How does the geographical distribution of regional development intersect with the landscape of demographic challenges, and how does it influence the descisions of internal migrants?

Dr Becky Arnold and Prof. Leo van Wissen

June 22, 2025

Abstract

Regional development is a key consideration for internal migrants as they pursue favourable economic conditions and high standards of living. Regional demography is thus entwined with regional development. Deficiencies in both perpetuate each other; low regional development results in higher outmigration (particularly amongst younger cohorts) which increases regional vulnerability, lowering key development markers. In this project regional development is quantified in three dimensions (economic, social, and living environment) to produce regional development profiles at the NUTS 3 level in the EU 27 countries. From that a regional development typology is created, dividing regions into vulnerable, underdeveloped, developed, or leading categories. Likewise a demographic typology is developed which incorporates overall population change, natural change due to births and deaths, and net migration. These typologies are used to investigate the nature of vulnerable regions and how developmental and demographic challenges intersect. Finally, the degree to which different types of migrants (by age, sex, and citizenship) prioritise these different dimensions is modelled.

1 Introduction

Regional development is one of the EU's principle strategies for increasing equality and living standards internationally. With a budget of € 226.05 billion for 2021 - 2027, the stated goal of the European Regional Development Fund is to 'reduce economic, social and territorial disparities' [1]. These disparities are well established; 'vulnerable' or 'left behind' regions have been a matter of considerable public discussion, as have the related rural-urban, demographic, and political divides many countries are experiencing. With the societal and financial stakes so high it is vitally important that the landscape of regional development and demographic challenges within Europe be systematically assessed.

The way in which vulnerable regions are defined is not fixed, but there are undoubtedly common themes. These regions are typically those with low, often worsening, regional development, and are generally facing demographic challenges, notably ageing and population decline, most often via brain drain. These demographic challenges exacerbate the challenges the region faces regarding regional development. An ageing population usually means lower employment and more strain on services. However depopulation and brain drain make the region less and less able to meet those needs. These increased regional development challenges in turn exacerbate demographic challenges by making the region less attractive to young potential-stayers and to migrants.

In this work a regional development typology and demographic typology are presented. Subnational regions within Europe (NUTS 3 level) are analysed and categorised according to these two typologies. The geographic distribution of regions of different types is investigated. The role of urbanisation in both of these domains is also studied. The relationship between regional development and demographic challenges is closely examined, and finally the relationship between regional development and migration patterns is explored using internal migration data.

2 Methodology

Regional demography and development are complex, multifaceted concepts, and every region is unique. In order to investigate these in a systematic way, regional development data is gathered within Europe, from which a regional development typology is created. Likewise, a demographic typology is created from demographic data.

2.1 Regional development typology

Conceptions of regional development have historically focused on economic indicators such as GDP per capita, however this has evolved over time. In this work, making use of [3], regional development is divided into three dimensions: economic, social, and living environment, and data is gathered for each of them. This data is primarily drawn from Eurostat and the OECD, with some supplementary data from sources such as the Global Data Lab [4]. This data is as follows:

- Economic
  - Educational Level
  - Primary Sector Fraction
  - Secondary Sector Fraction (Industrial)
  - Secondary Sector Fraction (Manufacturing)
  - Secondary Sector Fraction (Construction)
  - Tertiary Sector
  - Total Unemployment
  - Regional Innovation Score
  - GDP per Capita
  - Economic Resilience
- Social
  - LGBT Acceptance
  - Ethnic and Racial Acceptance
  - Migrant Acceptance
  - Life Satisfaction
  - Homicide Rate
  - Assault Rate
  - Robbery Rate
  - Burglary Rate
  - Theft Rate
  - Subjective Safety
  - Labour Force Participation Rate Gender Difference
  - Youth NEET Rate
  - Total Long-Term Unemployment
  - Gender Development Index (Education)
  - Gender Development Index (Income)
- Living environment
  - Flooding (Rivers)
  - Flooding (Coastal)
  - Physicians Rate

- Nurses Rate
- Hospital Beds
- Air Pollution
- Motorway Infrastructure
- Digital Infrastructure
- Heating Degree Days
- Cooling Degree Days
- CO2 per Capita
- Heat Stress (average exposure)

The data is collected for 2010-2021. During this timespan the NUTS regional boundaries have gone through multiple iterations, and in this work only the NUTS 2021 boundaries are used. That means that regions that no longer exist, e.g. due to being merged into other regions, are excluded even if data was gathered in them at the time. Newly created regions are included as far into the past as data has been collected for them. For some indicators data have been 'back calculated' by national statistical institutes. E.g. a new region is coined in 2021, and the relevant national statistical agency calculates what the unemployment rate was within those boundaries in previous years. As a result in some cases it is possible to produce regional development profiles for regions in years prior to their official creation.

