JOHANNES KEPLER UNIVERSITY LINZ

DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

Converting variables between non-hierarchical regions: The case of Austrian municipalities and ZIP-code areas

by

Dieter PENNERSTORFER

Working Paper No. 2111
April 2021

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

dieter.pennerstorfer@jku.at

Converting variables between non-hierarchical regions: The case of Austrian municipalities and ZIP-code areas*

Dieter Pennerstorfer†

April 29, 2021

Abstract

This report provides a method to convert variables between two non-hierarchical regional levels of aggregation. This approach is applied to municipalities and ZIP-code areas in Austria for the period between 2012 and 2020. Based on detailed information on the spatial distribution of the population at the 250 m grid-cell level and on the spatial extent of both municipalities and ZIP-code areas, I calculate weights to convert variables from one regional entity to the other. This report also provides the data necessary for this data conversion.

1 Introduction

In regional economics, data are often available at different levels of spatial aggregation, which is particularly cumbersome if the relationship between these regional levels is not hierarchical. This is, for example, the case for municipalities and ZIP-code areas in Austria: Municipalities are sometimes congruent with ZIP-code areas, but frequently municipalities comprise a number of ZIP-code areas or the same ZIP code is used in multiple municipalities. Even more complicated, one ZIP code might be assigned to only a share of inhabitants of one municipality, but also to a share of residents of another one. In the end, more than half of the Austrian population cannot be precisely assigned to one municipality based on their ZIP codes.

In this report, I outline a simple way to convert data between two non-hierarchical regional levels. Additionally, I provide the necessary data to convert variables from the municipal to the ZIP-code level (and vice versa) for Austria for the period between 2012 and 2020. Based on detailed information on the spatial distribution of the population at the 250 m grid-cell level and on the spatial extent of both municipalities and ZIP-code areas, I calculate (and provide) weights to convert variables from one regional entity to the other. The conversion between these two regional levels is accurate, if the distribution of the variable of interest across space is proportional to the distribution of the residential population.

The remainder of the report is organized as follows: Section 2 outlines the methodology how these weights are calculated. Section 3 describes the data sources and discusses descriptive

*I thank Astrid Pennerstorfer for sharing the data for this project.

†Johannes Kepler University Linz, Altenberger Straße 69, Linz, Austria; e-mail: *dieter.pennerstorfer@jku.at*

statistics on the relationship between municipalities and ZIP-code areas, while data accessibility and instructions to use these data in Stata are provided in Section 4.

2 Methodology

To convert variables from the municipal to the ZIP-code level (and vice versa), I calculate weights based on the spatial distribution of the population. The intuition of this approach can be described by a simple example: Assume that a researcher knows the ZIP codes of all university graduates, but wants to conduct an analysis at the municipal level. If one ZIP code is used exclusively by one municipality, than all university graduates with this ZIP code can be assigned to that municipality. If this ZIP code, however, is utilized by multiple municipalities, it is less straightforward how to split the university graduates among the municipalities that use the respective ZIP code. I suggest using weights based on the spatial distribution of the population. If, for example, 60 % of all inhabitants of this ZIP-code area live in municipality A and 40 % in municipality B, then 60 % of the university graduates with this ZIP code should be assigned to municipality A and 40 % to community B. This approach thus relies on the assumption that the variable of interest (e.g. the number of university graduates) is distributed proportionally across space as the residential population, and is the more accurate the more similarly these two variables are spatially distributed.

To calculate these weights I utilize information on the distribution of the population at a high spatial resolution, the 250 m × 250 m grid-cell level. Each grid cell is assigned to exactly one municipality and to one ZIP code. pop_{imz} denotes the population in grid cell i, located in municipality m and in ZIP-code area z. C_m denotes the exhaustive set of grid cells located in municipality m (irrespective of their ZIP code), while C_z indicates the set of grid cells with ZIP code z. C_{mz} describes the set of cells located in both municipality m and ZIP-code area z. The population can be aggregated from the grid-cell level to the municipality ($pop_m = \sum_{i \in C_m} pop_{imz}$) and the ZIP-code level ($pop_z = \sum_{i \in C_z} pop_{imz}$). $pop_{mz} = \sum_{i \in C_{mz}} pop_{imz}$ describes the population living in municipality m with ZIP code z.

