

Moffat, John; Roth, Duncan

**Working Paper**

## Cohort size and youth unemployment in Europe: a regional analysis

MAGKS Joint Discussion Paper Series in Economics, No. 40-2014

**Provided in Cooperation with:**

Faculty of Business Administration and Economics, University of Marburg

*Suggested Citation:* Moffat, John; Roth, Duncan (2014) : Cohort size and youth unemployment in Europe: a regional analysis, MAGKS Joint Discussion Paper Series in Economics, No. 40-2014, Philipps-University Marburg, Faculty of Business Administration and Economics, Marburg

This Version is available at:

<https://hdl.handle.net/10419/105077>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**No. 40-2014**

**Duncan Roth and John Moffat**

**Cohort size and youth unemployment in Europe: a regional  
analysis**

This paper can be downloaded from  
[http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index\\_html%28magks%29](http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/index_html%28magks%29)

Coordination: Bernd Hayo • Philipps-University Marburg  
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg  
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: [hayo@wiwi.uni-marburg.de](mailto:hayo@wiwi.uni-marburg.de)

# Cohort size and youth unemployment in Europe: a regional analysis\*

John Moffat (Durham University)†

Duncan Roth (Philipps-Universität Marburg)‡

July 2014

## Abstract

Will the projected decline in the youth share of European countries' populations alleviate the currently high levels of youth unemployment in Europe? Economic theory predicts that in the absence of perfectly competitive labour markets, changes in the relative size of age groups will cause changes in age-specific unemployment rates. In light of the expected development of the youth population's size over the coming decades, this paper utilises the existing heterogeneity in the structure of youth populations across European countries and regions to identify the effect of nationally and regionally defined age-cohort size on the probability of young individuals being unemployed. To account for the possibility that individuals self-select into areas of low unemployment, the empirical analysis employs an instrumental variables estimator to identify the causal effect of age-cohort size. The results show that individuals in larger cohorts are more likely to be unemployed and that this effect is more pronounced when analysis is conducted at the regional level. While shrinking youth cohorts therefore have the potential to contribute to improving the current youth unemployment situation, this mechanism should not be relied in isolation upon due to the relatively greater importance of changes in the macroeconomic environment.

JEL classification: J10, J21, R23

Key words: Cohort size, unemployment, regional labour markets, causal effect, instrumental variables, EU-SILC

---

\* The authors would like to thank Bernd Hayo as well as the participants of the MACIE brownbag seminar, the MAGKS workshop, the 63<sup>rd</sup> Congress of the French Economic Association and the 7<sup>th</sup> GfR Summer Conference for helpful comments and suggestions.

† Durham University, Department of Economics, Finance and Accounting

‡ Corresponding author: Philipps-Universität Marburg, Department of Economics, Am Plan 1-2, Tel.: +49-6421-28-23751, Fax: +49-6421-28-28912, Email: Duncan.roth@wiwi.uni-marburg.de

# 1 Introduction

According to the German Chancellor Angela Merkel, ‘youth unemployment is perhaps the most pressing problem facing Europe at the present time’ (BBC, 2013). In addition to foregone output, reduced tax revenue and higher expenditure on unemployment benefits which are costs to society of unemployment regardless of the age of the unemployed, youth unemployment has some particularly pernicious effects. Empirical evidence suggests that individuals who experience unemployment when young are left with long-lasting wage scars (Arulampalam, 2001; Gregg and Tominey, 2005) and have a higher probability of becoming unemployed again (Mroz and Savage, 2006). In addition to effects on labour market outcomes, unemployment experiences also have undesirable effects on non-economic outcomes such as happiness, nutrition and mental well-being (see Bell and Blanchflower, 2010, for a summary).

European countries differ substantially with respect to their youth unemployment experiences (see Cahuc et al, 2013, for a comparison of the different youth unemployment situations in France and Germany). Figure A1a in the appendix shows the heterogeneity in the national youth unemployment ratio<sup>1</sup> for the year 2010<sup>2</sup>. While many Central European countries have unemployment ratios belonging to the two lowest quintiles, youth unemployment is especially severe in Spain, Ireland, Sweden, Estonia and Latvia. Using less aggregated spatial units, Figures A1b and A1c display the youth unemployment ratio across the first and second level, respectively, of the Nomenclature of Territorial Units in Statistics (NUTS). Both figures illustrate that the nationally defined variable masks considerable heterogeneity in the extent of youth unemployment within countries.

Youth unemployment rates in Europe are traditionally higher than the corresponding rates for adults. One factor that determines the level of youth unemployment is the state of aggregate demand. Bell and Blanchflower (2010, 2011) show that in most European countries the youth unemployment rate increased considerably during the recent recession and more so than the overall unemployment rate. This is unsurprising as, during a recession, youth flows into employment will fall as it becomes more difficult for labour market entrants to find jobs while youth flows out of employment will increase disproportionately because young workers are likely to be among the first to be made redundant given their lack of work experience and firm-specific human capital.

In addition to the effects of aggregate demand, the extent of unemployment is also influenced by the supply-side of the labour market. This paper contributes to the understanding of the determinants of youth unemployment by using the existing variation in youth population structures across European regions and countries to identify the effect of changes in these structures on the probability of young individuals being unemployed. The heterogeneity in the size of the youth share across European countries and regions is illustrated in Figure A2.<sup>3</sup> In light of the forecast decline in

---

<sup>1</sup> The youth unemployment ratio is defined as the number of unemployed aged between 15 and 24 relative to the total population aged between 15 and 24. In contrast, the youth unemployment rate would be standardised on the youth labour force. The unemployment ratio is chosen since the subsequent empirical analysis is not restricted to labour market participants.

<sup>2</sup> As the dataset used in the empirical analysis covers the years 2007-2010, data from the year 2010 is chosen in favour of more recent statistics.

<sup>3</sup> The youth share is defined as the ratio of the 15-24 year olds relative to the population of 15-65 year olds. Data are from the year 2010. As explained in section 3, the measure of the youth population used in the empirical analysis does not only vary across spatial units (countries, NUTS 1 and NUTS 2 regions) but also across age groups and time.

the share of the youth population throughout Europe (Eurostat, 2014), it is important to understand whether a shrinking supply of young individuals will alleviate the youth unemployment problem and how the magnitude of such an effect compares to the importance of other determinants of youth unemployment.

In order to look at this question, we use the EU Statistics on Income and Living Conditions (EU-SILC) dataset which contains information from most European countries as well as from a large number of NUTS 1 and NUTS 2 regions at the individual and household level. To deal with the endogeneity of the youth population in a given spatial unit, which arises due to the potential for migration of individuals from areas of high unemployment to areas of low unemployment, an instrumental variables (IV) approach is taken. While earlier papers in the literature have used either the country or the region as the spatial unit to define a measure of the youth population, this paper uses the country, NUTS 1 and NUTS 2 region. This more flexible approach allows us to see whether the estimated effect is influenced by the choice of spatial unit and therefore provides guidance to future researchers on the appropriate geographic unit to be used in the analysis of cohort size effects. Our results show that being part of a large age group increases the probability of being unemployed regardless of whether the analysis is conducted at the national, the NUTS 1 or the NUTS 2 level but that the estimated effects are more pronounced at the regional than at the national level. If labour markets are defined at a sub-national level, which seems very likely in larger countries, the true effect of changes in the youth population structure will be underestimated by the use of nationally defined measures.

Section 2 outlines the economic argument why changes in population structures should be expected to have an effect on unemployment and reviews the existing literature. Section 3 discusses the dataset, the empirical model and the identification strategy. Results and robustness checks are presented in section 4. Section 5 concludes.

## **2 Population structures and unemployment**

The way in which changes in population age structures affect labour market outcomes represents a research topic that has been receiving attention since the entrance of the baby boom cohorts to the US labour market in the 1960s. The standard approach taken to modelling the effects of changes in the population age structure is to specify a production function in which differently aged individuals enter as distinct factors of production which are only imperfectly substitutable (for examples of this approach, see Freeman, 1979; Welch, 1979; Berger, 1983; Connelly, 1986; Stapleton and Young, 1988; Brunello, 2010). Welch (1979) motivates the assumption of imperfect substitutability of differently aged workers<sup>4</sup> by proposing a model in which each worker's career consists of distinct phases which are characterised by the performance of different tasks. Imperfect substitutability then follows from differently aged workers being at different stages of their respective careers (see also Skirbekk, 2003 for a literature survey on how differences in work experience and mental abilities affect the respective productivity of young and old workers in dealing with different types of tasks).

In a perfectly competitive economy, factors of production are paid their marginal value product. If the production process is characterised by diminishing marginal productivity, economic theory predicts that an increase (decrease) in a specific age group's size will be associated with a decrease

---

<sup>4</sup> In Welch's (1979) model, workers are differentiated by experience rather than age.

(increase) in the marginal productivity of its members and hence with lower (higher) wages.<sup>5</sup> The exact extent to which the effect of a change in the size of a specific age group will be concentrated on the members of that group will depend on the degree of substitutability between differently aged workers (with greater substitutability causing smaller falls in wages). Empirical studies from the US and EU mainly support the prediction that the wages of age- or age-education-specific groups is negatively affected by the relative size of that group (for the US see Freeman, 1979; Welch, 1979; Berger, 1983 and 1985; for the EU see Wright, 1991; Brunello, 2010; Moffat and Roth, 2013).

Fertig and Schmidt (2004), however, question the adequacy of analysing the effect of changes in the age distribution of the workforce on labour market outcomes using a framework of perfect competition, especially in the context of European economies. The authors argue that instead of merely reflecting marginal productivities, wages are likely to be the outcome of a bargaining process between employer associations and unions. In these circumstances wages will also be determined by non-economic factors, e.g. the preferences and bargaining power of various factions within the bargaining associations. If age-specific wages are set above the corresponding marginal value products, changes in the relative size of a specific age group may therefore not be reflected in adjustments to age-specific wages but rather to age-specific unemployment or employment rates.

The existing theoretical literature has mainly focused on the use of a right-to-manage framework in order to model labour market imperfections: in these models, monopolistic unions maximise a utility function which depends on the total wage bill (i.e. the sum of the age-specific product of wage and employment levels) by setting these quantities for each age group subject to the firm's profit maximisation condition. The existence of this constraint implies that for each age group unions face a trade-off between the wage and the level of employment as setting the former above its marginal value product will lead to a part of this age group becoming unemployed. The union will be willing to accept the existence of age-specific unemployment as long as the gains in utility from a higher wage for those in employment exceed the losses resulting from reduced employment levels. In addition, the union may discriminate between differently aged workers by placing different weights on the age-specific parts of the wage bill. Within this framework, Schmidt (1993) derives expressions for the equilibrium unemployment levels of different age groups which are functions of the relative size of the corresponding age group. Michaelis and Debus (2011) show that if the union adjusts the age-specific weights of its objective function in favour of the preferences of old workers relative to those of young workers – a likely scenario if the share of old workers in the labour force increases – the unemployment rate and wages of the old (young) increases (decreases). Both models provide support for the existence of a relationship between age-cohort size and age-specific unemployment either through a direct connection as in Schmidt (1993) or indirectly through a shift in union preferences as a result of demographic processes that increase the share of old workers in the workforce as in Michaelis and Debus (2011).

Relatively few studies have empirically investigated the impact of population structure on the probability of unemployment. Using time-series variation in West German data spanning the period

---

<sup>5</sup> Card and Lemieux (2001) and Brunello (2010) formally derive the inverse labour demand functions for age-specific labour inputs within the framework of a perfectly competitive economy from a constant elasticity of substitution (CES) aggregate production function. Under these assumptions the age-specific wage is inversely related to the size of that age group relative to the whole population. Modelling total labour input in form of a CES function of differently aged labour inputs ensures that the former are only imperfectly substitutable across age.

1967 to 1988, Zimmermann (1991) finds no evidence for a long-run relationship between the youth share and annual age-specific unemployment rates. However, his results suggest that in the short run an increase in the relative size of the youth population increases the male unemployment rates of most age groups. Applying similar time-series analysis to annual West German data on age-sex-specific cohort size – defined as the ratio of selected five-year age groups to the population aged between 16 and 64 – and age-sex-specific unemployment rates for the period 1967 to 1989, Schmidt (1993) performs co-integrating regressions between the levels of both variables. While the point estimates of the cohort size variable take on a positive value in the majority of sex-age groups, evidence for a co-integrating relationship can be found for only a small number of sex-age groups. In both cases the inability to identify any substantial evidence for a long-run relationship between cohort size and unemployment may be due to the comparatively short time span available for analysis.

Fertig and Schmidt (2004) rely on cross-sectional variation in the size of five-year age-cohorts across different European countries to identify the formers' effect on an individual's employment probability. The authors use the 1999 wave of the European Community Household Panel (ECHP) to estimate a binary dependent variable model but do not address the possibility of endogeneity of the cohort size variable arising from potential self-selection of individuals into geographic areas through migration. Their findings suggest that the effect of an increase in cohort size depends on its size: for relatively small cohorts, the marginal effects on the individual employment propensity are negative, while they are positive for larger cohorts. The authors interpret their findings as evidence that sufficiently large cohorts are able to exploit their bargaining power to limit the negative consequences of belonging to a large cohort.

Using aggregated data from 15 OECD countries over the period 1970 to 1994, Korenman and Neumark's (2000) cross-national study uses longitudinal as well as cross-sectional variation in the young-to-old population ratio at the country level to identify the former's effect on the annual youth unemployment and the youth employment rate. The authors apply IV estimation based on lagged birth rate data to address the potential endogeneity of their youth share variable. Their findings suggest that an increase in the relative size of the population of young individuals is associated with increases in the youth unemployment rate. In contrast, their cohort size variable has no statistically significant effect on the youth employment rate.

Most recently, Garloff et al. (2013) estimate the effect of the youth share on the aggregate unemployment rate in Germany using the Sample of Integrated Labour Market Biographies (SIAB) for the time period 1993-2009. While identification also relies on a combination of cross-sectional and longitudinal variation, in contrast to previous studies, the authors are able to utilise the regional variation in population structure by constructing the youth share variable at the level of 108 West German functional labour market (FLM) districts.<sup>6</sup> After instrumenting the youth share with a lagged value, they find that an increase in the youth share leads to increases in the annual unemployment rate and to decreases in the annual employment rate.

