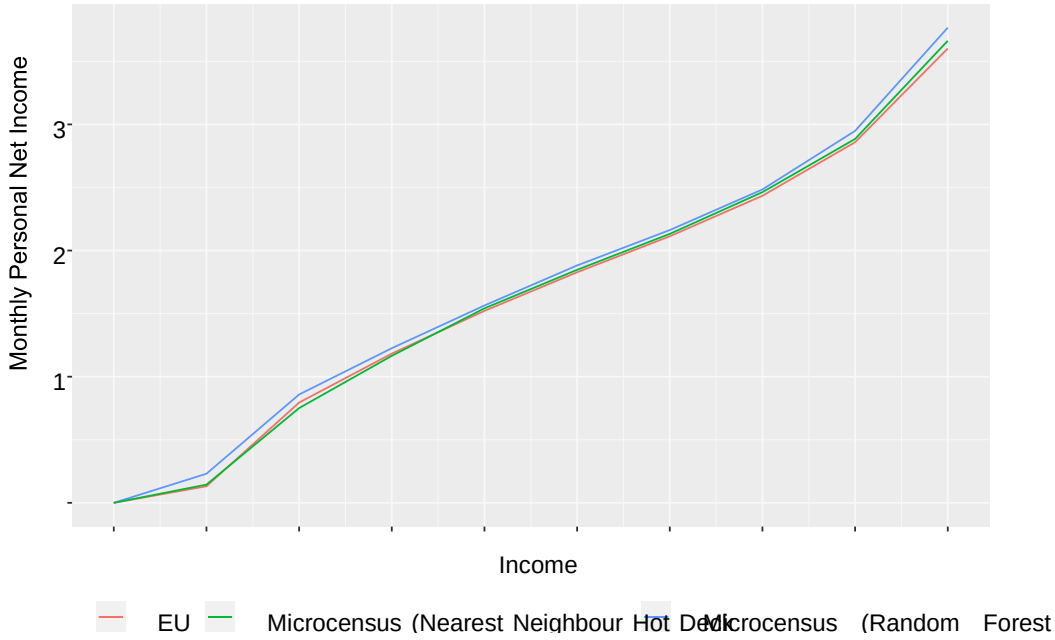

Table B2: Determinants of Perceived Environmental Quality (Binomial Logistic Regression)

	Environmental Quality		Air Quality		Noise Situation		Greenspace	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age	1.00	(1.00;1.01)	1.01	(1.00;1.01)	1.01*	(1.00;1.01)	1.01	(1.00;1.01)
Gender								
male	1.00		1.00		1.00		1.00	
female	1.66***	(1.36;2.03)	1.55***	(1.27;1.89)	1.23***	(1.06;1.42)	1.48***	(1.16;1.89)
Education	0.71***	(0.65;0.77)	0.92**	(0.85;0.99)	0.65*	(0.90;1.01)	0.84***	(0.76;0.93)
Income	1.00	(0.99;1.00)	1.00	(0.99;1.01)	1.00	(1.00;1.01)	1.00	(0.99;1.01)
Job Level	0.93**	(0.88;0.99)	0.97	(0.92;1.03)	0.98	(0.94;1.03)	1.00	(0.94;1.08)
Migration Background								
no	1.00		1.00		1.00		1.00	
yes	0.51***	(0.35;0.75)	0.83	(0.61;1.12)	1.18	(0.90;1.39)	0.62**	(0.41;0.93)
Urbanisation	1.03	(0.90;1.18)	1.48***	(1.29;1.69)	1.30***	(1.18;1.44)	1.42***	(1.20;1.68)
Type of Family								
couple w/o children	1.00		1.00		1.00		1.00	
couple w/ children	0.96	(0.74;1.25)	0.94	(0.71;1.23)	0.90	(0.74;1.09)	0.98	(0.72;1.32)
singles	1.11	(0.88;1.40)	1.13	(0.90;1.42)	1.04	(0.87;1.23)	1.14	(0.88;1.48)
single parents	1.27	(0.88;1.83)	1.24	(0.85;1.79)	0.96	(0.72;1.29)	0.95	(0.59;1.51)
other	1.25	(0.74;2.11)	0.81	(0.49;1.35)	1.06	(0.71;1.57)	1.15	(0.66;2.02)
Flats								
1 or 2	1.00		1.00		1.00		1.00	
3 or more	1.28**	(1.03;1.59)	1.63***	(1.31;2.02)	1.33***	(1.13;1.58)	1.21	(0.93;1.57)
Category of Flat								
cat. A	1.00		1.00		1.00		1.00	
cat. B or below	0.93	(0.62;1.40)	1.05	(0.68;1.63)	1.18	(0.57;1.39)	1.93***	(1.24;3.00)