Converting data from ZIP-code areas to municipalities

To convert data from ZIP-code areas to municipalities, I assign a weight for each possible combination of municipalities and ZIP-code areas. The weight of ZIP-code area z for municipality m, $\omega_{mz}^{z \to m}$ is defined as:

$$\omega_{mz}^{z \to m} = \frac{pop_{mz}}{pop_z} = \frac{\sum_{i \in C_{mz}} pop_{imz}}{\sum_{i \in C_z} pop_{imz}} \tag{1}$$

The weight $\omega_{mz}^{z \to m} \in [0, 1]$ and a weight $\omega_{mz}^{z \to m} = 0$ indicates that nobody lives in both municipality m and ZIP code z (as $pop_{mz} = 0$ in this case), while $\omega_{mz}^{z \to m} = 1$ if all residents of ZIP-code area z live in municipality m. When the weights are aggregated over all municipalities, they sum up to unity for each ZIP code ($\sum_m \omega_{mz}^{z \to m} = 1$). To convert a variable x_z, measured in levels (e.g. the number of university graduates) and available at the ZIP-code level, to the municipal level, one has to calculate for each municipality m the weighted sum over all ZIP-code areas:

$$\tilde{x}_m = \sum_z \omega_{mz}^{z \to m} x_z, \tag{2}$$

with $\tilde{x}_m$ denoting the respective variable converted to the municipality level.

Converting data from municipalities to ZIP-code areas

The conversion of variables from municipalities to ZIP-code areas is carried out exactly the other way round. The weight of municipality m for ZIP-code area z, $\omega_{mz}^{m \to z}$ is defined as:

$$\omega_{mz}^{m \to z} = \frac{pop_{mz}}{pop_m} = \frac{\sum_{i \in C_{mz}} pop_{imz}}{\sum_{i \in C_m} pop_{imz}} \tag{3}$$

The weights $\omega_{mz}^{m \to z} \in [0, 1]$ and sum up to unity for each municipality ($\sum_z \omega_{mz}^{m \to z} = 1$). A weight $\omega_{mz}^{m \to z} = 0$ indicates that nobody lives in both municipality m and ZIP code z (as $pop_{mz} = 0$ in this case), and $\omega_{mz}^{m \to z} = 1$ if all residents of municipality m live in ZIP-code area z. To convert a variable x_m, available at the level of municipalities, to the ZIP-code level, one has to calculate for each ZIP code z the weighted sum of all municipalities:

$$\tilde{x}_z = \sum_m \omega_{mz}^{m \to z} x_m, \tag{4}$$

with $\tilde{x}_z$ denoting the respective variable converted to the ZIP-code level.

If variables are available as shares or rates, I suggest transforming the variables into levels in a first step, and applying the procedure outlined above to convert the information from one regional entity to the other.

3 Data sources and description

The weights are based on three data sources: First, information on the spatial distribution of the population is provided at the 250 m × 250 m grid cell level, collected by Statistics Austria in 2015. Statistics Austria places regional statistical grid units over the entire territory of Austria, and grid cells are independent of administrative boundaries. Each person is assigned to exactly one cell based on the postal address of his / her primary residence ('Hauptwohnsitz'). Second, maps of the geographic location of the Austrian municipalities are provided by Statistics Austria annually for the period between 2012 and 2020. According to the Nomenclature of Territorial Units for Statistics (NUTS) of the EU, municipalities correspond to the level of local administrative units (LAU).[1] Grid cells are assigned to municipalities based on the location of the centroids of the grid cells.[2] Therefore, the entire population of one grid cell is assigned to exactly one municipality, even if the respective grid cell is partly located in multiple municipalities. Third, data on the ZIP-code areas are provided by the company GeoMagis annually for the period between 2012 and 2020 at the grid-cell level. Again, grid cells are assigned to exactly one

[1] While the capital Vienna is only a single municipality (and also a single LAU), Vienna is included at the district level. Vienna is divided into 23 districts.