Most of the papers discussed above indicate that changes in the size of differently defined age groups have an effect on age-specific unemployment. Two measures of population structure have

---

<sup>6</sup> Functional labour market (FLM) districts represent combinations of one or more administrative units at the NUTS 3 level. Information on the construction of the FLM districts can be found in Eckey et al. (2006).

been used: the relative age-cohort size (i.e. the number of individuals in a specific age group relative to the total population) or the youth share (i.e. the ratio of young to total population or the young-to-old population ratio). In our view, the former is the better measure because it better captures the idea of differently aged individuals representing different factors of production and also allows the use of variation in cohort size over age groups (rather than only over time and/or spatial units). We therefore follow this approach.

Previous studies also differ with respect to their identification strategies. While earlier studies rely on either cross-sectional or time-series variation in population structures, the emphasis has shifted to combining both dimensions of variation. There is also a greater awareness of the potential endogeneity of the cohort size or youth share variable arising from individual self-selection through migration. Due to the nature of our dataset we are able to exploit variation over time while our identification strategy, which involves the use of information on the relative size of past age groups as an instrument for the current age group, controls for the endogeneity of the cohort size variable.

An additional feature of recent studies is the use of the region rather than the country as the spatial unit for the construction of the population structure variable (see Garloff et al, 2013; Gregory and Patuelli, 2013)<sup>7</sup>. This, on the one hand, allows regional heterogeneity in population structures to be used to identify the impact of cohort size but is also arguably a more appropriate level to define labour markets. To test the importance of the geographical units used to define cohort size, the effects of the youth population structure will be estimated at the national but also at the NUTS 1 and NUTS 2 level.

Moreover, our paper differs from the majority of the reviewed literature by focussing specifically on unemployment outcomes of young individuals rather than that of the entire population. Unemployment rates among youths in Europe are higher than for the rest of the working age population, while the formers' relative lack of work experience and firm-specific knowledge means that young individuals generally face different labour market conditions than the more experienced adult population (Bell and Blanchflower, 2010 and 2011). The labour market response to changes in population structure may therefore differ systematically between youths and adults which means that the use of labour market outcomes for the entire population may obscure the effects on younger members. Finally, in order to control better for individual characteristics, our unit of analysis is the individual rather than aggregate measures.

## **3 Empirical model**

### **3.1 Data**

The data used in the empirical analysis come mainly from the 2010 longitudinal version of the EU-SILC.<sup>8</sup> This dataset includes a wide range of information about individuals and households from 27 European countries for the period 2007-2010. This dataset therefore lends itself to an analysis of individual unemployment probabilities as it allows various individual- and household-level characteristics which may determine the probability of unemployment to be controlled for in the

---

<sup>7</sup> Gregory and Patuelli (2013) use variation in population structure across 332 German NUTS 3 regions to estimate the former's effect on regional innovation output.

<sup>8</sup> Specifically, the data come from the *EUSILC LONGITUDINAL UDB 2010 – version 3 of August 2013*



empirical model. Additional data, such as national and regional unemployment rates and information on the size of lagged age groups (which are used as instruments in the subsequent empirical analysis), are taken from Eurostat's publicly available database.<sup>9</sup>

The EU-SILC dataset not only contains information about an individual's residence at the country level but also at the NUTS 1 level. Some countries also provide information at the NUTS 2 level.<sup>10</sup> The availability of this information allows the construction of variables measuring age-specific youth cohort size at a national as well as at a regional level. If an individual's labour market is better defined at a sub-national level, then using the relative size of age groups within a specific NUTS 1 or NUTS 2 region will provide a better approximation to the potential availability of age-specific labour inputs within a labour market than corresponding variables constructed at the national level. The availability of population variables constructed at different levels of spatial aggregation also allows testing whether the estimated effect of changes in age-cohort size differ across these levels. The use of regional information about the age distribution is in line with recent developments within the literature dealing with the economic consequences of demographic change.<sup>11</sup>

When conducting the empirical analysis using an age-cohort size variable that is defined at the national level, observations from up to 24 countries can be included.<sup>12</sup> But if the sample is restricted to those observations for which information about residence at the NUTS 2 level is also provided, the following 13 countries remain: Bulgaria, Cyprus, Czech Republic, Estonia, Finland, France, Iceland, Latvia, Lithuania, Luxembourg, Romania, Spain and the UK. Figure A3a provides a graphical illustration of those countries that are not available for analysis and those countries that are available in the full and in the restricted sample, respectively. At the NUTS 1 level, observations from 67 regions can be included in the full sample and 39 in the restricted sample, while the dataset includes observations from 97 NUTS 2 regions.<sup>13</sup> Figures A2b and A2c show the NUTS 1 and NUTS 2 regions, respectively that are available in the full and the restricted sample.

---

<sup>9</sup> The data can be obtained through the following link:

[http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search\\_database](http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database)

<sup>10</sup> NUTS regions correspond to administrative units within countries. E.g. at the NUTS 1 level, France is divided into eight large regions (Zones d'études et d'aménagement du territoire), while Austria consists of three groups of states (Gruppen von Bundesländern). Smaller countries (e.g. Cyprus, Denmark, Latvia) may consist of a single NUTS 1 region only. At the NUTS 2 level, France consists of 27 regions (régions), while Austria is divided into nine states (Bundesländer).

<sup>11</sup> For this kind of analysis, spatial units should ideally be self-contained, meaning that the resident population and the population working in the same spatial unit should coincide as much as possible. NUTS regions are administrative units and thus do not necessarily satisfy this criterion. As an example of the former type of spatial units, Eckey et al (2006) derive functionally delineated labour market regions for the case of Germany. These regions are based on one or more administrative NUTS 3 regions (*Kreise*) and satisfy conditions regarding population size and commuting time. While similar classifications are available for a subset of countries, no corresponding system exists for Europe as a whole.

<sup>12</sup> Since observations from the Netherlands and Slovenia do not include information on all control variables, they are not included in the sample. Limited availability on past population figures which are used as instruments in the empirical analysis precludes using observations from Malta, while individuals from Cyprus aged 25 in the year 2007 have to be excluded for the same reason.

<sup>13</sup> At the NUTS 1 level, observations from Portugal are excluded because no information on regional residence is provided. An insufficient sample size prevents the inclusion of observations from Scotland (UKM) and Northern Ireland (UKN). Due to changes in the delineations of the Italian regions Nord-Est (ITH) and Centro (ITI) in the year 2010 no lagged population data (required for construction of the instruments) exists for these newly created spatial units and consequently observations from these regions have to be excluded from the analysis at the NUTS 1 level. Only a subset of countries provides information about residence at the NUTS 2 level. Individual regions have to be dropped because of insufficient sample sizes or changes in the delineations.

### 3.2 Sample selection

This paper aims to estimate the causal effect of the relative size of age groups on the unemployment probability of young individuals. The sample therefore is restricted to males and females aged between 18 and 25 years.<sup>14</sup> In addition, individuals are included in the sample on the basis of their self-declared economic status.

The relevant variable in the dataset distinguishes between several categories which can be aggregated into three main groups<sup>15</sup>: a) employed individuals, b) unemployed individuals and c) individuals that are not participating in the labour force. The final category consists of those that are i) in the process of acquiring education, ii) retired, iii) permanently disabled and/or unfit to work, iv) in compulsory military service, v) fulfilling domestic tasks and vi) otherwise inactive. Since individuals in sub-groups iii) and iv) cannot participate in the labour market, they are excluded from the sample. All other individuals are included in the sample so that it is not restricted to labour market participants only.<sup>16</sup>

Although, as outlined in section 2, it is expected that in the presence of market imperfections changes in the age structure will have an impact on age-specific unemployment outcomes, it is possible that there are also behavioural responses to cohort crowding which may induce individuals to exit the labour market. For instance, it is well established that the relative size of age groups has an impact on the amount of investment into human capital (see Fertig, Schmidt and Sinning, 2009). These behavioural responses would be precluded, and may confound our results, if the sample were to be restricted to individuals who are active on the labour market.

### 3.3 Empirical model

Similar to Fertig and Schmidt (2003), our model's dependent variable  $unemp_{ijrt}$  is binary, taking the value 1 if individual  $i$ , in age group  $j$ , in spatial unit  $r$  at time  $t$  is unemployed and 0 otherwise. For the reasons outlined above, this definition does not restrict changes in an individual's unemployment status to those due to changes from employment to unemployment and vice-versa but also takes into consideration changes from non-participation to unemployment and vice-versa.

The model's central explanatory variable is the relative size of the age cohort to which an individual belongs. Specifically, it is defined as the number of individuals in age group  $j$  in spatial unit  $r$  in year  $t$ ,  $N_{jrt}$ , relative to the whole population aged between 16 and 65 in the same area and year,  $N_{rt}$ . Since spatial units refer either to countries, NUTS 1 or NUTS 2 regions, the age cohort size variable is constructed at the national level as well as at the two regional levels for ages 18-25.

Being age-specific, the cohort size variable captures the idea that differently aged individuals represent different factors of production. But since the assumption that only identically aged

---

<sup>14</sup> Individuals aged 16 or 17 are not included in the analysis as they are almost exclusively registered as being enrolled in education (91.42%). As a robustness check, we extend the age range to cover individuals between 16 and 25 years of age and are able to show that the results are not affected.

<sup>15</sup> Since the variable measuring an individual's economic status is a subjective measure, it appears to be vulnerable to measurement error. Therefore, answers to this question are compared to other answers, e.g. number of hours worked, to ensure consistency.

<sup>16</sup> For those individuals that are potentially available for the empirical analysis at the national level the distribution across the different categories of self-declared economic status is: employed (36.48%), unemployed (10.37%), in education (46.57%), retired (0.04%), disabled or unfit to work (0.69%), in military service (0.71%), fulfilling domestic tasks (2.88%) and otherwise inactive (2.26%).

individuals are perfectly substitutable in production appears highly restrictive we adopt a measure of cohort size which also considers slightly younger and older individuals (see Wright, 1991; Brunello, 2010; Moffat and Roth, 2013). Specifically, for any age group  $j$ , the numerator of the cohort size variable is defined as a weighted sum of the size of the own age group as well as of the number of individuals who are up to two years older and younger with the weights decreasing with the distance from the reference age  $j$ :<sup>17</sup>

$$(1) \quad CS_{jrt} = \frac{(1/9)N_{(j-2)rt} + (2/9)N_{(j-1)rt} + (3/9)N_{jrt} + (2/9)N_{(j+1)rt} + (1/9)N_{(j+2)rt}}{N_{rt}}$$

To ensure that representative values are estimated, the cohort size variable is constructed using individual sampling weights.

The model of the individual unemployment probability takes the following form where  $r$  refers to countries, NUTS 1 regions and NUTS 2 regions. The model is estimated separately at each of these three levels of spatial aggregation<sup>18</sup>:

$$(2) \quad unemp_{ijrt} = \alpha_0 + \alpha_1 CS_{jrt} + \beta X_{ijrt} + \varepsilon_{ijrt}$$

To ensure that the estimated effect of cohort size on the individual unemployment probability is not confounded by other factors, various control variables are added. At the individual level, an individual's age and level of education are included to control for changes in unemployment probabilities as an individual ages and/or acquires more human capital. In order to allow for nonlinear effects of age and education, both variables are entered as a set of mutually exclusive dummy variables ( $age19_{ijrt}$ - $age25_{ijrt}$  and  $isced1_{ijrt}$ - $isced5_{ijrt}$ ). An individual's gender ( $male_{ijrt}$ ) and marital status ( $married_{ijrt}$ ) are included, while the effect of population density is also controlled for ( $urban2_{ijrt}$  and  $urban3_{ijrt}$ ). The national ( $totalunemp\_country_{rt}$ ) and regional ( $totalunemp\_nuts1_{rt}$  and  $totalunemp\_nuts2_{rt}$ ) unemployment rates are added to control for aggregate demand effects.<sup>19</sup> Definitions of the included variables and basic descriptive statistics can be found in the appendix in Table A1 and Table A2, respectively.

In addition to the above-mentioned control variables, the model includes year and country dummies. When estimated at the regional level, NUTS 1 or NUTS 2 dummies are added to the model. While the use of time dummies allows us to control for year-specific shocks which are common to all spatial units, the country- and region-specific dummies ensure that the coefficients of age-cohort size are not confounded by time-invariant unobserved heterogeneity at the level of the corresponding spatial unit which could, for example, be caused by differences in national labour market institutions.

### 3.4 Identification strategy

The effect of cohort size on the individual unemployment probability is identified through variation in the former variable. At each level of aggregation (country, NUTS 1 and NUTS 2 regions), it is

<sup>17</sup> A discussion of the specified weights can be found in Wright (1991).

<sup>18</sup> The model includes a constant. We do not estimate the model of each spatial level as part of a system of equations because consistent estimation requires homoscedastic errors (see Cameron and Trivedi, 2009).

<sup>19</sup> While the national unemployment rate is included as a regressor regardless of the level of spatial aggregation, the regional NUTS 1 unemployment rate is only added if the cohort size variable is defined at the NUTS 1 level and accordingly for the NUTS 2 regional unemployment rate.

possible to construct a cohort size variable which varies across spatial units, age groups and time (the latter due to the availability of four waves of observations). This variation is used to identify how changes in population structures affect the unemployment probability of young individuals.<sup>20</sup> The availability of cross-sectional and longitudinal variation is preferable as a source of identification to the use of time-series variation in cohort size only. As emphasised by Korenman and Neumark (2000), the separate identification of the effect of cohort size on labour market outcomes from period effects may be difficult using time-series data only, especially if the underlying series is short.<sup>21</sup>

Identification of the causal effect of cohort size on the individual unemployment probability is complicated by the fact that individuals cannot be assumed to be randomly allocated into geographically defined age cohorts. In contrast, it is more likely that membership of an age cohort in a specific spatial unit is the result of individual self-selection into a region through migration. If individuals choose to migrate into areas with lower unemployment probabilities, the cohort size variable is endogenous and fitting the model by ordinary least squares (OLS), logit or probit will result in inconsistent estimates of the cohort size coefficient.

To address the endogeneity of the cohort size variable due to individual self-selection into spatial units, recent studies on the effects of changes in population structure on labour market outcomes have employed an IV identification strategy using birth rates or lagged population variables as instruments (for examples of this approach see Korenman and Neumark, 2000; Garloff et al, 2013 or Moffat and Roth, 2013). A similar approach will be used in this paper. In particular, two instruments will be constructed which both assume that within each spatial unit the relative size of a given age cohort at some point in the past has predictive power for the size of that cohort in the current year.