Despite extensive work, and the deliberate bias of selected indicators towards those with high coverage, the regional characteristics database is far from complete, especially at the NUTS 3 level we wish to build regional profiles of. Conventional statistical methods and machine learning are considered as methods to model the missing data and improve the completeness of the data set. Both are rejected because it is observed that the incompleteness of the data set is highly inhomogeneous. Data for a given indicator is often not collected in every year, or in every country. Even in countries where the data is collected, different countries often collect the data on different NUTS levels (e.g. some collecting that data at NUTS 2 and others at NUTS 3). A country also may not collect the data in all the regions it contains, even at their chosen level. This inhomogeneity makes producing a meaningful training data set for machine learning or robust classical modelling extraordinarily challenging, therefore both approaches are determined to be non-viable (at least in the short term).

One method that is used to increase the completeness of the data set is disaggregation. That is using data that is available at the NUTS 2 level to estimate what the data would be at the NUTS 3 level had it been collected. This is currently done in the simplest possible way, by assuming the value recorded in a given NUTS 2 region would be the same as in all the NUTS 3 regions it contains. For example, if educational level (percentage of the population aged 25-64 with tertiary educational attainment) is X % in a NUTS 2 region, we assume it is also X % in the relevant NUTS 3 regions.

Using the collected data, overall quality indices are calculated for each region for each year from 2010 to 2021 in each of the three dimensions (economic, social, living environment). The considerable sparsity and inhomogeneity of the data set presents a barrier to this, and so the method we have designed for this conversion is deliberately constructed to be resistant to that data weakness. The method for this will be described via a simplified example in which the economic dimension contains only two indicators, educational level and GDP per capita. To further simplify this example we consider data for only one year, and assume that there are only five regions, A, B, C, D, and E.

For the purposes of this explanation consider the example data set shown in Table 1.

Step 1: For each property rank the regions from best to worst. Note that if one of the properties was undesirable, e.g. unemployment rate, this ranking would be done in reverse order with the smallest first and largest last. See Table 2 to see this reflected in the example data set.

Step 2: Convert those rankings to fractions normalised by the number of valid datapoints for each property. In the example case educational level has three valid entries, so ranking 1, 2, 3, becomes 1, 0.5, 0. In contrast GDP per capita has four valid entries, so ranks 1, 2, 3, 4 become 1, 0.66, 0.33, 0. The consequences of this step for the example data set can be seen in Table 3.

Step 3: For each region calculate the average of all its properties now they have been converted to fractions. This is the index, and for the example case can be seen in the last column of Table 3.

Table 1: An example of a data set with five regions, and two indicators, for which there has been incomplete and inhomogeneous data collection.

Region	Educational Level	GDP per Capita
A	72	40000
B	15	No data
C	No data	10000
D	No data	25000
E	43	5000

Table 2: The example data set where the values in each column have been replaced with ranks from best to worst.

Region	Educational Level	GDP per Capita
A	1	1
B	3	No data
C	No data	3
D	No data	2
E	2	4

Step 4: Now this process has been used on the economic indicators to produce and economic index it is repeated for the social and living environment indicators to produce indices in each of those dimensions.

The resistance of this method to missing data is such that as long as our selection of indicators ensures every region has data for at least one property in each dimension then an index can be calculated. The overall regional development score of a region is its average score across these three dimensions. By splitting these overall scores into quartiles regions are classified as one of four types: vulnerable, underdeveloped, developed, or leading.

2.2 Demographic typology

Data is gathered from Eurostat on three demographic dimensions:

- Population change: Is the population of this region increasing or decreasing?
- Natural change: Are there more deaths than births (natural decline), or more births than deaths (natural increase)?
- Migratory change: Is outmigration higher than inmigration, or the other way around?

This results in $2 \times 2 \times 2 = 8$ possible demographic types. In practice almost all regions fall within four types, so those four form the basis of this paper's demographic typology, plus one 'other' category. The types are described below:

- Growing by births and migration: The population of this region is growing. Births are higher than deaths, causing a positive natural increase. Additionally, more people move to this region than away from it, so net migration is also positive.

Table 3: The example data set where the ranks in each column have been replaced with fractions, and the average fraction in shown in the last column.

Region	Educational Level	GDP per Capita	Economic Index
A	1	1	1
B	0	No data	0
C	No data	0.33	0.33
D	No data	0.66	0.66
E	0.5	0	0.25

- Growing through migration: The population of this region is growing. More people die in this region than are born, resulting in a natural decrease, however because more people move to this region than away from it (i.e. net migration is positive) the natural decrease is overcome.
- Declining despite migration: The population of this region is declining. More people die in this region than are born, resulting in a natural decrease. Net migration is positive, i.e. more people move to this region than away from it, but this effect is not strong enough to overcome the natural decrease, so the population as a whole declines.
- Declining by births and migration: The population of this region is declining. More people die in this region than are born, resulting in a natural decrease. Additionally, more people move away from this region than to it, so net migration is also negative.
- Other: All other demographic types.