[2] I use the geographic information system QGIS (Version 2.18) to match the (centroids of the) grid cells with the municipalities.

ZIP code based on the cells' centroids. From the total population of 8,580,317 in Austria in 2015, virtually all inhabitants (8,579,409 or 99.99 %) could be assigned to exactly one municipality and one ZIP-code area every year.[3]

The merged dataset therefore includes information on the population (from 2015) at the 250 m grid-cell level, along with the corresponding annual information on the municipality and ZIP code where the cell is located. For each year, the data is aggregated at the municipality × ZIP-code level, and supplemented by the weights $\omega_{mz}^{z \to m}$ and $\omega_{mz}^{m \to z}$, as outlined above.

Descriptive statistics on the congruence between municipalities and ZIP-code areas are given in Table 1. While the number of municipalities and ZIP-code areas are roughly the same, only about 22 % of the population lives in a location where the corresponding borders of the municipality and the ZIP code are identical ('one-for-one' relationship). The weights $\omega_{mz}^{z \to m} = \omega_{mz}^{m \to z} = 1$ in these cases. More than half of the population lives in grid cells where ZIP-code area and municipality are almost identical ('almost one-for-one' relationship, with $\omega_{mz}^{z \to m} \geq 0.95 \ \wedge \ \omega_{mz}^{m \to z} \geq 0.95$). About 45 % of the population lives in municipalities that comprise the entire territory of one or more ZIP-code areas, and are thus hierarchically above ZIP codes. For municipalities comprising about 75 % of the population I find that municipalities are in an almost hierarchical relationship with ZIP codes. Similar figures are found for the other way around, namely when ZIP-code areas are (almost) hierarchically above municipalities.[4] This descriptive evidence suggests that using (almost) one-for-one or (almost) hierarchical relationships to convert variables from one regional level of aggregation to the other is a rather inaccurate (and probably time-consuming) way to utilize information that is available only at the 'wrong' regional level.

Both municipalities and ZIP-code areas change over time. Between 2012 and 2020 the number of municipalities declined substantially from 2,379 to 2,117, whereas the number of ZIP-code areas increased slightly from 2,194 to 2,221. Additionally to the change in the number of entities, the regional units' borders also vary over time. Figure 1 illustrates the share of regional units that changed their borders relative to 2013,[5] as well as the population affected by these changes. In 2020, for example, about 10 % (8 %) of the population live in a municipality (ZIP-code area) that is not identical to the respective regional unit in 2013. While the time series on municipalities is dominated by a reform on the municipal structure in Styria, which became effective in 2015 and reduced the number of municipalities in this state from 542 to 287, ZIP-code areas are found to adjust more continuously over time.

[3] A small number of grid cells could not be assigned to regional entities, because the cells' centroids lie beyond the Austrian borders.

[4] The number of (exact) one-for-one relationships are much smaller in 2012. I suspect that the ZIP-code data is less precise in this year, and the weights for 2012 have to be used cautiously. The inaccuracy is probably small, as the number of 'almost one-for-one' or 'almost hierarchical' relationships are similar to 2013.

[5] 2013 is used as the reference year due to less precise data for 2012, as discussed above.

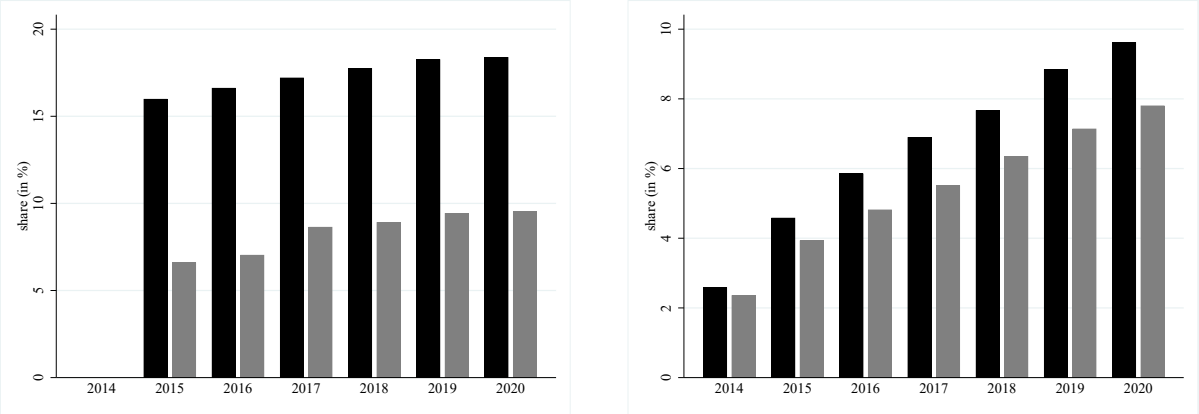
Table 1: Descriptive statistics on congruence between municipalities and ZIP-code areas