The validity of this identification strategy requires that the proposed instruments have no direct effect on the individual unemployment probability and can therefore be legitimately excluded from Equation (2). Since the number of excluded instruments exceeds the number of endogenous variables, a Hansen test statistic can be calculated to test the validity of the overidentifying restrictions. To be consistent with the construction of the cohort size variable in Equation (1), both instruments are also defined as weighted averages which take into account the size of surrounding age groups. Analogously to the construction of the age-cohort size variable, the instrumental variables are constructed at the national, the NUTS 1 and the NUTS 2 level.

The first instrument applies a constant age ( $CS\_ins\_cage$ ) and is defined as the number of individuals aged ten in the year in which the reference age group was ten years old,  $N_{10r(t-j+10)}$ , relative to the population aged 16 to 65 in that year,  $N_{r(t-j+10)}$ :

---

<sup>20</sup> The use of cross-regional data spanning several countries as compared to regional data for a single country may be criticised as differences in national labour market institutions could have an effect on how changes in the relative size of age groups translate into changes in age-specific unemployment probabilities. Cross-regional data from a single country may in principle be able to address this issue, but only to the extent that the country's labour market institutions remain unchanged over the sample period. In addition, as one of our robustness checks we include country-by-year dummies which should control for the effects of changes in national institutions.

<sup>21</sup> As an example, Korenman and Neumark (2000) refer to the entrance of young workers into the US labour market during the period 1973-1984 which was marked by unfavourable macroeconomic conditions. Assuming a labour market entry age of 20 years, these individuals were born during the US baby boom period. Using solely longitudinal variation, it will be difficult to separate the effects of cohort crowding at the time of labour market entry on subsequent labour market outcomes from those deriving from an unfavourable macroeconomic condition.

$$(3) \quad CS\_ins\_cage_{jrt} = \frac{(\frac{1}{9})N_{8r(t-j+10)} + (\frac{2}{9})N_{9r(t-j+10)} + (\frac{3}{9})N_{10r(t-j+10)} + (\frac{2}{9})N_{11r(t-j+10)} + (\frac{1}{9})N_{12r(t-j+10)}}{N_{r(t-j+10)}}$$

In contrast, the second instrument applies a constant lag ( $CS\_ins\_clag$ ) and is defined as the number of individuals fourteen years younger than the reference age group fourteen years ago,  $N_{(j-14)r(t-14)}$ , relative to the population aged 16 to 65 in that year,  $N_{r(t-14)}$ .<sup>22</sup>

$$(4) \quad CS\_ins\_clag_{jrt} = \frac{(\frac{1}{9})N_{(j-14)r(t-14)} + (\frac{2}{9})N_{(j-14)r(t-14)} + (\frac{3}{9})N_{(j-14)r(t-14)} + (\frac{2}{9})N_{(j-14)r(t-14)} + (\frac{1}{9})N_{(j-14)r(t-14)}}{N_{r(t-14)}}$$

Unlike Fertig and Schmidt (2003) we do not estimate a binary dependent variable model despite the binary nature of the variable *unemp*. In the presence of endogenous regressors, consistent estimation of a binary dependent variable model using instrumental variables requires that strong distributional assumptions be met.<sup>23</sup> Instead, we implement a less structural approach by fitting a linear probability model (LPM) with an IV estimator (see Cameron and Trivedi, 2009). While the binary nature of the dependent variable is not taken account of by such an approach, it imposes fewer distributional assumptions.

Because the model is overidentified, the two-step generalized method of moments (GMM) estimator is used which increases efficiency through optimal choice of the weighting matrix.<sup>24</sup> Finally, estimation is weighted using the individual sampling weights and cluster-robust standard errors are estimated.<sup>25</sup>

The outlined approach aims to identify the causal effect of the current size of an individual's age cohort on the same individual's probability of being unemployed. However, if an individual belongs to a large cohort, it is likely that the same individual was also born as part of a large birth cohort and the size of the latter might also exert an effect on current labour market outcomes. In particular, if the number of births increases relative to population size, it can be expected that the average family size will also increase. Since family size has been found to negatively affect the level of education that children acquire (Booth and Kee, 2009), it is likely that it is also related to outcomes such as unemployment and that birth cohort size is therefore a determinant of the unemployment probability.

We argue that any potential contamination of the coefficient of current labour market cohort size will be limited due to two factors. First, we control for the level of education and thus for the channel through which birth cohort size (via family size) is expected to affect future labour market outcomes. Second, the comparatively small number of age groups (18-25) and years (2007-2010) ensures that within spatial units the individuals considered in this sample will have been born into similarly sized birth cohorts. The oldest individual in the sample was born in the year 1982 and the youngest in 1992. Inspection of the ratio of births to population size over the period 1982-1992 and

<sup>22</sup> The first year of sampling is the year 2007. Since regional population data is available for most NUTS 1 and NUTS 2 regions from the year 1993 onwards, fourteen years represents the longest feasible lag.

<sup>23</sup> Consistent estimation of a binary dependent variable model with endogenous regressors requires that the error terms of the structural and the first-stage equation are jointly normally distributed and homoscedastic (see Cameron and Trivedi, 2009).

<sup>24</sup> In the presence of heteroskedastic errors, two-step GMM is the more efficient estimator compared to the two-stage least squares (2SLS) estimator (see Cameron and Trivedi, 2005).

<sup>25</sup> Clustering takes place at the individual level. When heteroskedasticity-robust standard errors are used, they are slightly smaller than the cluster-robust versions, which are presented here. Otherwise their use does not affect the results or conclusions. Results are available from the authors upon request.

across the countries of the restricted sample shows that approximately 89% of the overall variation in this variable results from between-country variation.<sup>26</sup> If belonging to a large birth cohort has an effect on subsequent outcomes but if differences in birth cohort size result from differences between spatial units rather than from differences within spatial units, then the use of country and region dummies will account for these effects.<sup>27</sup>

In summary, we have the following hypotheses:

*Hypothesis 1: If labour markets are not fully flexible, we expect that increases in age-specific cohort size will increase the probability of individuals in that age group being unemployed. IV estimation should therefore yield positive coefficients for the age-cohort size variable at every level of spatial aggregation.*

*Hypothesis 2: Less aggregated spatial units provide a better approximation to the size of local labour markets. The effects of age-cohort size on the unemployment probability are therefore expected to be more pronounced at the regional than at the national level.*

*Hypothesis 3: OLS estimation fails to account for self-selection of individuals into spatial units with lower unemployment. Coefficients estimated by OLS are therefore expected to be smaller than the corresponding IV estimates at each level of spatial aggregation.*

*Hypothesis 4: Self-selection into spatial units is likely to increase as these units become smaller. For this reason, we expect differences between OLS and IV estimates to be larger at the regional than at the national level.*

## 4 Results

Since Equation (2) is estimated separately at the national, the NUTS 1 and the NUTS 2 level, Table 1 reports three sets of regression results which include coefficients estimated by OLS and by two-step GMM as well as various instrument-related test statistics.<sup>28</sup> As discussed in section 3.1 the number of observations decreases as the spatial units become more disaggregated. To ensure comparability, the results displayed in Table 1 are derived from a restricted sample that uses at each spatial level only those observations for which information about residence is available at the NUTS 2 level. Table A3 in the appendix shows the results obtained when estimating Equation (2) using the full sample of observations available at each level of spatial aggregation.

Estimation by GMM yields positive and significant coefficients for the age-cohort size variable at every level of spatial aggregation.<sup>29</sup> These results are consistent with the hypothesis that in the

---

<sup>26</sup> In the absence of comparable birth rate data at the NUTS 1 and NUTS 2 level, we perform an identical decomposition of the number of individuals aged ten relative to the total population over the time period 1992-2002. Using the restricted sample, the share of between variation is approximately 78% and 77%, respectively.

<sup>27</sup> Note that the ratio of births to population is not the same as either the cohort size variable or its instruments. While the birth ratio varies only over time and across spatial units, the cohort size variables of this analysis also have an age dimension.

<sup>28</sup> The model contains year dummies, country dummies, NUTS 1 region dummies (only for the analysis at the NUTS 1 level) and NUTS 2 region dummies (only for the analysis at the NUTS 2 level). Full results are available from the authors upon request.

<sup>29</sup> When a 2SLS estimator is used the results are: 6.430\*\*\* (2.249), 10.887\*\*\* (3.066) and 11.682\*\*\* (3.271) at the country, the NUTS 1 and NUTS 2 level, respectively. Since the 2SLS estimator does not use an optimal weighting matrix,

absence of perfectly competitive markets, changes in the relative size of youth age groups will lead to adjustments in age-specific unemployment. Moreover, we also find that the effect of age-cohort size is larger at the regional than at the national level with the estimated coefficients taking on the values 11.704 and 10.833 at the NUTS 2 and NUTS 1 level, respectively, compared to 6.453 at the national level. Comparable results supporting the first two hypotheses are obtained when the full sample of observations is used (Table A3). This increases the number of spatial units from 13 to 24 at the country level and from 39 to 67 at the NUTS 1 level while in both cases the number of available observations approximately doubles.

---

GMM estimates are in principle more precise. In this case, the efficiency gains for the cohort size coefficients are only marginal, though.

**Table 1: Estimated OLS and GMM coefficients of Equation (2) at the country, NUTS 1 and NUTS 2 level (restricted sample)**

| Dependent variable:<br><i>Unemp</i>                       | Country             |                      | NUTS 1              |                      | NUTS 2              |                      |
|---|---------------------|----------------------|---------------------|----------------------|---------------------|----------------------|
|   | OLS                 | GMM                  | OLS                 | GMM                  | OLS                 | GMM                  |
| CS  | -1.545<br>(1.734)   | 6.453***<br>(2.249)  | -1.820<br>(1.323)   | 10.833***<br>(3.065) | -0.891<br>(1.002)   | 11.704***<br>(3.270) |
| Male  | 0.016**<br>(0.007)  | 0.016**<br>(0.007)   | 0.016**<br>(0.007)  | 0.016**<br>(0.007)   | 0.017**<br>(0.007)  | 0.016**<br>(0.007)   |
| Married   | -0.029*<br>(0.015)  | -0.030**<br>(0.015)  | -0.027*<br>(0.015)  | -0.026*<br>(0.015)   | -0.025*<br>(0.015)  | -0.026*<br>(0.015)   |
| Age19   | 0.060***<br>(0.009) | 0.055***<br>(0.009)  | 0.061***<br>(0.009) | 0.054***<br>(0.009)  | 0.059***<br>(0.008) | 0.052***<br>(0.009)  |
| Age20   | 0.099***<br>(0.011) | 0.092***<br>(0.011)  | 0.099***<br>(0.011) | 0.098***<br>(0.011)  | 0.097***<br>(0.011) | 0.096***<br>(0.011)  |
| Age21   | 0.105***<br>(0.012) | 0.098***<br>(0.012)  | 0.105***<br>(0.012) | 0.111***<br>(0.012)  | 0.103***<br>(0.012) | 0.109***<br>(0.012)  |
| Age22   | 0.104***<br>(0.012) | 0.101***<br>(0.012)  | 0.104***<br>(0.012) | 0.116***<br>(0.012)  | 0.101***<br>(0.012) | 0.112***<br>(0.012)  |
| Age23   | 0.114***<br>(0.013) | 0.113***<br>(0.012)  | 0.115***<br>(0.012) | 0.133***<br>(0.013)  | 0.112***<br>(0.012) | 0.130***<br>(0.013)  |
| Age24   | 0.128***<br>(0.012) | 0.127***<br>(0.012)  | 0.127***<br>(0.013) | 0.149***<br>(0.014)  | 0.125***<br>(0.013) | 0.147***<br>(0.013)  |
| Age25   | 0.117***<br>(0.012) | 0.116***<br>(0.012)  | 0.116***<br>(0.012) | 0.135***<br>(0.013)  | 0.113***<br>(0.012) | 0.132***<br>(0.013)  |
| Isced1  | 0.178**<br>(0.089)  | 0.165*<br>(0.089)    | 0.165*<br>(0.087)   | 0.173*<br>(0.089)    | 0.175**<br>(0.079)  | 0.180**<br>(0.079)   |
| Isced2  | 0.041<br>(0.085)    | 0.021<br>(0.085)     | 0.036<br>(0.083)    | 0.043<br>(0.086)     | 0.044<br>(0.076)    | 0.047<br>(0.076)     |
| Isced3  | -0.069<br>(0.084)   | -0.080<br>(0.084)    | -0.072<br>(0.082)   | -0.066<br>(0.085)    | -0.060<br>(0.075)   | -0.057<br>(0.075)    |
| Isced4  | -0.063<br>(0.088)   | -0.070<br>(0.088)    | -0.065<br>(0.085)   | -0.065<br>(0.089)    | -0.048<br>(0.078)   | -0.053<br>(0.079)    |
| Isced5  | -0.120<br>(0.084)   | -0.124<br>(0.084)    | -0.120<br>(0.082)   | -0.111<br>(0.084)    | -0.105<br>(0.074)   | -0.099<br>(0.074)    |
| Urban2  | -0.001<br>(0.009)   | 0.002<br>(0.009)     | -0.006<br>(0.010)   | -0.007<br>(0.010)    | -0.010<br>(0.010)   | -0.011<br>(0.010)    |
| Urban3  | 0.009<br>(0.009)    | 0.012<br>(0.009)     | 0.004<br>(0.009)    | 0.004<br>(0.009)     | 0.004<br>(0.009)    | 0.005<br>(0.010)     |
| Totalunemp_country  | 0.010***<br>(0.001) | 0.012***<br>(0.001)  | 0.004<br>(0.004)    | 0.006<br>(0.004)     | 0.013***<br>(0.004) | 0.018***<br>(0.004)  |
| Totalunemp_nuts1  | -                   | -                    | 0.006<br>(0.004)    | 0.007*<br>(0.004)    | -                   | -                    |
| Totalunemp_nuts2  | -                   | -                    | -                   | -                    | -0.003<br>(0.004)   | -0.005<br>(0.004)    |
| Constant  | 0.067<br>(0.091)    | -0.087<br>(0.096)    | 0.091<br>(0.090)    | -0.201*<br>(0.108)   | 0.132<br>(0.086)    | -0.144<br>(0.108)    |
| N   | 45,623              | 45,623               | 45,623              | 45,623               | 45,623              | 45,623               |
| R <sup>2</sup>  | 0.057               | 0.055                | 0.065               | 0.056                | 0.077               | 0.062                |
| Underidentification test <sup>a</sup>                     | -                   | 2,022.473<br>(0.000) | -                   | 858.376<br>(0.000)   | -                   | 507.508<br>(0.000)   |
| Weak identification test <sup>b</sup>                     | -                   | 4,203.009            | -                   | 624.060              | -                   | 336.006              |
| Hansen's J statistic<br>(overidentification) <sup>c</sup> | -                   | 21.526<br>(0.000)    | -                   | 0.304<br>(0.582)     | -                   | 1.324<br>(0.250)     |