2.3 Evaluating migrant priorities

In the last part of this paper we examine how migrants prioritise the three dimensions of regional development (economic, social, living environment) when making their migration decisions. This is done by modelling the relationship between regional development and internal migration.

In this model a region i has the economic, social-cultural, and living environment properties E_i, S_i, L_i. The vector (E_i, S_i, L_i) is the region's development profile, $\boldsymbol{D_i}$. The priority people of type k place on the economic, social and living environment dimensions is ρ_{Ek}, ρ_{Sk}, and ρ_{Lk}. A person type, k may be e.g. 50-54 year old females. The total priority of each person type sums to 1, and is distributed between the three dimensions, i.e.:

$$\sum_{j=E,S,L} \rho_{jk} = 1. \tag{1}$$

The gross attractiveness of a region i to people of type k is G_{ik}. It is defined as:

$$G_{ik} = \boldsymbol{D_i} \cdot \boldsymbol{\rho_k} \tag{2}$$

We normalise gross attractiveness to produce the main attractiveness parameter A_{ik} which defines the probability a person of type k will select region i.

$$A_{ik} = \frac{G_{ik}}{\sum_i G_{ik}} \tag{3}$$

so

$$\sum_i A_{ik} = 1. \tag{4}$$

The number of migrants that select destination i is modelled as

$$N_{ik} = N_k A_{ik}, \tag{5}$$

where N_k is the number of migrants of type k in the country.

To model migrant's priorities (i.e. ρ_{Ek}, ρ_{Sk}, and ρ_{Lk}) internal migration data from the Norwegian population register is gathered. Machine learning techniques, making use of Markov chain Monte Carlo (MCMC) algorithms, are used to compare this internal migration data to the regional economic, social, and living environment scores described in section 2.1. The MCMC methodology uses random walkers to explore parameter space and identify minima, making it well suited for studying multidimensional non-linear functions. This methodology returns the sets of priorities that best reproduce the observed internal migrations given the region's development profiles.

3 Results

3.1 Where are vulnerable regions?

To understand the nature of vulnerable regions we first look at where they are located geographically. Fig. 1 shows a map of the regions colour coded by their development type in the most recent data (2021).

As one would expect, eastern Europe and southern Europe have significantly lower development scores than the west and north.

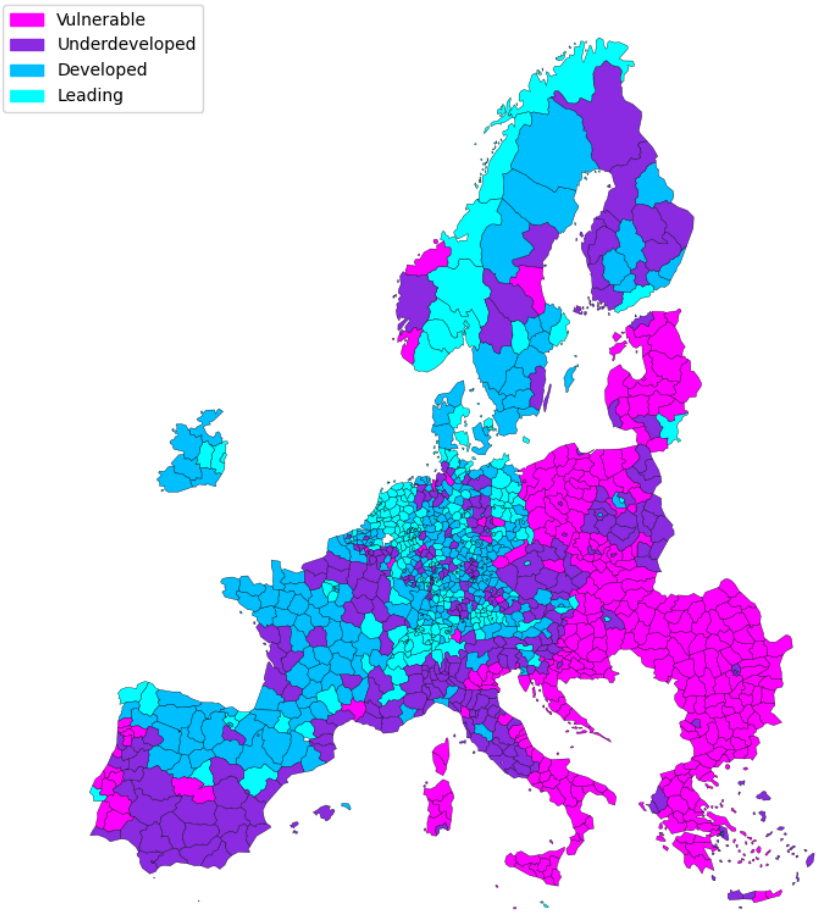

Figure 1: The geographic distribution of regional development.