n = 3,749

*p<0.1; **p<0.05; ***p<0.01

Table B3: Determinants of Perceived Environmental Exposure (Ordinal Logistic Regression)

	Odour / Fumes		Dust / Soot		Noise		Heat	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age	1.00	(0.99;1.01)	1.01***	(1.01;1.02)	0.99**	(0.99;1.00)	0.99***	(0.99;1.00)
Gender								
male	1.00		1.00		1.00		1.00	
female	1.26	(0.94;1.68)	1.1	(0.91;1.34)	1.17*	(0.99;1.40)	1.26***	(1.12;1.42)
Education	0.93	(0.83;1.05)	1.04	(0.97;1.12)	1.08**	(1.01;1.15)	0.94**	(0.90;0.99)
Income	0.99	(0.98;1.00)	1	(0.99;1.01)	1	(0.99;1.01)	1	(1.00;1.00)
Job Level	0.97	(0.89;1.05)	1.03	(0.98;1.09)	0.97	(0.92;1.02)	1	
Migration Background								
no	1.00		1.00		1.00		1.00	
yes	1.28	(0.86;1.92)	1.37**	(1.05;1.09)	1.87	(0.68;1.11)	0.96	(0.80;1.15)
Urbanisation	1.39***	(1.15;1.68)	1.42***	(1.24;1.63)	1.54***	(1.38;1.73)	1.21***	(1.12;1.31)
Type of Family								
couple w/o children	1.00		1.00		1.00		1.00	
couple w/ children	0.79	(0.54;1.18)	0.9	(0.70;1.60)	0.95	(0.77;1.17)	1.04	(0.89;1.20)
singles	0.89	(0.64;1.24)	1.21*	(0.98;1.50)	0.92	(0.76;1.12)	1.14*	(0.98;1.32)
single parents	1.32	(0.78;2.21)	1.50**	(1.04;2.17)	1.02	(0.75;1.39)	1.38**	(1.07;1.77)
other	0.63	(0.31;2.89)	0.99	(0.60;1.64)	0.66**	(0.44;0.99)	0.82	(0.59;1.14)
Flats								
1 or 2	1.00		1.00		1.00		1.00	
3 or more	1.11	(0.82;1.50)	1.22*	(0.97;1.55)	1.42***	(1.18;1.71)	1.07	(0.93;1.23)
Category of Flat								
cat. A	1.00		1.00		1.00		1.00	
cat. B or below	1.24	(0.58;2.67)	1.11	(0.75;1.65)	0.97	(0.68;1.38)	1.12	(0.84;1.51)

n = 3,749

*p<0.1; **p<0.05; ***p<0.01

Appendix B.2: Bivariate Analysis

Table B4: Determinants of Perceived Environmental Quality (Binomial Logistic Regression)