Cluster-robust standard errors are given in parentheses, \*/\*\*/\*\* denote significance at the 10%/5%/1% level

<sup>a</sup> Kleibergen-Paap rank LM statistic (p-value in brackets)

<sup>b</sup> Kleibergen-Paap rank Wald F statistic (critical values are available only for i.i.d. errors)

<sup>c</sup> Overidentification test of all instruments



Table 2 shows the marginal effect of age-cohort size evaluated for an increase in cohort size of 0.0032 units, or 0.32 percentage points, which corresponds to one standard deviation of the cohort size variable at the NUTS 1 level (see Table A2 in the appendix) together with the standard error of the marginal effect, the p-value and the 95% confidence interval (Table A4 shows the estimated marginal effects when the full sample is used). At the country level, an increase in youth cohort size by one standard deviation increases the probability of being unemployed by approximately 2.06 percentage points. This effect is larger at the regional level: the predicted increase is 3.47 percentage points at the NUTS 1 and 3.75 percentage points at the NUTS 2 level, respectively. The difference between the estimated marginal effects therefore manifests itself most clearly when comparing the national and the regional level, while the effects are very similar at the NUTS 1 and the NUTS 2 level.<sup>30</sup>

**Table 2: Marginal effects of cohort size at the country, NUTS 1 and NUTS 2 level**

| Spatial unit | Marginal effect | Standard error | p-value | 95% confidence interval |
|--------------|-----------------|----------------|---------|-------------------------|
| Country      | 0.0206          | 0.0072         | 0.004   | [0.0065; 0.0348]        |
| NUTS 1       | 0.0347          | 0.0098         | 0.000   | [0.0154; 0.0539]        |
| NUTS 2       | 0.0375          | 0.0105         | 0.000   | [0.0170; 0.0580]        |

At each level of spatial aggregation and regardless of whether the restricted or the full sample (Table A3) is used, the OLS point estimates of the cohort size coefficient are smaller than their GMM counterparts. This finding is consistent with the hypothesis that in the presence of self-selection into areas of lower unemployment OLS estimates of the cohort size coefficient are downward biased. Since re-locating from one region to another within the same country can be assumed to be associated with lower costs than moving between different countries, we expect the downward bias of OLS to be more pronounced when the model is estimated at the regional level compared to the country level.<sup>31</sup> Our results provide support for this hypothesis as the differences between the GMM and the OLS point estimates decrease with the level of spatial aggregation.<sup>32</sup> Moreover, we find when using the full sample to estimate Equation (2) at the national level that the OLS point estimates are always positive and for some specifications also significant.<sup>33</sup> While they continue to be smaller than their GMM counterparts, these findings show that the need for estimators able to account for individual self-selection into spatial units becomes more pronounced when regions are used as the basis for analyzing the relationship between population structure and unemployment.

The validity of the identification strategy of using the relative size of a given age cohort in the past as an instrument for its size in the present is supported by the first-stage test results. The test statistic for underidentification is high at each level of spatial aggregation and consequently the null

<sup>30</sup> As can be seen from Table 2, the 95% confidence intervals overlap when comparing the marginal effect estimated at the country level with that estimated at either the NUTS 1 or the NUTS 2 level.

<sup>31</sup> For example, in 2007 a total of 958,266 immigrants arrived in Spain compared to an additional 591,625 individuals who also migrated internally between Spanish NUTS 2 regions (data is obtained from Eurostat's publicly accessible database at [http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search\\_database](http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database)).

<sup>32</sup> The differences between the GMM and the OLS point estimates are 7.998 (country), 12.653 (NUTS 1), 12.595 (NUTS 2) in the restricted sample and 5.829 (country) and 9.308 (NUTS 1) in the full sample.

<sup>33</sup> With the full sample the OLS estimate takes a value of 1.482 (1.170) which is not significant (see Table A3). However, when the alternative specifications for the age-cohort size variable, CS\_A (Equation (5)) and CS\_B (Equation (6)), are employed, the coefficient of the latter is significant at the 0.10 level: 1.132 (1.000) and 1.196\* (0.710). When the sample is extended to include also individuals aged 16 or 17, the coefficients from each specification are significant: 1.766\*\* (0.900), 1.352\* (0.781) and 1.141\*\* (0.570). These findings are robust to the inclusion of country-by-year dummies. Full results are available from the authors upon request.

hypothesis of the endogenous regressor and the excluded instruments being uncorrelated can be strongly rejected. While the test statistics for weak identification are similarly large at every spatial level, critical values only exist for the case of independently and identically distributed errors. Finally, the Hansen J statistic provides a test for the validity of the overidentifying restrictions (OIR). While this statistic is sufficiently small at the NUTS 1 and the NUTS 2 level so that the null hypothesis of correctly excluded instruments cannot be rejected at conventional levels of significance, a similar rejection at the country level is not possible when the restricted sample is used. Despite the large value of the OIR test statistic at the national level, we nevertheless argue that the estimated GMM coefficient provides a reliable measure of the causal effect of age-cohort size on the probability of being unemployed. This is because the GMM point estimates at the country level are very similar regardless of whether the restricted or the full sample is used. However, the value of the OIR test statistic is sufficiently small in the full sample that the null hypothesis of correctly excluded instruments cannot be rejected at even the 0.1 level of significance.<sup>34</sup>

The coefficients of the remaining variables are similar when comparing OLS and GMM results at a given spatial level as well as when comparing the results of a specific estimator across different levels of spatial aggregation. Male individuals are significantly more likely to be unemployed than females, while the probability of being in unemployment is lower for married individuals. Being unemployed is significantly more likely for individuals in each age group relative to the base category (age 18) which may be due to the fact that a large share of individuals are still acquiring secondary education at that age.

Compared to the base category (pre-primary education), individuals with completed primary education (ISCED 1) are significantly more likely to be in unemployment. In contrast, the point estimates of the three highest of the remaining educational groups are negative but not statistically significant. With respect to the degree of urbanization, individuals living in thinly populated areas (Urban3) are more likely to be unemployed than individuals in the base category (densely populated areas) but the estimated coefficients are not statistically significant.

Finally, increases in the national unemployment rate are mainly found to have a positive and significant effect on the probability of young individuals being unemployed but when the analysis is conducted at the NUTS 1 level, the estimated coefficient is smaller and statistically insignificant. A similar result is found for the NUTS 1 unemployment rate. The ability to separately identify the effects of the national and the NUTS 1 unemployment rate is likely to be reduced by the fact that for a subset of the sample country and NUTS 1 region are identical.<sup>35</sup> The estimated effect of the NUTS 2 unemployment rate is small and statistically insignificant.

Once individual self-selection into areas of low unemployment has been accounted for in the estimation, our results show that the youth population structure is a significant determinant of the probability that young individuals are unemployed. The estimated effects are particularly

---

<sup>34</sup> When the full sample is used to estimate the effect of cohort size at the national level, the OIR test statistic is generally sufficiently small that the null hypothesis of correctly excluded instruments cannot be rejected at the 0.1 level of significance. For the alternative measures of cohort size, CS\_A (Equation (5)) and CS\_B (Equation (6)), the values are 1.946 (0.163) and 2.106 (0.147). This finding is robust to the inclusion of country-by-year dummies. The OIR test statistic takes the value 1.659 (0.198) when the main specification of cohort size is used. For the alternative specifications the values are 1.933 (0.164) and 2.174 (0.140), respectively. Full results are available from the authors upon request.

<sup>35</sup> Both variables are highly jointly significant with a  $\chi^2$ -statistic test statistic of 81.32.

pronounced when NUTS 1 or NUTS 2 regions are used as the basis for constructing measures of the youth population. Specifically, we find that a decrease in age-cohort size by one standard deviation leads to a decrease in the probability of being unemployed by 3.75 (NUTS 2) and 3.47 (NUTS 1) percentage points when conducting the analysis at the regional level, while the estimated decrease is 2.06 percentage points at the national level. To put the significance of youth population structure as a determinant of the unemployment probability into perspective, we compare it to the estimated effect of the national unemployment rate. We find that an increase in the national unemployment rate by one standard deviation is predicted to increase the probability of being unemployed by 8.96 percentage points: more than twice the size of the effect that a corresponding change in age-cohort size has on the unemployment probability.<sup>36</sup> The development of the macroeconomic environment will therefore be crucial for the extent to which the projected changes in youth population structure can be expected to alleviate the current youth unemployment problem in Europe.

In order to probe the robustness of the above findings, we modify the empirical model by a) adding further control variables, b) altering the specifications of the cohort size variable and c) extending the age range of sampled individuals. First, we add country-by-year dummies to the model of Equation (2). The time period that is covered by the dataset partly coincides with the financial crisis which affected European countries differently. Inclusion of these interactions allows us to control for any country-and-year-specific shocks, such as the effects stemming from the financial crisis or from changes in national labour market institutions, thereby ensuring that the estimated cohort size coefficients are not confounded by these effects. Table A5 contains the results for the restricted sample. At each spatial level the GMM estimates are approximately 10% smaller than their counterparts in Table 1 but the results continue to support the outlined hypotheses: the GMM coefficients are positive and statistically significant and larger at the regional than the national level. The OLS estimates are smaller but the difference between them and the GMM estimates falls when the analysis is conducted at the national level.<sup>37</sup>

The construction of the age-cohort size variable used above follows that of Wright (1991). While applying weights to surrounding age groups that decrease with the distance from the reference group implies that the substitutability between differently aged workers decreases with the age difference, their specific values are arbitrary. To ensure that our results are not dependent on the weights attached to surrounding age groups in the specification of the cohort size variable we define two alternative measures - a three-year weighted average ( $CS\_A_{jrt}$ ) as well as the own-age cohort size ( $CS\_B_{jrt}$ ) - and re-estimate Equation (2):<sup>38</sup>

$$(5) \quad CS\_A_{jrt} = \frac{(1/4)N_{(j-1)rt} + (1/2)N_{jrt} + (1/4)N_{(j+1)rt}}{N_{rt}}$$

$$(6) \quad CS\_B_{jrt} = \frac{N_{jrt}}{N_{rt}}$$

<sup>36</sup> This effect is obtained by multiplying the GMM coefficient of the national unemployment rate from the NUTS 2 specification (0.018) with the standard deviation of the national unemployment rate (4.98, see Table A2). The effect is equal to 5.98 percentage points when the GMM coefficient from the country-level analysis (0.012) is used.

<sup>37</sup> Due to the inclusion of country-by-year dummies the coefficient of the national unemployment rate is sensitive to the choice of the base categories of these interactions and therefore cannot be given a meaningful interpretation.

<sup>38</sup> The construction of the instruments – Equations (3) and (4) – is adjusted accordingly.

Table A6 shows the estimated coefficients of the alternative cohort size specifications.<sup>39</sup> The size of the estimated coefficients is similar to those presented in Table 1. Moreover, the results obtained from using the two alternative cohort size specifications display the same general pattern as those from the main cohort size specification: the GMM coefficients are all positive and statistically significant and larger than their OLS counterparts. In addition, the coefficients estimated at the regional level exceed those estimated at the national level.

Next we test whether the inclusion of individuals aged 16 or 17 affects the findings. Table A7 presents the results obtained from estimating Equation (2) using an extended age range.<sup>40</sup> These are generally similar to those from the main specification: the GMM estimates are positive and significant, while the OLS estimates are close to zero and statistically insignificant. In contrast to the results of the main specification, the difference between the country-level and the regional coefficients are not as pronounced.

The validity of the results presented in Table 1 is supported by the preceding analysis as the main findings are not affected by the inclusion of country-by-year dummies, alternative specifications of the cohort size variable or the extension of the age range to also include individuals aged 16 or 17.

## 5 Conclusion

In light of the currently high levels of youth unemployment across European countries and regions the aim of this paper has been to identify the effect that changes in the youth population structure have on the probability of young individuals being unemployed. As the share of the youth population is forecast to decline during the coming decades, it is important to understand whether and to what extent this demographic development will alleviate the current youth unemployment problem in Europe.

The central contributions of this paper are its focus on the unemployment outcomes of young individuals as well as its use of different spatial units to define the measure of youth population. The latter allows us to determine whether the estimated effect of age-cohort size on the unemployment probability of young individuals depends on the choice of the spatial level. The use of regionally defined cohort size variables also offers the advantage of utilising the existing intra-country variation in the youth population structure to identify its effect and is also more likely to provide a better approximation to an individual's relevant labour market.

To identify the causal effect of cohort size, IV estimation is used in order to take account of the likelihood that individuals self-select into areas of lower unemployment. Our results show that at each level of spatial aggregation age-cohort size has a positive and significant effect on the probability of young individuals being unemployed. Consistent with the hypothesis of regional spatial units providing better representations of an individual's labour market we find that the effect of cohort size is more pronounced at the regional than at the national level. Specifically, a one-standard-deviation decrease in cohort size is predicted to decrease the unemployment probability by 2.06 percentage points at the country level compared to 3.47 and 3.75 percentage points at the

---

<sup>39</sup> The results are for the restricted sample. Comparable results are found when the full set of observations is used. These results are available from the authors upon request.

<sup>40</sup> The results use a restricted sample of observations that are available at the NUTS 2 level. Comparable results are found when the full set of observations is used. These results are available from the authors upon request.

NUTS 1 and NUTS 2 level, respectively. When estimating the model by OLS we find that the coefficients are smaller than their GMM counterparts and mainly insignificant. As OLS estimation does not account for individual self-selection, the estimates are expected to be downward biased. Moreover, our results show that the difference between the GMM and the OLS estimates decreases when the model is estimated at the country level indicating that an identification strategy based on instrumental variables is especially important when using NUTS 1 or NUTS 2 regions as the spatial basis for measures of youth population structure. We also find some evidence that the effect of cohort size is more pronounced in countries with higher levels of wage rigidity.