Underlying these supra-national trends is significant inequality on the national scale for many countries. The north-south divide in both Spain and Italy is particularly apparent[1]. Urbanisation also clearly plays a role; major urban centres are often visible as leading regions surrounded by less developed regions. The impact of the closely-connected property, population density, is also self-evident. The borders for NUTS regions are chosen based partly on population, with the goal that all NUTS regions of the same level should have similar populations to improve comparability. Smaller NUTS regions therefore have higher average population density than larger ones. In Fig. 1 it is clear that more developed regions are generally smaller than less developed regions. This trend is more starkly evident in Fig. 2, in which the natural logarithm of population density is plotted against development score. Pearson's P test finds that these parameters are significantly correlated, with a value of 0.25 and a p-value of < 0.0001.

The geographic distribution of different migration typologies is shown in Fig. 3. The east and south of Europe are dominated by regions with declining populations. In some countries the decline of depopulating regions is being partially mitigated by positive net migration. However in some counties, notably Spain, it is more common for depopulating regions to be suffering from both negative natural change and negative net migration.

[1]One false positive of national inequality is visible in Norway, which contains two regions that are found to be vulnerable and one which is found to be underdeveloped. This is mostly an artifact of data scarcity in those three regions due to changing NUTS boundaries; in the living environment dimension only CO_2 per captia and number of hospital beds per 1000 people could be obtained. Largely due to its geographic nature and its climate Norway performs amongst the worst in the dataset for the former of these, resulting in an overwhelmingly low living environment score, which badly impacts the overall score of these regions.

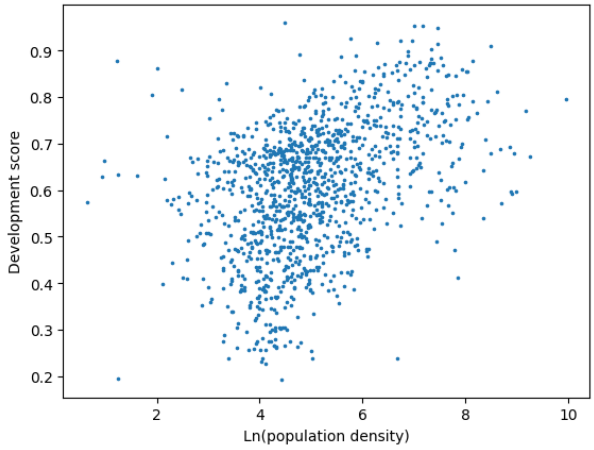

Figure 2: The natural logarithm of population density plotted against overall regional development score.

As is found for regional development, inequality on the national scale is obvious despite supranational trends. The unevenly distributed national and international landscape of demographic challenges increases the complexity for potential policies to address these challenges.

3.2 What is the relationship between urbanisation and vulnerability?

The classical picture of a stereotypical vulnerable or "left behind" region is usually rural. In this section we examine the degree to which that picture (using this text's definition of vulnerability) is and isn't accurate.

Data on NUTS 3 region's degrees of urbanisation are drawn from the OECD's work on functional urban areas (FUAs) [2]. That work classifies regions as one of five types: remote non-metropolitan regions (NMR-R), non-metropolitan regions with access to a small/medium city NMR-S, non-metropolitan regions with access to a metropolitan region (NMR-M), metropolitan regions (MR-M) and large metropolitan regions (MR-L). For readability these are referred to as very rural, somewhat rural, intermediate, somewhat urban and very urban respectively in this text.

In Fig. 4 the relationship between urbanisation and regional development status is shown. Somewhat rural regions are the most likely to be vulnerable, whereas very rural regions are only the second most likely. This implies the relationship between urbanisation and low regional development is not so simple as more rural = less developed. Further to that point there are comparable fractions of underdeveloped regions in every urbanisation category. Developed regions exist predominantly in intermediate and somewhat urban regions, but are well represented at the other three urbanisation levels too. The most clear trend is for leading regions. Very urban regions are more likely to also be leading than any other development category. The other leading regions are primarily found in somewhat urban regions and, to a lesser degree, intermediate regions.