							Municipalities				ZIP codes			
	Municipalities	ZIP codes	One-for-one		Almost one-for-one		Hierarchical		Almost hierarchical		Hierarchical		Almost hierarchical	
Year	No	No	No	Pop	No	Pop	No	Pop	No	Pop	No	Pop	No	Pop
2012	2,379	2,194	381	842,586	930	4,317,736	834	2,352,313	1,349	6,097,025	884	2,580,192	1,395	5,851,940
				9.82 %		50.33 %		27.42 %		71.07 %		30.07 %		68.21 %
2013	2,376	2,203	631	1,789,887	1,016	4,644,054	1,054	3,787,765	1,411	6,316,979	1,120	3,769,739	1,426	6,008,154
				20.86 %		54.13 %		44.15 %		73.63 %		43.94 %		70.03 %
2014	2,376	2,213	654	1,844,527	1,043	4,708,435	1,074	3,859,955	1,435	6,373,849	1,136	3,799,140	1,444	6,043,094
				21.50 %		54.88 %		44.99 %		74.29 %		44.28 %		70.44 %
2015	2,124	2,215	652	1,857,971	1,043	4,768,899	1,094	3,942,478	1,474	6,603,260	1,053	3,638,752	1,360	5,889,416
				21.66 %		55.59 %		45.95 %		76.97 %		42.41 %		68.65 %
2016	2,122	2,218	663	1,880,682	1,057	4,798,759	1,102	3,944,036	1,484	6,625,843	1,061	3,651,239	1,369	5,903,897
				21.92 %		55.93 %		45.97 %		77.23 %		42.56 %		68.81 %
2017	2,122	2,217	670	1,900,446	1,069	4,832,833	1,108	3,962,775	1,489	6,641,200	1,068	3,655,224	1,379	5,931,202
				22.15 %		56.33 %		46.19 %		77.41 %		42.60 %		69.13 %
2018	2,120	2,217	673	1,929,756	1,076	4,869,604	1,108	3,987,416	1,493	6,672,958	1,073	3,670,491	1,384	5,947,063
				22.49 %		56.76 %		46.48 %		77.78 %		42.78 %		69.32 %
2019	2,118	2,217	676	1,918,926	1,085	4,903,240	1,113	3,995,780	1,498	6,687,166	1,078	3,678,324	1,392	5,977,637
				22.37 %		57.15 %		46.57 %		77.94 %		42.87 %		69.67 %
2020	2,117	2,221	684	1,933,685	1,096	4,934,773	1,120	4,014,888	1,508	6,720,500	1,085	3,689,852	1,398	5,980,821
				22.54 %		57.52 %		46.80 %		78.33 %		43.01 %		69.71 %

Notes: A municipalities m and and ZIP-code area z are defined to have a 'one-for-one' relationship if $\omega_{mz}^{z\to m} = \omega_{mz}^{m\to z} = 1$. They are categorized as having an 'almost one-for-one' relationship if $\omega_{mz}^{z\to m} \geq 0.95 \ \wedge \ \omega_{mz}^{m\to z} \geq 0.95$. A municipality m is classified as 'hierarchical' relative to ZIP-code areas if $\omega_{mz}^{z\to m} \in \{0,1\} \ \forall \ z$ and as 'almost hierarchical' if $\omega_{mz}^{z\to m} = 0 \ \vee \ \omega_{mz}^{z\to m} \in [0.95,1] \ \forall \ z$. Conversely, a ZIP-code area z is classified as 'hierarchical' relative to municipalities if $\omega_{mz}^{m\to z} \in \{0,1\} \ \forall \ m$ and as 'almost hierarchical' if $\omega_{mz}^{m\to z} = 0 \ \vee \ \omega_{mz}^{m\to z} \in [0.95,1] \ \forall \ m$. No denotes the number of regional entities and Pop indicates the population (both in absolute terms and relative to the entire population) that fall into the respective category.

