Sub-national projections of migration, population and economy: The Demography-Economy-Land Use interaction (DELi) model

Socioeconomic and demographic projections are essential tools for policy-making. While they are typically available only at coarse geographical aggregations, disaggregated projections at finer geographical resolution are required to assess future impacts of scenarios on a variety of issues, such as urbanization, internal migration, climate, environment, and natural hazards. The most simple calculation rule for downscaling projections keeps observed regional shares constant over time. However, such a simple approach is incapable of projecting regional variability. It can solely replicate the status quo of a given observed moment in time and ignores ongoing or expected processes of divergence or convergence of regions. Furthermore, it does not endogenize any dynamic interaction between demography and regional economy.

To overcome this limitation, we develop a novel regionalization model that regionalizes long-term, country-level socioeconomic and demographic projections in an integrated manner. This Demography-Economy-Land Use interaction (DELi) model builds on a set of linked equations that integrate assumptions regarding future regional growth. It provides sub-national estimates at NUTS3 level of net migration, employment, GDP and land use for all 27 EU Member States until the year 2100. The model estimates these elements dynamically and recursively, while ensuring consistency with any set of given socioeconomic and demographic projections at the country level. The novel approach incorporates functional causality between economic factors and human migration based on observed strong empirical associations between sub-national net migration rates and GDP per capita levels.

The projections can be generated based on a variety of assumptions and scenarios, which can be switched on and off, so that for the same exogenous country level projections, different regional distribution of population, GDP and land use can be obtained. Examples of these assumptions and scenarios relate to demography and economy (e.g., regional economic growth might increase depending on changes in education level or quality of institutions).

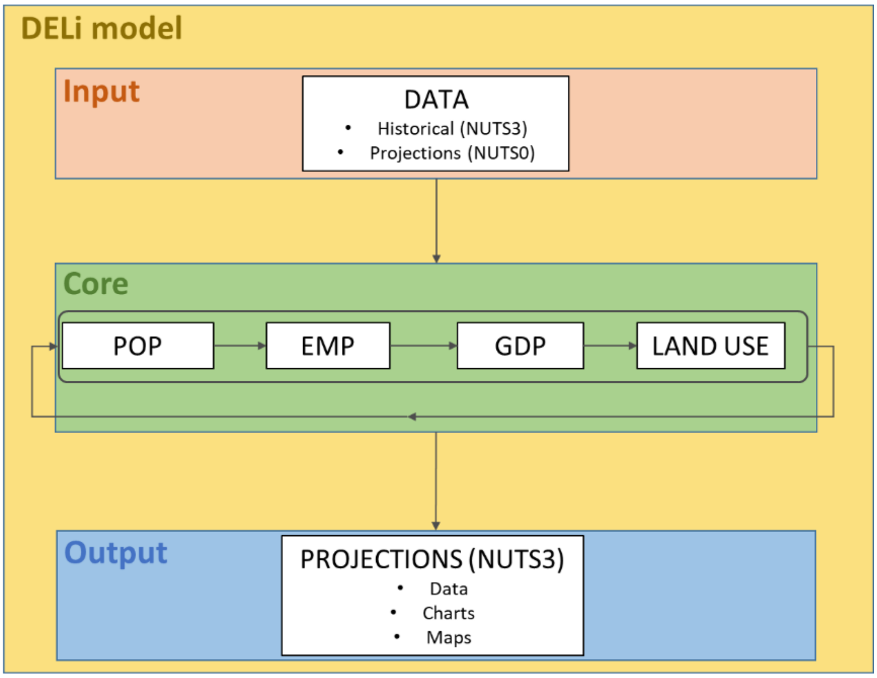

Figure 1. Demography-Economy-Land Use interaction (DELi) model layers

The DELi model is organized in three layers (see Figure 1). The *input layer* harmonizes historical data at the NUTS3 level and prepares them for the core layer. The *core layer* generates projections of population, employment, GDP and land use. At every time step, net migration, population, employment, GDP and land use are sequentially estimated, using values from previous steps in the sequence and values from previous time-steps (e.g. for estimates of population, only values until time t are available, for employment and GDP at time t+1, also values of population at time t+1 are available). The *output layer* provides tables, charts and maps from the output of the core layer.