These observed trends imperfectly reflect the common picture or how urbanisation and regional development intersect. While it is true that in general rural regions are less developed than urban ones, the full reality is much more complex. Notably, there are a significant number of rural regions that are developed or even leading. These regions provide interesting potential case studies as policy makers search for solutions to rural decline. Investigating regions that have found ways to thrive may yield valuable insights into what works and what doesn't.

The relationship between urbanisation and demographic trends is now discussed. Fig. 5 presents how the different demographic types are split between different levels of urbanisation. The biggest breakpoint is between very and somewhat rural regions on the one hand (which are more likely to have declining populations than increasing ones), and intermediate, somewhat urban, and very urban regions on the other hand (which are more likely to have increasing populations than declining ones). For the second type the story is relatively simple: more urban regions are less likely to face depopulation. The proportion of regions that are growing in all respects increases with urbanisation, and the proportion of declining

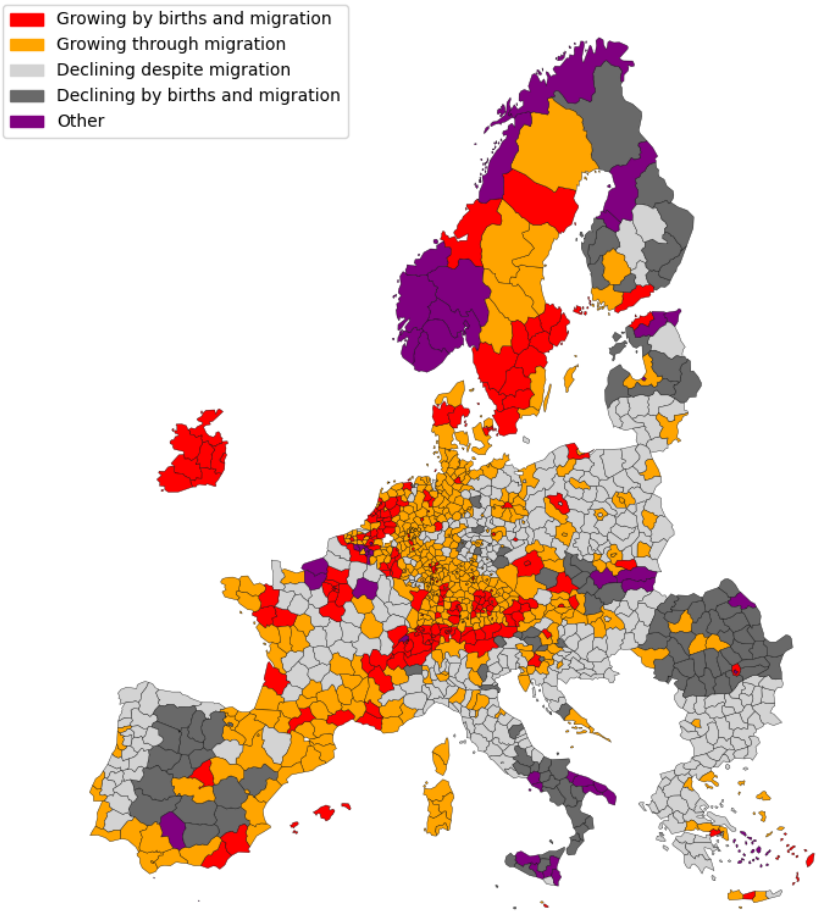

Figure 3: The geographic distribution of demographic typologies.

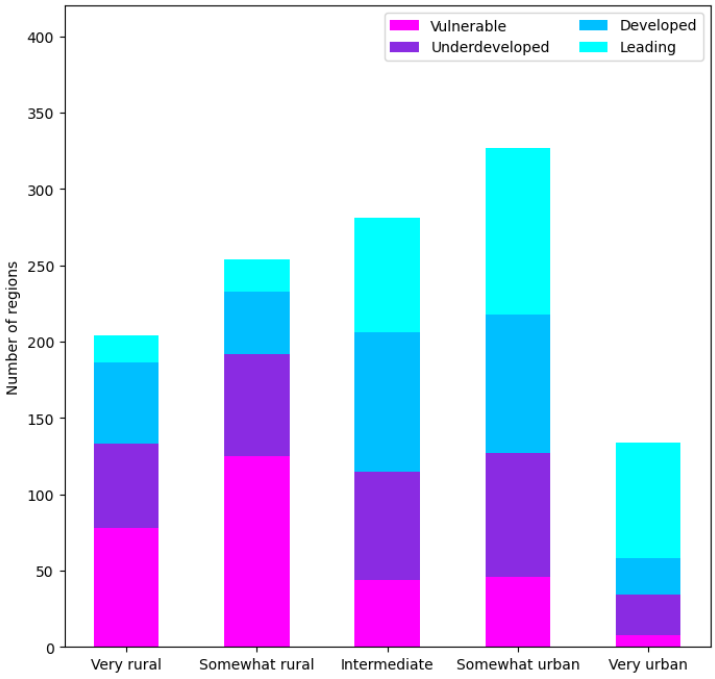

Figure 4: The relationship between development status and urbanisation.