Figure 1: Stability of regional entities over time

(a) Municiaplities (b) ZIP-code areas



Notes: Black bars denote the share of municipalities (left panel) and the share of ZIP-code areas (right panel) that have changed their borders relative to the year 2013. Gray bars indicate the share of population living in these regional entities.

4 Stata code and data availability

The necessary data to convert variables between these two regional entities provided below are aggregated at the municipality × ZIP-code level for each year from 2012 to 2020. These data are available in Stata format and named 'wp2111_Weights_Municipalities_ZIP_2012.dta' to 'wp2111_Weights_Municipalities_ZIP_2020.dta'. Combinations of municipalities and ZIP-code areas without common inhabitants (where $\omega_{mz}^{z \to m} = \omega_{mz}^{m \to z} = 0$) are left out for brevity. Relevant information on the data can be accessed by using the Stata-command `notes`. To convert a variable from the ZIP-code to the municipality level e.g. for the year 2015, the ID for the ZIP-code areas should be named 'PLZ_2015'. In the reverse case, the variable comprising the IDs of the municipalities should be named 'GKZ_2015' and should include the official ID ('Gemeindekennziffer—GKZ'), allocated by Statistics Austria.

I show an example to convert a variable from one regional level of aggregation to the other. Assume you have a variable on the number of university graduates at the level of ZIP codes in 2015, labeled 'Grad_zip'. The data 'Grad_2015_zip.dta' includes this variable, as well as a variable to identify the ZIP codes, named 'PLZ_2015':

```
use Grad_2015_zip.dta, clear
merge 1:m PLZ_2015 using wp2111_Weights_Municipalities_ZIP_2015.dta
notes
gen Grad_mun_tilde = Grad_zip * weight_z_to_m_2015
collapse (sum) Grad_mun_tilde, by(GKZ_2015)
```

The variable 'Grad_mun_tilde' denotes the number of university graduates, converted from the ZIP-code to the municipality level.

Contrariwise, assume that the information on the number of university graduates is available at the municipal level, labeled 'Grad_mun', and stored in 'Grad_2015_mun.dta', together with a variable on the municipality IDs named 'GKZ_2015':

```
use Grad_2015_mun.dta, clear
```

```
merge 1:m GKZ_2015 using wp2111_Weights_Municipalities_ZIP_2015.dta
notes
gen Grad_zip_tilde = Grad_mun * weight_m_to_z_2015
collapse (sum) Grad_zip_tilde, by(PLZ_2015)
```

The variable 'Grad_zip_tilde' denotes the number of university graduates converted from the municipal to the ZIP-code level.

Data access

Year	Resource
2012	wp2111_Weights_Municipalities_ZIP_2012.dta
2013	wp2111_Weights_Municipalities_ZIP_2013.dta
2014	wp2111_Weights_Municipalities_ZIP_2014.dta
2015	wp2111_Weights_Municipalities_ZIP_2015.dta
2016	wp2111_Weights_Municipalities_ZIP_2016.dta
2017	wp2111_Weights_Municipalities_ZIP_2017.dta
2018	wp2111_Weights_Municipalities_ZIP_2018.dta
2019	wp2111_Weights_Municipalities_ZIP_2019.dta
2020	wp2111_Weights_Municipalities_ZIP_2020.dta

If you use the data provided above, please cite this working paper:
Pennerstorfer, Dieter (2021). Converting variables between non-hierarchical regions: The case of Austrian municipalities and ZIP-code areas. JKU Working Paper No. 2111.