The core layer is the crucial element of the model and generates new sub-national projections for every time-step of the projection period. Within each time-step, population, employment, GDP and land use are estimated sequentially and recursively based on the most recent (historical or projected) available data. These components are then rescaled based on simple rescaling functions to ensure that the sub-national projections are compatible with country-level projections.

Population is the **first** variable generated at every time-step. Age-specific population is endogenized by accounting for the effect of economic factors on net migration at sub-national level. The endogenous population projections are calculated based on age- and sex-specific fertility and mortality trends at NUTS3 level as provided by EUROSTAT, while net migration is estimated econometrically. This means that simply one of the three components of population change is endogenized. Information on fertility and mortality are taken as exogenous components from the official EUROSTAT population projections, whereas the migration patterns of individuals between regions become endogenous. In order to keep the analysis trackable and simple, the following parsimonious specification is chosen for the estimation:

$$nmr_r^a = \beta_0 + \beta_1 GDPpc_r + FE + \delta_{Capital} + \mu_{Metro} + \epsilon_r, \qquad (1)$$

where a describes the five-year age group, r denotes the region at NUTS3 level, nmr stands for the net migration rate and $GDPpc$ is the GDP per capita. In addition the specification includes country (FE), capital ($\delta_{Capital}$) and metro area (μ_{Metro}) fixed effects.

The coefficients of the econometric estimation of net migration and the GDP per capita projected by the model in a subsequent step are used to re-estimate the net migration rates for each five-year age group and NUTS3 region for the projection period. This generates a loop between the demographic and economic modules of the model.

Employment is the **second** variable generated at every time-step. It is calculated based on working age population and employment rates:

$$Emp_{r,t} = Pop_{wa,r,t} * ER_{r,t}, \qquad (2)$$

where t describes the time interval, r denotes the region at NUTS3 level, Emp describes employment, Pop_{wa} stands for the working age population (20-64 years of age) and ER is the employment rate. Historical levels of ER are available at the NUTS2 level. The ER is assumed to be the same for all NUTS3 regions belonging to the same mother NUTS2 region. Projections of ER are based on the last observed historical value and projected based on the estimated growth of ER at the NUTS0 level.

GDP is the **third** variable generated at every time-step. The regional GDP per capita is estimated based on an econometric model, where geographical factors, human resources and agglomeration economy determine GDP per capita growth. The econometric model relies on an augmented growth equation following Barro et al. (1991). It estimates the impact of factors affecting average GDP per capita growth, capturing the so-called beta convergence (i.e. regions with initially lower GDP are expected to record a faster growth) and additional idiosyncratic factors explaining the heterogeneity of GDP per capita growth across regions. The following equation describes the model:

$$\frac{1}{x} log\left(\frac{GDPpc_{r,t+x}}{GDPpc_{r,t}}\right) = \beta_1 \log(GDPpc_{r,t}) + \beta_2\, log\left(GDPpc_{r,t}\right)^2 + \beta_3 Edu_{r,t} +$$

$$\beta_4 EQI_t + \beta_5 Agg_{r,t} + \beta_6 OADR_{r,t} + \beta_7 Pop_{wa,r,t} + FE + \mu_{Metro} + \epsilon_{r,t+x}\,, \quad (3)$$

where t describes the time interval, r denotes the region at NUTS3 level and $GDPpc$ is the GDP per capita. The set of covariates includes the share of population with tertiary education (Edu), the European Quality of Government Index (EQI), agglomeration (Agg), the old age dependency ratio ($OADR$) and the share of working age population (Pop_{wa}). In addition, the specification includes country (FE) and metro area (μ_{Metro}) fixed effects. The growth rate of GDP per capita for each region and year is applied to region-specific GDP per capita of the previous year. Future values of the regressors are updated depending on changes in population and land use.