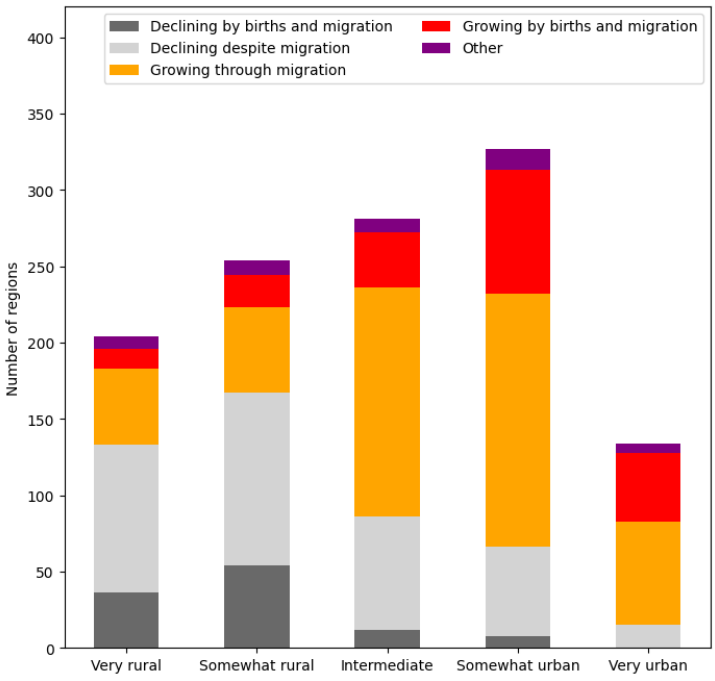

Figure 5: The relationship between demographic typology and urbanisation.

regions decreases with urbanisation.

The proportion of regions in different demographic categories is broadly similar in very rural compared to somewhat rural regions. It is noteworthy that despite the majority of rural regions having declining populations net migration is mostly positive even in these regions. This implies the problem of rural depopulation is not solely, or even mainly, that these regions are unattractive. Instead this implies that it is a purer demographic problem of inhomogeneous population pyramids; the population of rural regions is older than that of urban regions. However, migration choices are not random, they are effected by a variety of demographic factors, most prominently age. Most rural regions have net inmigration, so they are attractive, but attractive to *who*? As was the case for regional development there is a small but significant fraction of rural regions that are thriving. By examining these regions, both through a demographic and regional development lens, insight may be wrought into the particular aspects of regional development most suited to even out the attractiveness of rural regions between age groups.

3.2.1 How do demographic challenges and regional development intersect?

Here the relationship between vulnerability in terms of regional development and demographic challenges is examined more directly. Fig. 6 presents how these different typologies relate to each other, and shows broadly a familiar story. The less developed a region is the more likely it is to be suffering from depopulation, the more developed a region is the more likely it is to be increasing in population. The extremely strong link between these two dimensions highlights how the vicious spiral of poor development resulting in population decline (which itself harms regional development) self-perpetuates.

3.3 What do internal migrants prioritise when selecting destinations?

As noted previously a critical question at the intersection of regional development and migration is what aspects of regional development different migrants prioritise when deciding whether or not to move to a region or not. This in investigated in this section by comparing the regional development data previously discussed and internal migration data within Norway.

The Norwegian population register is used to gather this data on internal migrations from 2010-2021 (the same timeframe as the regional development data is collected in). MCMC fitting is used