While being a significant determinant of the unemployment probability of young individuals, we find that the effect of a one-standard-deviation decrease in cohort size would be more than offset by a comparable increase in the national unemployment rate. Consequently, we conclude that the projected decline in the youth population share has the potential to contribute to alleviating the currently high levels of youth unemployment in Europe. However, an improvement simply by means of demographic processes should not be relied upon due to the relatively greater importance of changes in the macroeconomic environment.

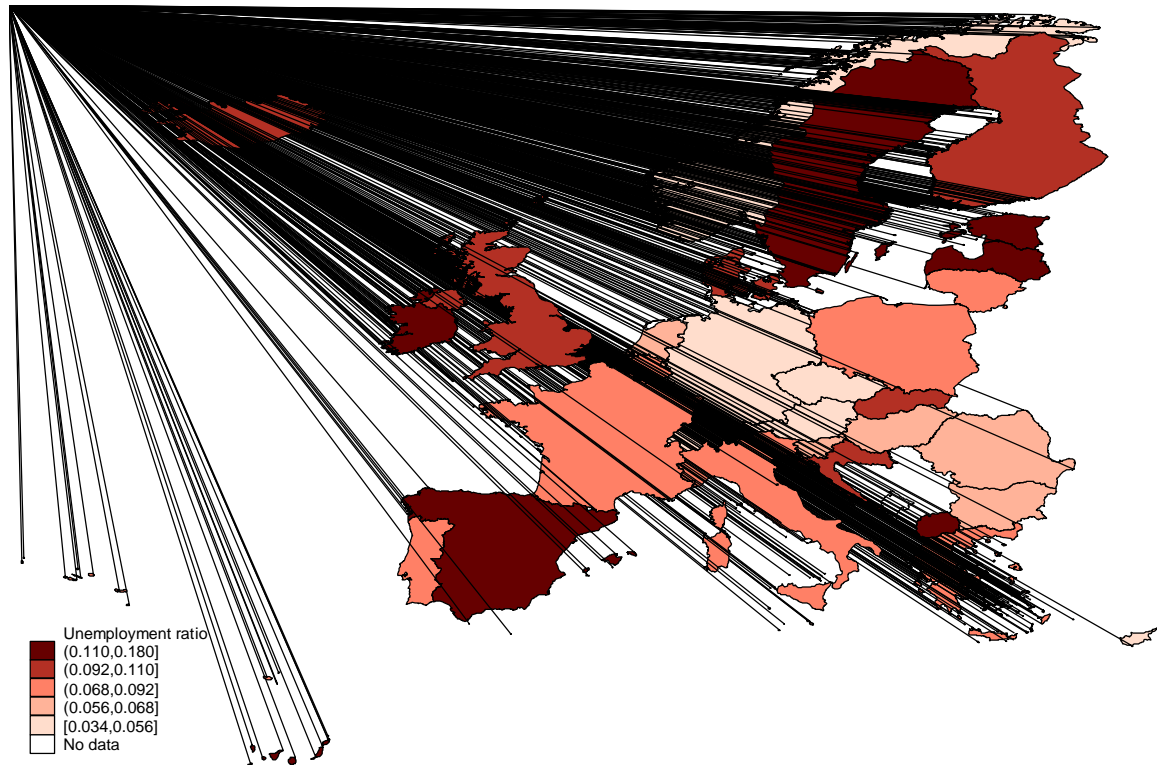
## 6 References

- Arulampalam W (2001) Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages. *Econ. J.* 111:F585-F606. doi: 10.2307/798307.
- BBC. (2013) Merkel Warns on Youth Unemployment Problem [Online]. Available: <http://www.bbc.co.uk/news/business-23160043> [Accessed 05/03 2014].
- Bell D, Blanchflower D (2010) Youth Unemployment: Déjà Vu? IZA Discussion Paper 4705.
- Bell D, Blanchflower D (2011) Youth Unemployment in Europe and the United States. *Nordic Econ. Pol. Rev.* 11:11-38.
- Berger M (1983) Changes in Labor Force Composition and Male Earnings: A Production Approach. *J. Human Res.* 18:177-196. doi: 10.2307/145481.
- Berger M (1985) The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence. *J. Polit. Economy* 93:561-573. doi: 10.2307/1832008.
- Booth A, Kee H (2009) Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment. *J. Population Econ.* 22:367-397. doi: 10.1007/s00148-007-0181-4.
- Brunello G (2010) The Effects of Cohort Size on European Earnings. *J. Population Econ.* 23:273-290. doi: 10.1007/s00148-009-0250-y.
- Cahuc P, Carcillo S, Rinne U, Zimmermann K (2013) Youth Unemployment in Old Europe: The Polar Cases of France and Germany. *IZA Journal of European Labor Studies* 2:18. doi: 10.1186/2193-9012-2-18.
- Cameron AC, Trivedi PK (2005) *Microeconometrics: Methods and Applications*, Cambridge University Press. Cambridge.
- Cameron AC, Trivedi PK (2009) *Microeconometrics Using Stata*, Stata Press. College Station, Texas.
- Card D, Lemieux T (2001) Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *Quart. J. Econ.* 116:705-746. doi: 10.1162/00335530151144140.
- Connelly R (1986) A Framework for Analyzing the Impact of Cohort Size on Education and Labor Earning. *J. Human Res.* 21:543-562. doi: 10.2307/145766.
- Eckey H-F, Kosfeld R, Türck M (2006) Abgrenzung Deutscher Arbeitsmarktregionen. *Raumforschung und Raumordnung* 64:299-309. doi: 10.1007/bf03183178.
- Eurostat. (2014) *Europop2010 – Population Projections at National Level* [Online]. Available: <http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database> [Accessed 02/01 2014].

- Fertig M, Schmidt C (2004) Gerontocracy in Motion: European Cross-Country Evidence on the Labor Market Consequences of Population Ageing. *In: Wright R (ed.) Scotland's Demographic Challenge*. Scottish Economic Policy Network. Stirling; Glasgow
- Fertig M, Schmidt C, Sinning M (2009) The Impact of Demographic Change on Human Capital Accumulation. *Lab. Econ.* 16:659-668. doi: 10.1016/j.labeco.2009.08.005.
- Freeman R (1979) The Effect of Demographic Factors on Age-Earnings Profiles. *J. Human Res.* 14:289-318. doi: 10.2307/145573.
- Garloff A, Pohl C, Schanne N (2013) Do Small Labor Market Entry Cohorts Reduce Unemployment? *Demographic Res.* 29:379-406. doi: 10.4054/DemRes.2013.29.15.
- Gregg P, Tominey E (2005) The Wage Scar from Male Youth Unemployment. *Lab. Econ.* 12:487-509. doi: 10.1016/j.labeco.2005.05.004.
- Gregory T, Patuelli R (2013) Regional Age Structure, Human Capital and Innovation - Is Demographic Ageing Increasing Regional Disparities? ZEW Discussion Papers 13-057.
- Korenman S, Neumark D (2000) Cohort Crowding and Youth Labor Markets: A Cross-National Analysis. *In: Blanchflower DG & Freeman RB (eds.) Youth Employment and Joblessness in Advanced Countries*. University of Chicago Press. Chicago
- Michaelis J, Debus M (2011) Wage and (Un-)Employment Effects of an Ageing Workforce. *J. Population Econ.* 24:1493-1511. doi: 10.1007/s00148-010-0325-9.
- Moffat J, Roth D (2013) The Cohort Size-Wage Relationship in Europe. MAGKS Papers on Economics 46-2013.
- Mroz TA, Savage TH (2006) The Long-Term Effects of Youth Unemployment. *J. Human Res.* 41:259-293. doi: 10.2307/40057276.
- Schmidt C (1993) Ageing and Unemployment. *In: Johnson P & Zimmerman K (eds.) Labour Markets in an Ageing Europe*. Cambridge University Press. Cambridge
- Skirbekk V (2003) Age and Individual Productivity: A Literature Survey. MPIDR Working Paper 2003-098.
- Stapleton D, Young D (1988) Educational Attainment and Cohort Size. *J. Lab. Econ.* 6:330-361. doi: 10.2307/2534828.
- Welch F (1979) Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust. *J. Polit. Economy* 87:S65-S97. doi: 10.2307/1829909.
- Wright R (1991) Cohort Size and Earnings in Great Britain. *J. Population Econ.* 4:295-305. doi: 10.1007/bf02426373.
- Zimmermann K (1991) Ageing and the Labor Market. *J. Population Econ.* 4:177-200. doi: 10.1007/bf00602428.

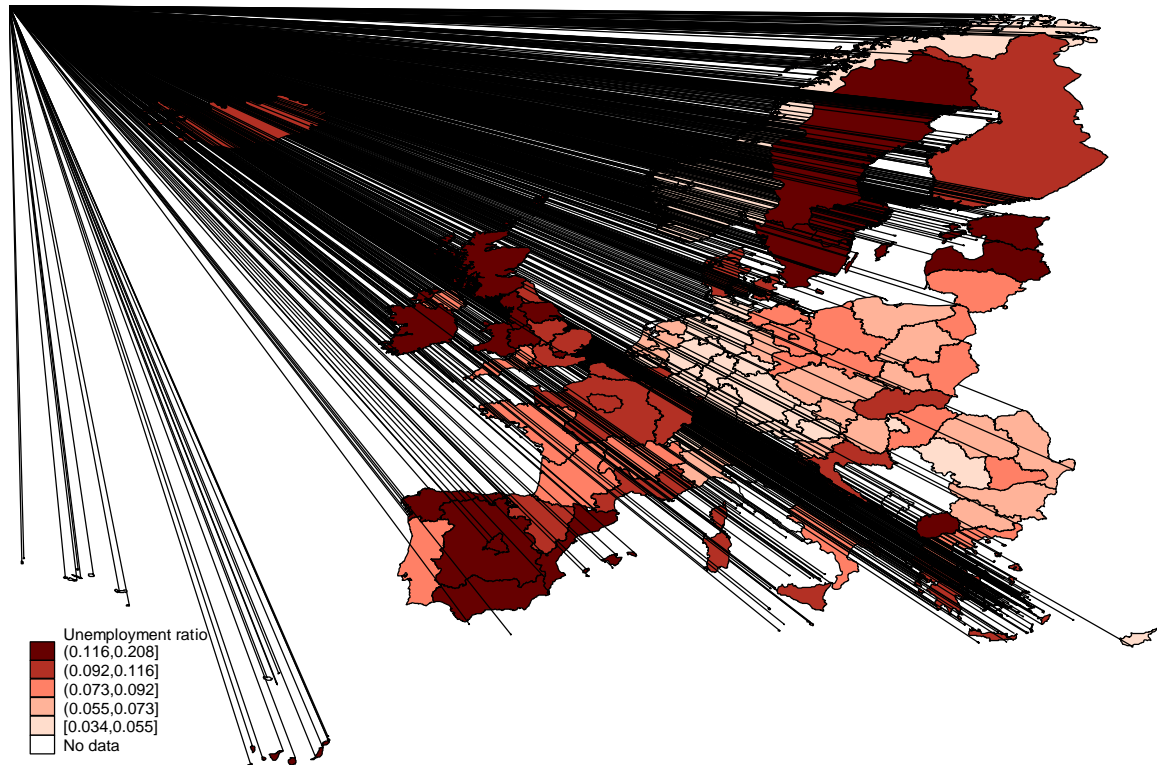
## Appendix

Figure A1a: The 2010 youth unemployment ratio (country level)



Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

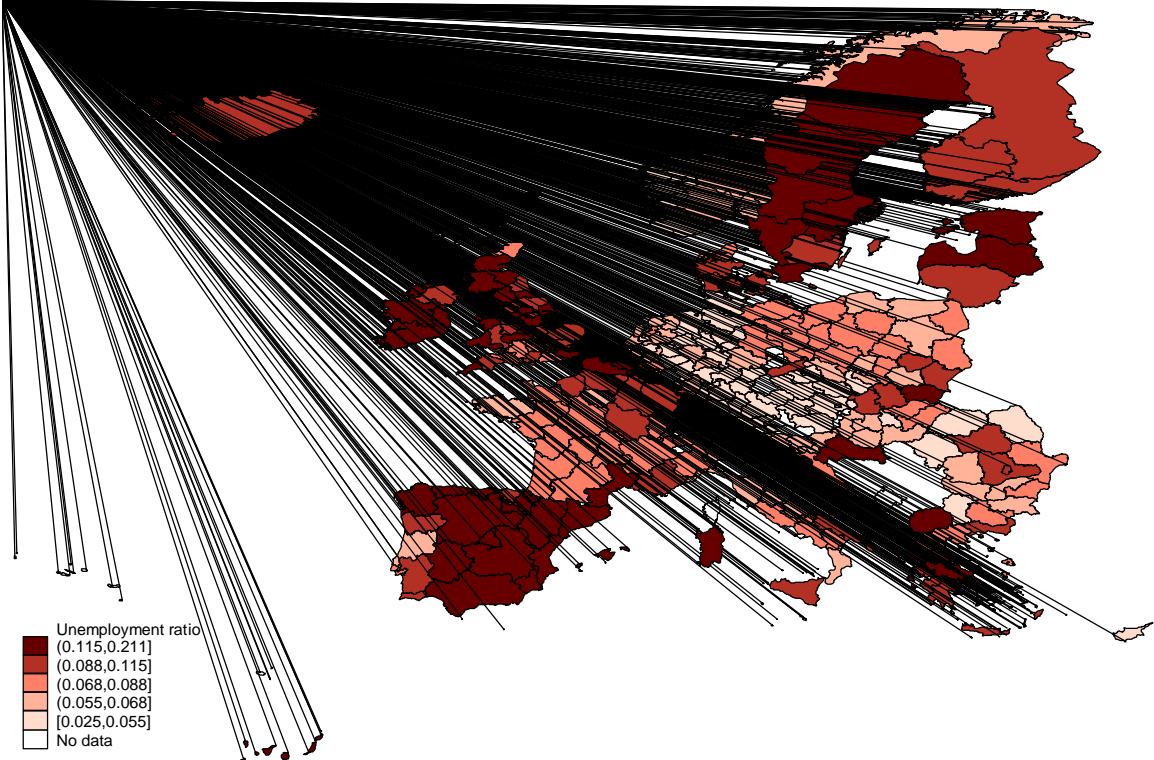
Figure A1b: The 2010 youth unemployment ratio (NUTS 1 level)



Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

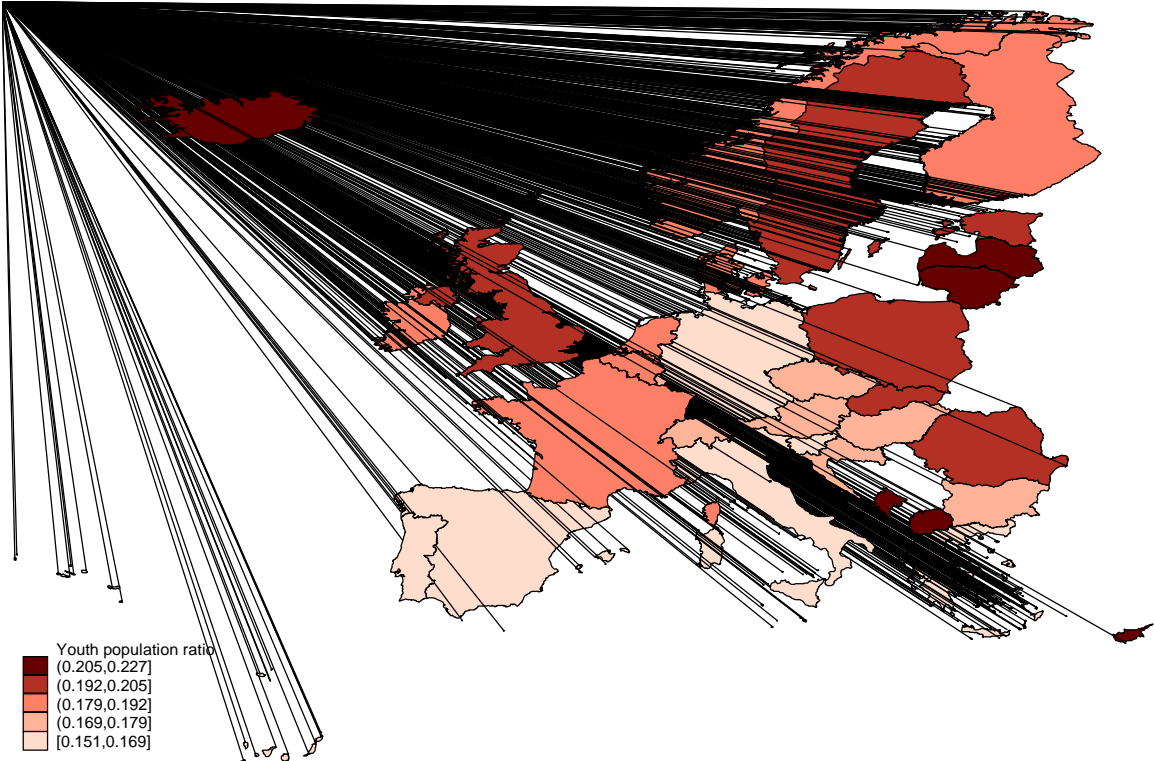


**Figure A1c: The 2010 youth unemployment ratio (NUTS 2 level)**



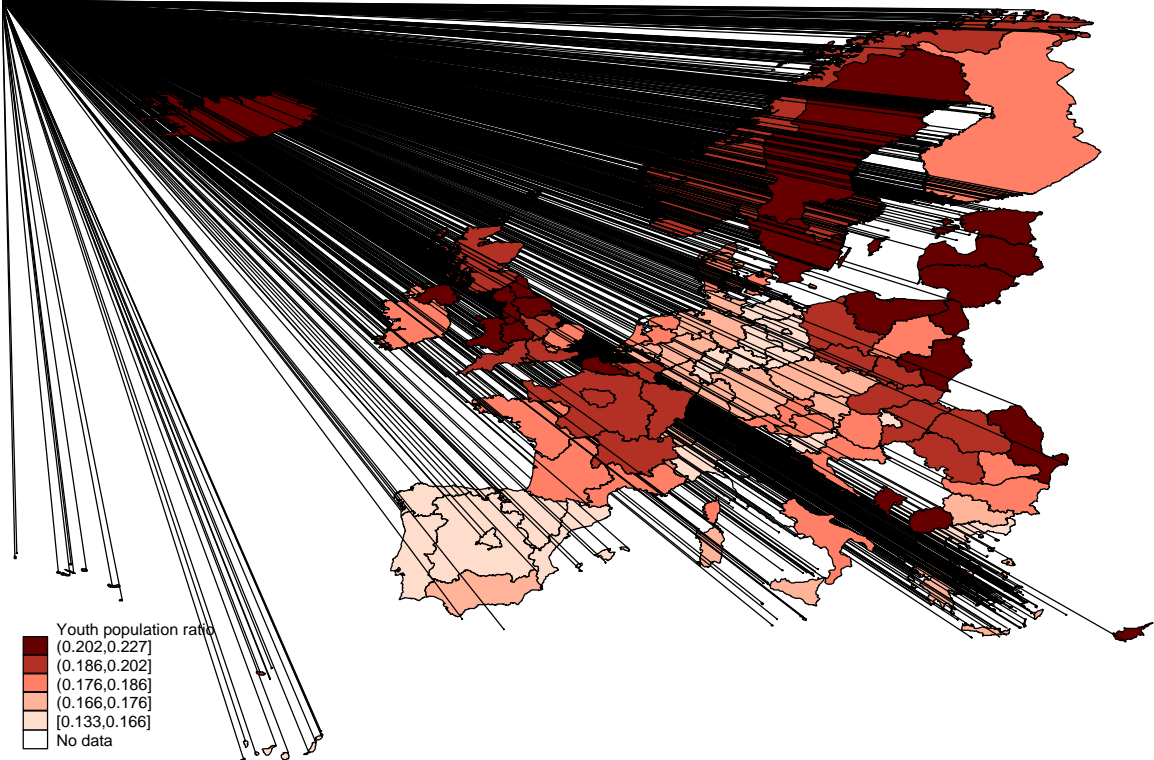
Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

**Figure A2a: The 2010 youth population ratio (country level)**



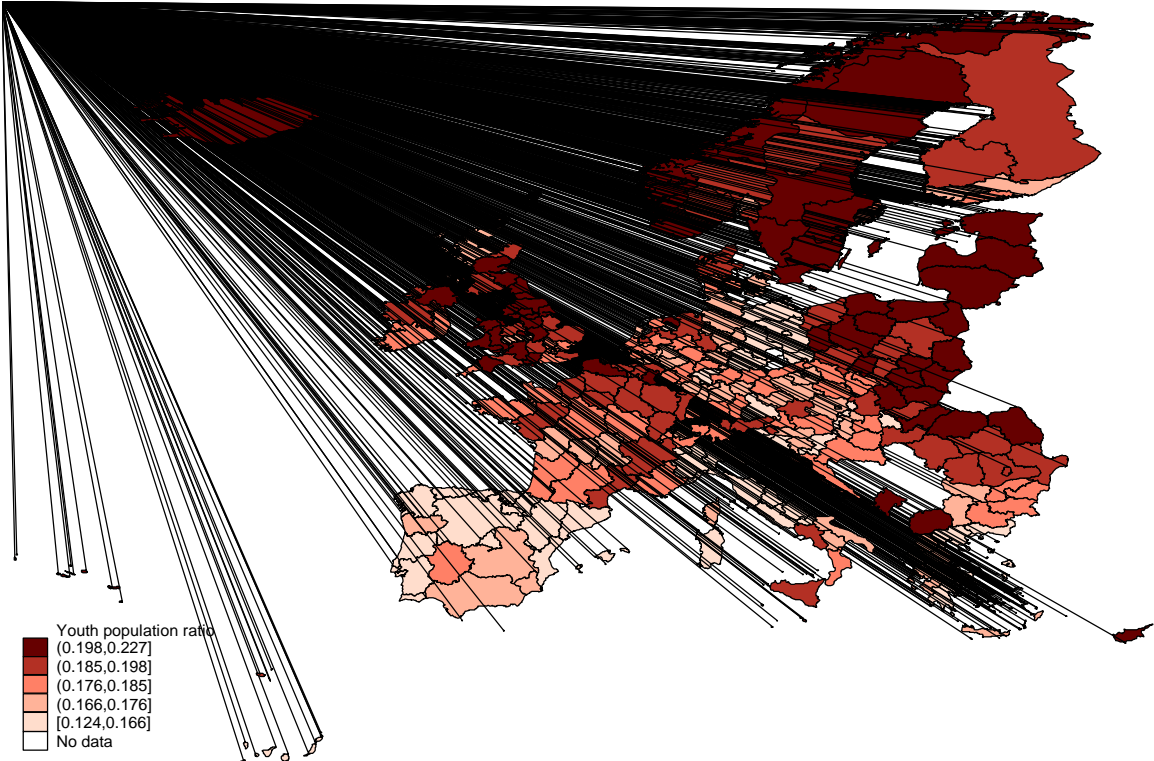
Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

**Figure A2b: The 2010 youth population ratio (NUTS 1 level)**



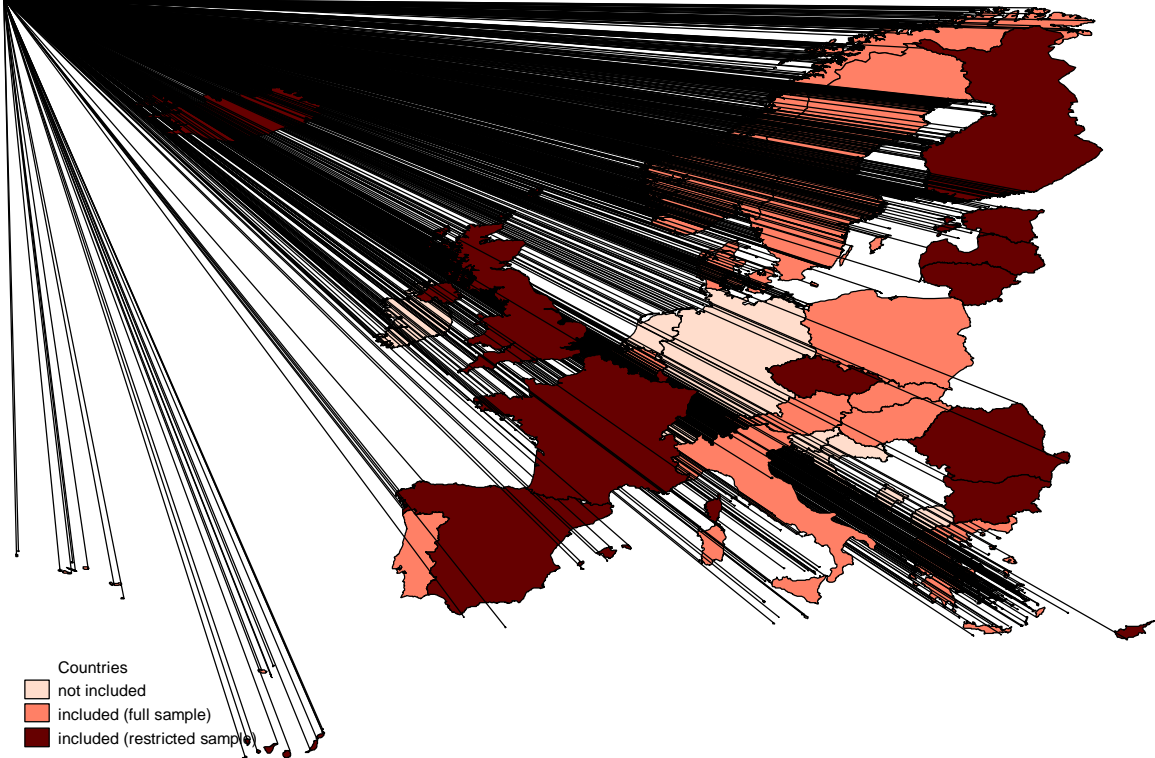
Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

**Figure A2c: The 2010 youth population ratio (NUTS 2 level)**



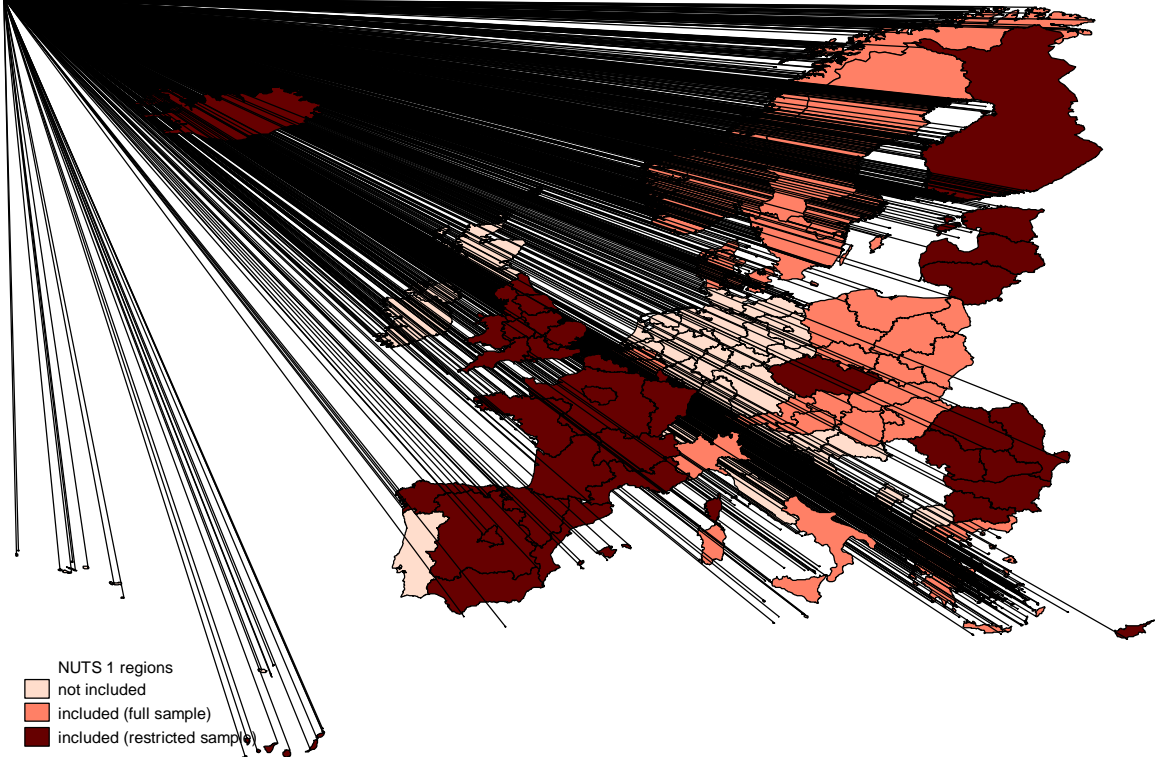
Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