Finally, demand for land use is the **fourth** variable generated at every time-step by the model. Different categories of land use can be classified according to the CORINE Land Cover nomenclature, including artificial surfaces, agricultural areas, forest and seminatural areas, wetlands and water bodies. Changes of two sub-categories of the artificial surfaces are modelled, namely the urban residential built-up and built-up for industry, commerce and services. Land use can be completely saturated in certain regions, implying that no land for development is available. For such regions, a constant land use is maintained over time. For the other regions, the two types of built-up are estimated econometrically.

Future demand of urban residential built-up per capita is estimated based on a refined specification of the model presented in Schiavina et al. (2022):

$$URpc_{r,t+1} = \beta_0 + \beta_1 URpc_{r,t} + \beta_4 \log(ALDpc) + \beta_4 \log(ALDpc)^2 + \delta_{Urb} + \delta_{Rur} + \epsilon_{r,t+1}, \quad (4)$$

where t describes the time interval, r denotes the region at NUTS3 level, $URpc$ stands for urban residential built-up per capita and $ALDpc$ is the available land for development per capita. In addition, the specification includes fixed effects for urban (δ_{Urb}) and rural (δ_{Rur}) areas.

The estimation of the built-up for industry, commerce and services follows a similar approach to that of urban residential built-up, but uses employment instead of population as normalization factor. The expected built-up per employee is obtained by a simple autoregressive model:

$$ICSpe_{r,t+1} = \beta_0 + \beta_1 ICSpe_{r,t} + \epsilon_{r,t+1}, \quad (5)$$

where t describes the time interval, r denotes the region at NUTS3 level and $ICSpe$ stands for the built-up for industry, commerce and services per employee.

In a final step, the available land for development (ALD) in each region can be calculated at every time-step by adding the difference between new and old built up, for both urban residential built-up and built-up for industry, commerce and services:

$$ALD_{r,t+1} = ALD_{r,t} + \left(UR_{r,t+1} - UR_{r,t}\right) + \left(ICS_{r,t+1} - ICS_{r,t}\right) \quad (6)$$

After the estimation of land use for period t is completed, the core layer continues the loop by estimating population, employment, GDP and land use for period t+1. The loop ends when reaching the last year of projections. This provides disaggregated sub-national estimates at NUTS3 level of net migration, population, employment, GDP and land use for all 27 EU Member States, which can be used for the analysis of variety of issues, including internal migration and urbanization.

Barro, R. J., Sala-I-Martin, X., Blanchard, O. J., & Hall, R. E. (1991). Convergence Across States and Regions. Brookings Papers on Economic Activity, 1991(1), 107. https://doi.org/10.2307/2534639.

Charron, N., Dijkstra, L. & Lapuente, V. (2015). Mapping the regional divide in Europe: A measure for assessing quality of government in 206 European regions. Social indicators research, 122, pp. 315-346.

Directorate-General for Economic and Financial Affairs. (2021). The 2021 Ageing Report: Economic and Budgetary Projections for the EU Member States (2019-2070). In European Commission (Vol. 2). http://ec.europa.eu/economy_finance/publications/european_economy/2011/pdf/ee-2011-4_en.pdf.

Feranec, J., Hazeu, G., Kosztra, B., & Arnold, S. (2016). CORINE land cover nomenclature. European Landscape Dynamics: CORINE Land Cover Data; Feranec, J., Soukup, T., Hazeu, G., Jaffrain, G., Eds, 17-25.

Pigaiani, C. & Batista E Silva, F. (2021), The LUISA Base Map 2018, EUR 30663 EN, Publications Office of the European Union, https://doi:10.2760/503006.

Schiavina, M., Melchiorri, M., Pesaresi, M., Politis, P., Freire, S., Maffenini, L., Florio, P., Ehrlich, D., Goch, K., Tommasi, P. & Kemper, T. (2022). GHSL data package 2022. Publications Office of the European Union: Luxembourg. researchgate.net/profile/Marcello-Schiavina/publication/361565927_GHSL_Data_Package_2022/links/62b9bc3893242c74cad1ba02/GHSL-Data-Package-2022.pdf.