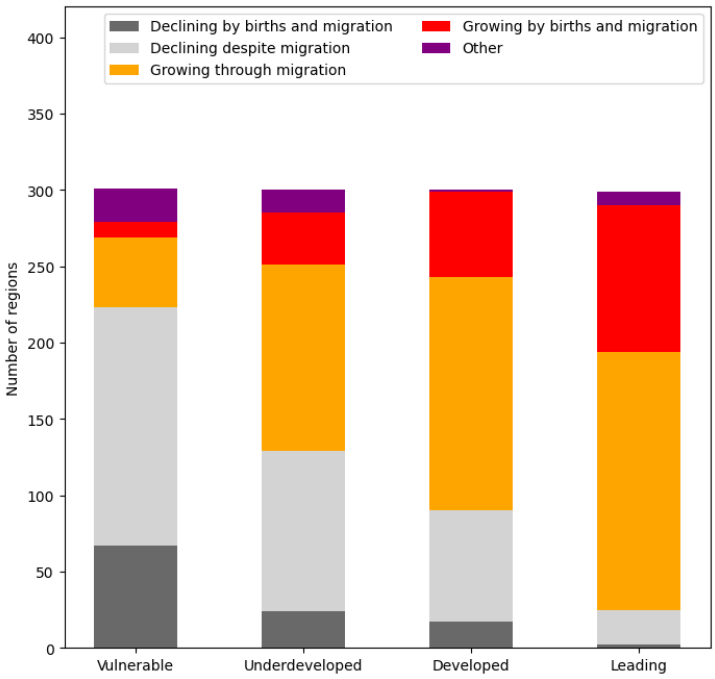

Figure 6: The relationship between demographic typology and development typology.

in conjunction with the model described in section 2.3 to estimate the ρ_{Ek}, ρ_{Sk}, and ρ_{Lk} which best reproduce N_{ik} given the known E_i, S_i, and L_i scores. The results for the priority given to the economic dimension for men (blue) and women (red) as a function of age are shown in Fig. 7. Likewise, the priority given to the social dimension is shown in Fig. 8, and living environment in Fig. 9.

Far and away, the most notable feature of these plots is the major departure in priorities observable for people from the age of twenty up until their early/mid thirties. During this life stage economic priorities completely dominate all others. It appears this stage ends slightly earlier for women than men; women in their late twenties show a rapid decrease in economic concerns, while the economic priority of men remains high. At ages above thirty however there is a complete collapse for both sexes in economic prioritisation. This may reflect a life course arc of career building and establishment at ages 20-30, followed by milestones such as the formation of long term relationships, parenthood, and home ownership. This change in life stages explains the rapid shift in priorities. People looking to "settle down" who have already achieved a measure of economic stability would reasonably prioritise finding a pleasant place to live for the long term.

The priorities given to the social dimension (Fig. 8) shows that the social dimension is prioritised almost exclusively at ages below five years. In reality toddlers do not make migration decisions, so this datapoint most accurately reflects the migration decisions of new parents. These individuals place almost no weight on either the economic or living environment dimensions. This means that regions wishing to mitigate depopulation by attracting young families would do well to focus on those policies that most directly improve the social dimension of regional development. This dimension is mostly concerned with safety and equality.

It is important to consider that the sample these priorities are drawn from: they are calculated based on net migration, i.e. movers not stayers. This underlines an important distinction: these are the estimated priorities of people of a given age and sex *who have made the decision to move*, not the priorities of *all* individuals of a given age and sex. It may be that people that prioritise e.g. the economic dimension at all ages move to economic powerhouses in their 20s and then are unlikely to leave regardless of life course events. From a regional development standpoint, restricting the sample to those individuals that are 'up for grabs' by using net migration as the input data is helpful. This allows valuable resources to be directed to policies that are most productive for combating demographic challenges via migration.

At ages above five and below twenty people (or more accurately the parents of these people) still place huge importance on the social dimension, around two thirds, but it no longer totally dominates.

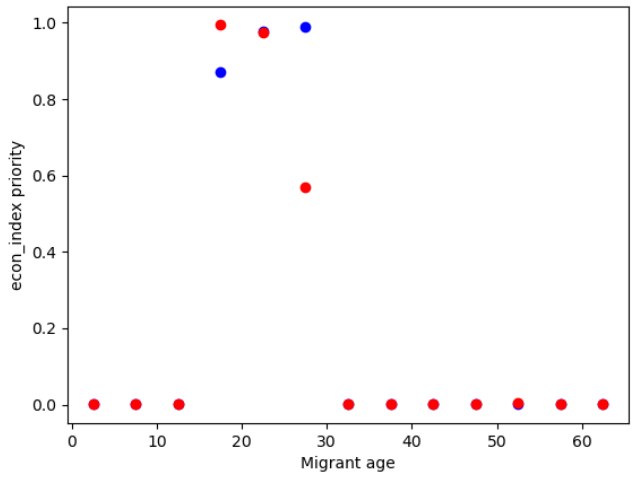

Figure 7: The degree to which Norwegians of different sexes (blue for men, red for women) prioritise the economic dimension as a function of age.

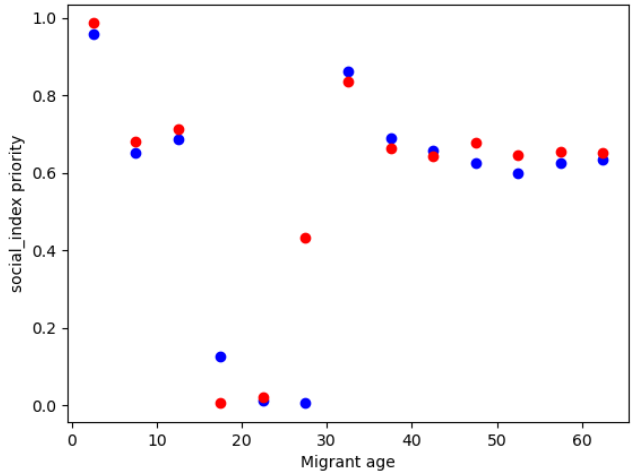

Figure 8: The degree to which Norwegians of different sexes (blue for men, red for women) prioritise the social dimension as a function of age.