Figure A3a: Countries included in the analysis



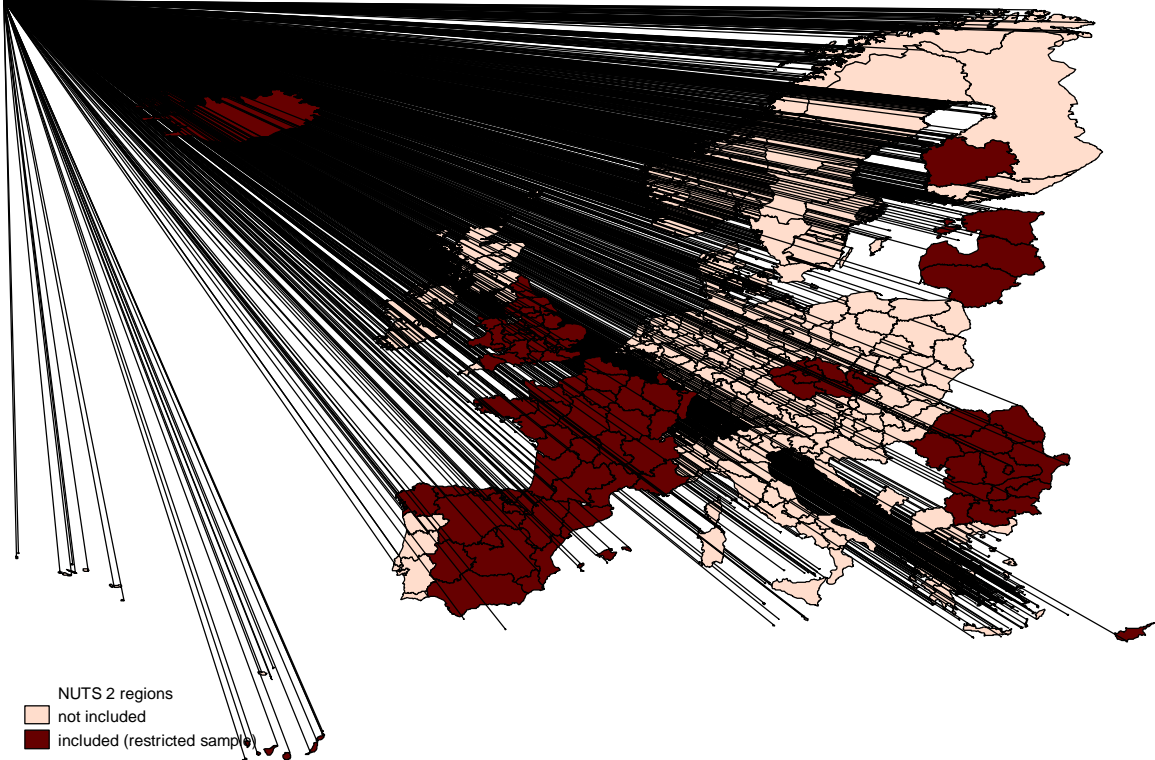
Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

Figure A3b: NUTS 1 regions included in the analysis



Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

Figure A3c: NUTS 2 regions included in the analysis



Source: Eurostat (© EuroGeographics for the administrative boundaries), own calculations

**Table A1: Definition of variables**

| Variable name                    | Description  | Source   |
|----------------------------------|--|----------|
| Unemp <sub>ijrt</sub>            | Equals 1 if the individual is unemployed and 0 otherwise   | EU-SILC  |
| CS <sub>jrt</sub>                | Size of an age-specific group relative to the total population (weighted to include individuals up to two years older and younger)   | EU-SILC  |
| CS_ins_cage <sub>jrt</sub>       | The number of individuals aged ten in the year when the reference age group was ten relative to the total population in that year (weighted to include individuals up to two years older and younger)            | Eurostat |
| CS_ins_clag <sub>jrt</sub>       | The number of individuals fourteen years younger than the reference age group fourteen years ago relative to the total population in that year (weighted to include individuals up to 2 years older and younger) | Eurostat |
| Male <sub>ijrt</sub>             | Equals 1 if the individual is male and 0 otherwise   | EU-SILC  |
| Married <sub>ijrt</sub>          | Equals 1 if the individual is married and 0 otherwise  | EU-SILC  |
| Age19 <sub>ijrt</sub>            | Equals 1 if the individual is 19 years old and 0 otherwise (base category: age 18)   | EU-SILC  |
| Age20 <sub>ijrt</sub>            | Equals 1 if the individual is 20 years old and 0 otherwise   | EU-SILC  |
| Age21 <sub>ijrt</sub>            | Equals 1 if the individual is 21 years old and 0 otherwise   | EU-SILC  |
| Age22 <sub>ijrt</sub>            | Equals 1 if the individual is 22 years old and 0 otherwise   | EU-SILC  |
| Age23 <sub>ijrt</sub>            | Equals 1 if the individual is 23 years old and 0 otherwise   | EU-SILC  |
| Age24 <sub>ijrt</sub>            | Equals 1 if the individual is 24 years old and 0 otherwise   | EU-SILC  |
| Age25 <sub>ijrt</sub>            | Equals 1 if the individual is 25 years old and 0 otherwise   | EU-SILC  |
| Isced1 <sub>ijrt</sub>           | Equals 1 if the individual has a completed primary education and 0 otherwise (base category isced0: pre-primary education)   | EU-SILC  |
| Isced2 <sub>ijrt</sub>           | Equals 1 if the individual has a lower secondary education and 0 otherwise   | EU-SILC  |
| Isced3 <sub>ijrt</sub>           | Equals 1 if the individual has a completed secondary education and 0 otherwise   | EU-SILC  |
| Isced4 <sub>ijrt</sub>           | Equals 1 if the individual has a post-secondary, non-tertiary education and 0 otherwise  | EU-SILC  |
| Isced5 <sub>ijrt</sub>           | Equals 1 if the individual has a completed tertiary education  | EU-SILC  |
| Urban2 <sub>ijrt</sub>           | Equals 1 if the individual resides in an intermediately populated area and 0 otherwise (base category urban1: densely populated area)  | EU-SILC  |
| Urban3 <sub>ijrt</sub>           | Equals 1 if the individual resides in a thinly populated area and 0 otherwise  | EU-SILC  |
| Totalunemp_country <sub>ct</sub> | The national unemployment rate   | Eurostat |
| Totalunemp_nuts1 <sub>rt</sub>   | The regional NUTS 1 unemployment rate (only included when cohort size is constructed at the NUTS 1 level)  | Eurostat |
| Totalunemp_nuts2 <sub>rt</sub>   | The regional NUTS 2 unemployment rate (only included when cohort size is constructed at the NUTS 2 level)  | Eurostat |

**Table A2: Descriptive statistics**

| <b>Variable name</b>  | <b>Observations</b> | <b>Mean</b> | <b>Standard deviation</b> | <b>Minimum</b> | <b>Maximum</b> |
|-----------------------|---------------------|-------------|---------------------------|----------------|----------------|
| Unemp                 | 45,623              | 0.1047      | 0.3062                    | 0.0000         | 1.0000         |
| CS (country)          | 45,623              | 0.0200      | 0.0029                    | 0.0142         | 0.0258         |
| CS (NUTS 1)           | 45,623              | 0.0201      | 0.0032                    | 0.0089         | 0.0313         |
| CS (NUTS 2)           | 45,623              | 0.0204      | 0.0036                    | 0.0070         | 0.0489         |
| CS_ins_cage (country) | 45,623              | 0.0204      | 0.0029                    | 0.0145         | 0.0271         |
| CS_ins_cage (NUTS 1)  | 45,623              | 0.0204      | 0.0032                    | 0.0119         | 0.0271         |
| CS_ins_cage (NUTS 2)  | 45,623              | 0.0205      | 0.0032                    | 0.0105         | 0.0280         |
| CS_ins_clag (country) | 45,623              | 0.0206      | 0.0031                    | 0.0147         | 0.0277         |
| CS_ins_clag (NUTS 1)  | 45,623              | 0.0205      | 0.0033                    | 0.0117         | 0.0277         |
| CS_ins_clag (NUTS 2)  | 45,623              | 0.0206      | 0.0033                    | 0.0102         | 0.0278         |
| Male                  | 45,623              | 0.5139      | 0.4998                    | 0.0000         | 1.0000         |
| Married               | 45,623              | 0.0543      | 0.2266                    | 0.0000         | 1.0000         |
| Age                   | 45,623              | 21.2840     | 2.2781                    | 18.0000        | 25.0000        |
| Age18                 | 45,623              | 0.1405      | 0.3476                    | 0.0000         | 1.0000         |
| Age19                 | 45,623              | 0.1378      | 0.3447                    | 0.0000         | 1.0000         |
| Age20                 | 45,623              | 0.1370      | 0.3439                    | 0.0000         | 1.0000         |
| Age21                 | 45,623              | 0.1286      | 0.3348                    | 0.0000         | 1.0000         |
| Age22                 | 45,623              | 0.1220      | 0.3272                    | 0.0000         | 1.0000         |
| Age23                 | 45,623              | 0.1144      | 0.3183                    | 0.0000         | 1.0000         |
| Age24                 | 45,623              | 0.1109      | 0.3140                    | 0.0000         | 1.0000         |
| Age25                 | 45,623              | 0.1087      | 0.3113                    | 0.0000         | 1.0000         |
| Isced                 | 45,623              | 2.8757      | 0.9822                    | 0.0000         | 5.0000         |
| Isced0                | 45,623              | 0.0028      | 0.0533                    | 0.0000         | 1.0000         |
| Isced1                | 45,623              | 0.0382      | 0.1916                    | 0.0000         | 1.0000         |
| Isced2                | 45,623              | 0.2971      | 0.4570                    | 0.0000         | 1.0000         |
| Isced3                | 45,623              | 0.5277      | 0.4992                    | 0.0000         | 1.0000         |
| Isced4                | 45,623              | 0.0106      | 0.1026                    | 0.0000         | 1.0000         |
| Isced5                | 45,623              | 0.1235      | 0.3290                    | 0.0000         | 1.0000         |
| Urbanisation          | 45,623              | 1.9654      | 0.9147                    | 1.0000         | 3.0000         |
| Urban1                | 45,623              | 0.4362      | 0.4959                    | 0.0000         | 0.0000         |
| Urban2                | 45,623              | 0.1622      | 0.3687                    | 0.0000         | 0.0000         |
| Urban3                | 45,623              | 0.4016      | 0.4902                    | 0.0000         | 0.0000         |
| Totalunemp_country    | 45,623              | 9.6193      | 4.9817                    | 2.3000         | 20.1000        |
| Totalunemp_nuts1      | 45,623              | 9.5370      | 5.2732                    | 2.3000         | 28.7000        |
| Totalunemp_nuts2      | 45,623              | 9.5946      | 5.3753                    | 1.9000         | 28.7000        |

**Table A3: Estimated OLS and GMM coefficients of equation (2) at the country, NUTS 1 and NUTS 2 level (full sample)**

| Dependent variable:<br><i>Unemp</i>                       | Country             |                      | NUTS 1              |                      | NUTS 2              |                      |
|---|---------------------|----------------------|---------------------|----------------------|---------------------|----------------------|
|   | OLS                 | GMM                  | OLS                 | GMM                  | OLS                 | GMM                  |
| CS  | 1.482<br>(1.170)    | 7.311***<br>(1.425)  | 0.026<br>(0.972)    | 8.417***<br>(1.630)  | -0.891<br>(1.002)   | 11.704***<br>(3.270) |
| Male  | 0.017***<br>(0.005) | 0.017***<br>(0.005)  | 0.019***<br>(0.005) | 0.019***<br>(0.005)  | 0.017**<br>(0.007)  | 0.016**<br>(0.007)   |
| Married   | -0.018**<br>(0.009) | -0.020**<br>(0.009)  | -0.017*<br>(0.010)  | -0.019*<br>(0.010)   | -0.025*<br>(0.015)  | -0.026*<br>(0.015)   |
| Age19   | 0.063***<br>(0.006) | 0.059***<br>(0.006)  | 0.065***<br>(0.006) | 0.060***<br>(0.006)  | 0.059***<br>(0.008) | 0.052***<br>(0.009)  |
| Age20   | 0.098***<br>(0.008) | 0.097***<br>(0.008)  | 0.102***<br>(0.008) | 0.100***<br>(0.008)  | 0.097***<br>(0.011) | 0.096***<br>(0.011)  |
| Age21   | 0.104***<br>(0.009) | 0.105***<br>(0.009)  | 0.108***<br>(0.009) | 0.110***<br>(0.009)  | 0.103***<br>(0.012) | 0.109***<br>(0.012)  |
| Age22   | 0.103***<br>(0.008) | 0.106***<br>(0.009)  | 0.106***<br>(0.009) | 0.111***<br>(0.009)  | 0.101***<br>(0.012) | 0.112***<br>(0.012)  |
| Age23   | 0.112***<br>(0.009) | 0.115***<br>(0.009)  | 0.116***<br>(0.009) | 0.123***<br>(0.010)  | 0.112***<br>(0.012) | 0.130***<br>(0.013)  |
| Age24   | 0.120***<br>(0.009) | 0.124***<br>(0.009)  | 0.126***<br>(0.009) | 0.134***<br>(0.010)  | 0.125***<br>(0.013) | 0.147***<br>(0.013)  |
| Age25   | 0.110***<br>(0.009) | 0.113***<br>(0.009)  | 0.114***<br>(0.009) | 0.120***<br>(0.009)  | 0.113***<br>(0.012) | 0.132***<br>(0.013)  |
| Isced1  | 0.151**<br>(0.073)  | 0.148**<br>(0.073)   | 0.141*<br>(0.074)   | 0.143*<br>(0.076)    | 0.175**<br>(0.079)  | 0.180**<br>(0.079)   |
| Isced2  | 0.029<br>(0.071)    | 0.026<br>(0.071)     | 0.025<br>(0.072)    | 0.028<br>(0.074)     | 0.044<br>(0.076)    | 0.047<br>(0.076)     |
| Isced3  | -0.052<br>(0.070)   | -0.056<br>(0.071)    | -0.058<br>(0.072)   | -0.056<br>(0.074)    | -0.060<br>(0.075)   | -0.057<br>(0.075)    |
| Isced4  | -0.048<br>(0.071)   | -0.052<br>(0.071)    | -0.052<br>(0.072)   | -0.052<br>(0.074)    | -0.048<br>(0.078)   | -0.053<br>(0.079)    |
| Isced5  | -0.091<br>(0.070)   | -0.092<br>(0.071)    | -0.098<br>(0.072)   | -0.094<br>(0.073)    | -0.105<br>(0.074)   | -0.099<br>(0.074)    |
| Urban2  | -0.001<br>(0.006)   | -0.001<br>(0.006)    | -0.004<br>(0.007)   | -0.004<br>(0.007)    | -0.010<br>(0.010)   | -0.011<br>(0.010)    |
| Urban3  | 0.015***<br>(0.006) | 0.015***<br>(0.006)  | 0.010<br>(0.006)    | 0.010*<br>(0.006)    | 0.004<br>(0.009)    | 0.005<br>(0.010)     |
| Totalunemp_country  | 0.011***<br>(0.001) | 0.012***<br>(0.001)  | 0.004<br>(0.004)    | 0.005<br>(0.004)     | 0.013***<br>(0.004) | 0.018***<br>(0.004)  |
| Totalunemp_nuts1  | -                   | -                    | 0.006*<br>(0.004)   | 0.007*<br>(0.004)    | -                   | -                    |
| Totalunemp_nuts2  | -                   | -                    | -                   | -                    | -0.003<br>(0.004)   | -0.005<br>(0.004)    |
| Constant  | -0.086<br>(0.075)   | -0.197**<br>(0.077)  | -0.044<br>(0.077)   | -0.203**<br>(0.080)  | 0.132<br>(0.086)    | -0.144<br>(0.108)    |
| N   | 96,178              | 96,178               | 89,181              | 89,181               | 45,623              | 45,623               |
| R <sup>2</sup>  | 0.040               | 0.040                | 0.053               | 0.049                | 0.077               | 0.062                |
| Underidentification test <sup>a</sup>                     | -                   | 6,204.774<br>(0.000) | -                   | 2,818.869<br>(0.000) | -                   | 507.508<br>(0.000)   |
| Weak identification test <sup>b</sup>                     | -                   | 7,817.691            | -                   | 2,341.367            | -                   | 336.006              |
| Hansen's J statistic<br>(overidentification) <sup>c</sup> | -                   | 1.803<br>(0.179)     | -                   | 0.136<br>(0.713)     | -                   | 1.324<br>(0.250)     |