	Environmental Quality		Air Quality		Noise Situation		Greenspace	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age	1.01***	(1.01;1.02)	1.01**	(1.00;1.01)	1.01***	(1.00;1.01)	1.01***	(1.00;1.02)
Gender								
male	1.00		1.00		1.00		1.00	
female	1.80***	(1.49;2.18)	1.61	(1.33;1.95)	1.25***	(1.08;1.43)	1.56***	(1.24;1.97)
Education	0.67***	(0.62;0.72)	0.94*	(0.89;1.00)	0.98	(0.93;1.03)	0.85***	(0.78;0.93)
Income*	0.95***	(0.94;0.97)	0.99*	(0.97;1.00)	1.00	(0.99;1.00)	0.98***	(0.97;0.99)
Job Level	0.82***	(0.79;0.86)	0.92***	(0.88;0.97)	0.95***	(0.92;0.98)	0.92***	(0.87;0.97)
Migration Background								
no	1.00		1.00		1.00		1.00	
yes	0.51***	(0.35;0.73)	1.05	(0.78;1.41)	1.28**	(1.04;1.59)	0.71*	(0.48;1.06)
Urbanisation	0.96	(0.86;1.07)	1.66***	(1.48;1.86)	1.43***	(1.31;1.56)	1.35***	(1.17;1.55)
Type of Family								
couple w/o children	1.00		1.00		1.00		1.00	
couple w/ children	0.72***	(0.57;0.91)	0.73**	(0.57;0.94)	0.76***	(0.64;0.90)	0.77*	(0.58;1.02)
singles	1.24**	(1.00;1.55)	1.41***	(1.13;1.75)	1.19**	(1.01;1.40)	1.35**	(1.05;1.74)
single parents	1.25	(0.89;1.76)	1.34	(0.93;1.91)	0.98	(0.73;1.31)	1	(0.63;1.60)
other	0.94	(0.58;1.52)	0.73	(0.45;1.17)	0.98	(0.67;1.43)	0.97	(0.55;1.70)
Flats								
1 or 2	1.00		1.00		1.00		1.00	
3 or more	1.12	(0.93;1.34)	2.24***	(1.88;2.69)	1.72***	(1.98;7.67)	1.72***	(1.98;7.67)
Category of Flat								
cat. A	1.00		1.00		1.00		1.00	
cat. B or below	1.12	(0.75;1.68)	1.08	(0.71;1.65)	1.19	(0.87;1.64)	1.96***	(1.25;3.07)

n = 7,021 / *n = 3,749

*p<0.1; **p<0.05; ***p<0.01

Table B5: Determinants of Perceived Environmental Exposure (Ordinal Logistic Regression)

	Odour / Fumes		Dust / Soot		Noise		Heat	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age	1.00	(1.00;1.01)	1.01**	(1.00;1.01)	0.99***	(0.99;1.00)	0.99***	(0.99;1.00)
Gender								
male	1.00		1.00		1.00		1.00	
female	1.39**	(1.05;1.84)	1.17*	(0.98;1.39)	1.19**	(1.02;1.39)	1.26***	(1.13;141)
Education	0.95	(0.85;1.05)	1.04	(0.99;1.10)	1.14***	(1.08;1.21)	0.97	(0.93;1.02)
Income*	0.98***	(0.96;0.99)	0.99	(0.99;1.00)	0.99	(0.00;1.00)	0.99*	(0.99;1.00)
Job Level	0.92**	(0.86;0.98)	0.99	(0.96;1.04)	1.01	(0.97;1.04)	0.99	(0.96;1.02)
Migration Background								
no								
yes	1.52**	(1.03;2.25)	1.60***	(1.25;2.06)	1.27**	(1.01;1.61)	1.10	(0.92;1.31)
Urbanisation	1.45***	(1.23;1.72)	1.62***	(1.45;1.81)	1.76***	(1.60;1.94)	1.22***	(1.14;1.30)
Type of Family								
couple w/o children	1.00		1.00		1.00		1.00	
couple w/ children	0.72*	(0.51;1.03)	0.77**	(0.62;0.97)	0.93	(0.77;1.13)	1.08	(0.94;1.23)
singles	0.99	(0.72;1.35)	1.36***	(1.11;1.66)	1.11	(0.92;1.34)	1.24***	(1.07;1.43)
single parents	1.48	(0.87;2.52)	1.49**	(1.03;2.15)	1.25	(0.92;1.70)	1.60***	(1.25;2.05)
other	0.67	(0.35;1.30)	0.97	(0.60;1.56)	0.87	(0.57;1.31)	0.91	(0.66;1.25)
Flats								
1 or 2	1.00		1.00		1.00		1.00	
3 or more	1.60***	(1.23;2.10)	1.88***	(1.58;2.24)	2.18***	(1.86;2.55)	1.31***	(1.17;1.47)
Category of Flat								
cat. A	1.00		1.00		1.00		1.00	
cat. B or below	1.23	(0.59;2.58)	1.09	(0.75;1.59)	0.84	(0.60;1.17)	1.10	(0.82;1.46)

n = 7,021 / *n = 3,749

*p<0.1; **p<0.05; ***p<0.01

Appendix B.3: Air Pollution in Vienna and Tyrol

Table B6: Determinants of Perceived Exposure to Air Pollution in Vienna and Tyrol (Ordinal Logistic Regression)