The majority of the rest of priority is given to the living environment dimension. During the ages 20-30 both the social and living environment priorities collapse in place of economic prioritisation, which has already been discussed, As noted then, the economic dominance declines slightly earlier for women than men. This is mirrored in an increase in social priorities for women but not men at that age, but there is no corresponding increase in living environment priority. This gender and timing difference may be explained by the fact women are more commonly younger than their male romantic partners, and these ages are often when couples have children. A couple having a child and mutually prioritising the social dimension in their contemporaneous migration would explain the earlier increase in women prioritising the social but not living environment, given what this paper has already determined about the priorities of new parents.

At ages above 35 the priorities of men and women become much more similar. Social prioritisation drops from a small peak of above 0.8 to stabilise a little above 0.6. Priority given to the living environment dimension increases until an age of around fifty, after which it stabilises. Nevertheless, living environment priority never goes above 0.4.

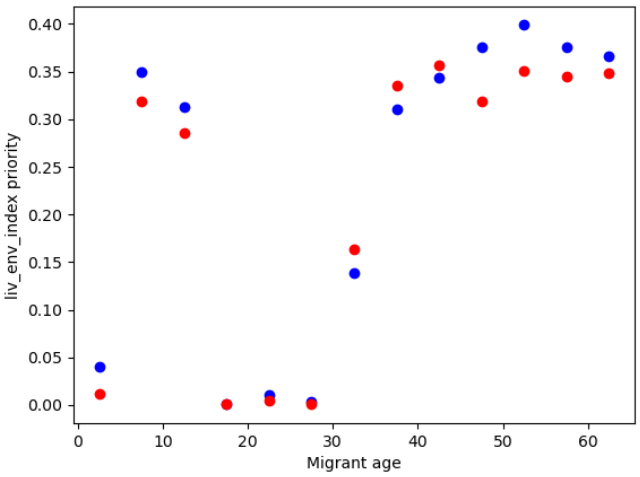

Figure 9: The degree to which Norwegians of different sexes (blue for men, red for women) prioritise the living environment dimension as a function of age.

4 Discussion and future work

The landscapes of regional development and demographic challenges are complex. On the whole, the classical picture of 'left behind regions' (low development, depopulating, rural, low population density) is supported, but this paper shows that there is also a much more complex picture underneath these broad strokes. We find regions that are rural and thriving, and conversely very urban regions with poor development. Likewise there are rural regions with growing populations and net inmigration, and there are urban regions facing demographic challenges. The strong relationship between low development and demographic challenges is upheld, but even in this there are exceptions. These regions that defy the odds are worthy of further study, so that their successes and failures can be understood and potentially replicated/avoided.

The ways in which migrants of different ages and sexes prioritise the different dimensions of regional development when making migration decisions are also examined. People in their twenties overwhelmingly prioritise the economic dimension. New parents overwhelmingly prioritise the social dimension. At ages above 35 around two thirds of weight is given to the social dimension, and around one third to the living environment dimension.

These priorities are modelled on Norwegian register data. This data has extremely high accuracy, but only covers one country. It is possible that people in different countries prioritise different things when making their migration decisions. To investigate whether or not this is the case internal migration data will be obtained from other countries such as Sweden, the Netherlands, and Spain. The migration priorities of people in those counties will be modelled and compared to those in Norway. These results will be used to create a more comprehensive picture of migrant priorities. The implications of those priorities will inform effective regional development policy to combat demographic challenges via migration.

References

[1] European Commission. URL: `https://ec.europa.eu/regional_policy/funding/erdf_en`.

[2] Milenko Fadic et al. *Classifying small (TL3) regions based on metropolitan population, low density and remoteness*. Tech. rep. June 2019.

[3] Andy Pike, Andres Rodriguez-Pose, and John Tomaney. *Local and Regional Development*. 2nd ed. London, England: Routledge, July 2016.

[4] Jeroen Smits and Iñaki Permanyer. "The Subnational Human Development Database". In: *Scientific Data* 6.1 (Mar. 2019), p. 190038. ISSN: 2052-4463. DOI: 10.1038/sdata.2019.38. URL: https://doi.org/10.1038/sdata.2019.38.