Cluster-robust standard errors are given in parentheses, \*/\*\*/\*\* denote significance at the 10%/5%/1% level

<sup>a</sup> Kleibergen-Paap rank LM statistic (p-value in brackets)

<sup>b</sup> Kleibergen-Paap rank Wald F statistic (critical values are available only for i.i.d. errors)

<sup>c</sup> Overidentification test of all instruments

**Table A4: Marginal effects of cohort size at the country, NUTS 1 and NUTS 2 level (full sample)**

| <b>Spatial unit</b> | <b>Marginal effect</b> | <b>Standard error</b> | <b>p-value</b> | <b>95% confidence interval</b> |
|---------------------|------------------------|-----------------------|----------------|--------------------------------|
| Country             | 0.0234                 | 0.0046                | 0.000          | [0.0145; 0.0323]               |
| NUTS 1              | 0.0269                 | 0.0052                | 0.000          | [0.0167; 0.0372]               |
| NUTS 2              | 0.0375                 | 0.0105                | 0.000          | [0.0170; 0.0580]               |



**Table A5: Estimated OLS and GMM coefficients of Equation (2) including country-by-year dummies at the country, NUTS 1 and NUTS 2 level (restricted sample)**

| Dependent variable:<br><i>Unemp</i>                       | Country             |                      | NUTS 1              |                      | NUTS 2              |                      |
|---|---------------------|----------------------|---------------------|----------------------|---------------------|----------------------|
|   | OLS                 | GMM                  | OLS                 | GMM                  | OLS                 | GMM                  |
| CS  | -1.320<br>(1.887)   | 5.800**<br>(2.267)   | -1.787<br>(1.381)   | 9.719***<br>(3.050)  | -0.838<br>(1.042)   | 10.568***<br>(3.190) |
| Male  | 0.016**<br>(0.007)  | 0.016**<br>(0.007)   | 0.016**<br>(0.007)  | 0.016**<br>(0.007)   | 0.017**<br>(0.007)  | 0.016**<br>(0.007)   |
| Married   | -0.030*<br>(0.015)  | -0.032**<br>(0.015)  | -0.028*<br>(0.015)  | -0.027*<br>(0.015)   | -0.026*<br>(0.015)  | -0.026*<br>(0.015)   |
| Age19   | 0.060***<br>(0.009) | 0.055***<br>(0.009)  | 0.060***<br>(0.009) | 0.055***<br>(0.009)  | 0.059***<br>(0.008) | 0.052***<br>(0.009)  |
| Age20   | 0.100***<br>(0.011) | 0.092***<br>(0.011)  | 0.100***<br>(0.011) | 0.099***<br>(0.011)  | 0.098***<br>(0.011) | 0.097***<br>(0.011)  |
| Age21   | 0.105***<br>(0.012) | 0.098***<br>(0.012)  | 0.106***<br>(0.012) | 0.111***<br>(0.012)  | 0.103***<br>(0.012) | 0.108***<br>(0.012)  |
| Age22   | 0.105***<br>(0.012) | 0.102***<br>(0.012)  | 0.105***<br>(0.012) | 0.116***<br>(0.012)  | 0.102***<br>(0.012) | 0.112***<br>(0.012)  |
| Age23   | 0.115***<br>(0.013) | 0.113***<br>(0.012)  | 0.115***<br>(0.012) | 0.132***<br>(0.013)  | 0.112***<br>(0.012) | 0.129***<br>(0.013)  |
| Age24   | 0.129***<br>(0.012) | 0.129***<br>(0.012)  | 0.128***<br>(0.013) | 0.148***<br>(0.014)  | 0.126***<br>(0.013) | 0.146***<br>(0.013)  |
| Age25   | 0.118***<br>(0.012) | 0.116***<br>(0.012)  | 0.117***<br>(0.012) | 0.134***<br>(0.013)  | 0.114***<br>(0.012) | 0.130***<br>(0.013)  |
| Isced1  | 0.181**<br>(0.089)  | 0.171*<br>(0.089)    | 0.168*<br>(0.087)   | 0.175**<br>(0.089)   | 0.178**<br>(0.079)  | 0.183**<br>(0.079)   |
| Isced2  | 0.043<br>(0.085)    | 0.026<br>(0.085)     | 0.039<br>(0.083)    | 0.046<br>(0.085)     | 0.046<br>(0.076)    | 0.050<br>(0.075)     |
| Isced3  | -0.067<br>(0.084)   | -0.075<br>(0.084)    | -0.070<br>(0.082)   | -0.063<br>(0.084)    | -0.058<br>(0.075)   | -0.054<br>(0.074)    |
| Isced4  | -0.055<br>(0.088)   | -0.061<br>(0.088)    | -0.057<br>(0.085)   | -0.054<br>(0.088)    | -0.041<br>(0.078)   | -0.040<br>(0.078)    |
| Isced5  | -0.118<br>(0.084)   | -0.119<br>(0.084)    | -0.118<br>(0.082)   | -0.110<br>(0.084)    | -0.104<br>(0.074)   | -0.097<br>(0.074)    |
| Urban2  | -0.001<br>(0.009)   | 0.001<br>(0.009)     | -0.006<br>(0.010)   | -0.007<br>(0.010)    | -0.010<br>(0.010)   | -0.011<br>(0.010)    |
| Urban3  | 0.009<br>(0.009)    | 0.011<br>(0.009)     | 0.004<br>(0.009)    | 0.004<br>(0.009)     | 0.004<br>(0.009)    | 0.005<br>(0.010)     |
| Totalunemp_country  | 0.024***<br>(0.003) | -0.177***<br>(0.045) | 0.018***<br>(0.005) | -0.217***<br>(0.047) | 0.026***<br>(0.005) | 0.074***<br>(0.015)  |
| Totalunemp_nuts1  | -                   | -                    | 0.006<br>(0.004)    | 0.008*<br>(0.004)    | -                   | -                    |
| Totalunemp_nuts2  | -                   | -                    | -                   | -                    | -0.003<br>(0.004)   | -0.004<br>(0.004)    |
| Constant  | 0.063<br>(0.091)    | 1.320***<br>(0.329)  | 0.088<br>(0.089)    | 1.447***<br>(0.333)  | 0.132<br>(0.085)    | -0.421***<br>(0.153) |
| N   | 45,623              | 45,623               | 45,623              | 45,623               | 45,623              | 45,623               |
| R <sup>2</sup>  | 0.058               | 0.057                | 0.066               | 0.060                | 0.079               | 0.067                |
| Underidentification test <sup>a</sup>                     | -                   | 2,313.319<br>(0.000) | -                   | 974.281<br>(0.000)   | -                   | 601.596<br>(0.000)   |
| Weak identification test <sup>b</sup>                     | -                   | 5,729.100            | -                   | 718.308              | -                   | 401.354              |
| Hansen's J statistic<br>(overidentification) <sup>c</sup> | -                   | 17.758<br>(0.000)    | -                   | 0.114<br>(0.736)     | -                   | 0.793<br>(0.373)     |

Cluster-robust standard errors are given in parentheses, \*/\*\*/\*\* denote significance at the 10%/5%/1% level

<sup>a</sup> Kleibergen-Paap rank LM statistic (p-value in brackets)

<sup>b</sup> Kleibergen-Paap rank Wald F statistic (critical values are available only for i.i.d. errors)

<sup>c</sup> Overidentification test of all instruments

**Table A6: Regression results from alternative cohort size specifications (restricted sample)**

| Dependent variable:<br><i>Unemp</i>                       | Country           |                      | NUTS 1            |                     | NUTS 2             |                      |
|---|-------------------|----------------------|-------------------|---------------------|--------------------|----------------------|
|   | OLS               | GMM                  | OLS               | GMM                 | OLS                | GMM                  |
| CS_A  | -1.245<br>(1.512) | 6.118***<br>(2.121)  | -1.553<br>(1.075) | 9.930***<br>(2.988) | --0.361<br>(0.795) | 10.708***<br>(3.197) |
| N   | 45,623            | 45,623               | 45,623            | 45,623              | 45,623             | 45,623               |
| R <sup>2</sup>  | 0.057             | 0.055                | 0.065             | 0.054               | 0.077              | 0.058                |
| Underidentification test <sup>a</sup>                     | -                 | 1,186.453<br>(0.000) | -                 | 681.671<br>(0.000)  | -                  | 378.647<br>(0.000)   |
| Weak identification test <sup>b</sup>                     | -                 | 3,586.422            | -                 | 458.379             | -                  | 232.625              |
| Hansen's J statistic<br>(overidentification) <sup>c</sup> | -                 | 18.222<br>(0.000)    | -                 | 0.185<br>(0.667)    | -                  | 0.871<br>(0.351)     |
| CS_B  | -0.025<br>(1.081) | 5.478***<br>(1.905)  | -0.453<br>(0.695) | 8.518***<br>(2.700) | -0.082<br>(0.467)  | 9.231***<br>(2.934)  |
| N   | 45,623            | 45,623               | 45,623            | 45,623              | 45,623             | 45,623               |
| R <sup>2</sup>  | 0.057             | 0.055                | 0.065             | 0.049               | 0.077              | 0.042                |
| Underidentification test <sup>a</sup>                     | -                 | 1,259.063<br>(0.000) | -                 | 353.126<br>(0.000)  | -                  | 188.781<br>(0.000)   |
| Weak identification test <sup>b</sup>                     | -                 | 1,629.527            | -                 | 216.590             | -                  | 107.434              |
| Hansen's J statistic<br>(overidentification) <sup>c</sup> | -                 | 13.158<br>(0.003)    | -                 | 0.381<br>(0.537)    | -                  | 0.748<br>(0.387)     |

Cluster-robust standard errors are given in parentheses, \*/\*\*/\*\* denote significance at the 10%/5%/1% level

<sup>a</sup> Kleibergen-Paap rank LM statistic (p-value in brackets)

<sup>b</sup> Kleibergen-Paap rank Wald F statistic (critical values are available only for i.i.d. errors)

<sup>c</sup> Overidentification test of all instruments

**Table A7: Regression results for age range 16-25 (restricted sample)**

| Dependent variable:<br><i>Unemp</i>                       | Country           |                      | NUTS 1            |                      | NUTS 2            |                     |
|---|-------------------|----------------------|-------------------|----------------------|-------------------|---------------------|
|   | OLS               | GMM                  | OLS               | GMM                  | OLS               | GMM                 |
| CS  | -0.308<br>(1.303) | 6.139***<br>(1.548)  | -0.800<br>(1.010) | 8.805***<br>(1.946)  | -0.212<br>(0.771) | 8.858***<br>(2.021) |
| N   | 56,030            | 56,030               | 56,030            | 56,030               | 56,030            | 56,030              |
| R <sup>2</sup>  | 0.061             | 0.058                | 0.068             | 0.061                | 0.078             | 0.068               |
| Underidentification test <sup>a</sup>                     | -                 | 2,594.462<br>(0.000) | -                 | 1,262.152<br>(0.000) | -                 | 800.140<br>(0.000)  |
| Weak identification test <sup>b</sup>                     | -                 | 8,043.478            | -                 | 1,119.241            | -                 | 647.162             |
| Hansen's J statistic<br>(overidentification) <sup>c</sup> | -                 | 29.466<br>(0.000)    | -                 | 0.322<br>(0.570)     | -                 | 0.889<br>(0.346)    |

Cluster-robust standard errors are given in parentheses, \*/\*\*/\*\* denote significance at the 10%/5%/1% level

<sup>a</sup> Kleibergen-Paap rank LM statistic (p-value in brackets)

<sup>b</sup> Kleibergen-Paap rank Wald F statistic (critical values are available only for i.i.d. errors)

<sup>c</sup> Overidentification test of all instruments