	Vienna				Tyrol			
	Odour / Fumes		Dust / Soot		Odour / Fumes		Dust / Soot	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age	1.03*	(1.00;1.06)	0.99	(0.97;1.01)	1.02	(0.98;1.06)	1.01	(0.98;1.03)
Gender								
male	1.00		1.00		1.00		1.00	
female	0.95	(0.45;1.24)	1.37	(0.87;2.13)	0.54	(0.18;1.61)	2.26**	(1.07;4.81)
Education	0.82	(0.60;1.19)	0.89	(0.65;1.22)	0.84	(0.63;1.11)	0.73	(0.53;1.02)
Income	0.96**	(0.91;1.00)	1.00	(0.98;1.02)	0.92***	(0.87;0.97)	1.01	(0.98;1.05)
Job Level	1.36	(0.92;2.00)	1.01	(0.81;1.26)	1.12	(0.66;1.91)	1.02	(0.74;1.40)
Migration Background								
no	1.00		1.00		1.00		1.00	
yes	1.57	(0.65;3.80)	1.39	(0.83;2.33)	3.86***	(1.39;10.70)	1.61	(0.72;3.61)
Urbanisation	-	-	-	-	2.02**	(1.05;3.90)	1.62**	(1.05;2.49)
Type of Family								
couple w/o children	1.00		1.00		1.00		1.00	
couple w/ children	0.75	(0.22;2.55)	1.63	(0.78;3.40)	0.93	(0.29;2.97)	0.72	(0.35;1.49)
singles	1.57	(0.60;4.12)	3.14***	(1.69;5.93)	2.4	(0.72;8.01)	1.1	(0.47;2.61)
single parents	2.06	(0.56;7.55)	3.16***	(1.39;7.19)	1.82	(0.39;8.41)	0.82	(0.28;2.43)
other	2.57	(0.43;15.33)	1.63	(0.37;7.24)	0.27	(0.03;2.86)	1.47	(0.42;5.09)
Flats								
1 or 2	1.00		1.00		1.00		1.00	
3 or more	1.54	(0.34;6.90)	0.88	(0.41;1.88)	0.62	(0.25;1.52)	1.28	(0.69;2.37)
Category of Flat								
cat. A	1.00		1.00		1.00		1.00	
cat. B or below	1.14	(0.17;7.70)	0.61	(0.13;2.80)	1.21	(0.17;8.71)	0.34	(0.04;2.76)

nVienna = 514; nTyrol = 376

*p<0.1; **p<0.05; ***p<0.01

Appendix B.4: Noise Exposure in Vienna

Table B7: Determinants of Perceived Noise Exposure in Vienna (Ordinal Logistic Regression)

	OR	95% CI
Age	0.99	(0.98;1.01)
Gender		
male	1.00	
female	1.37	(0.93;2.03)
Education	0.99	(0.85;1.18)
Income	0.98*	(0.97;1.00)
Job Level	0.94	(0.76;1.15)
Migration Background		
no	1.00	
yes	0.55**	(0.32;0.94)
Type of Family		
couple w/o children	1.00	
couple w/ children	0.66	(0.37;1.15)
singles	1.20	(0.73;1.95)
single parents	0.75	(0.36;1.56)
other	0.64	(0.24;1.75)
Flats		
1 or 2	1.00	
3 or more	1.20	(0.61;2.33)
Category of Flat		
cat. A	1.00	
cat. B or below	1.26	(0.47;3.37)
n = 514	*p<0.1; **p<0.05; ***p<0.01	