Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 40-2014

John Moffat and Duncan Roth

Cohort size and youth labour-market outcomes: the role of measurement

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

Cohort size and youth labour-market outcomes: the role of measurement error

John Moffat[1] - Duncan Roth[2]

Abstract

Using data from 49 European regions covering 2005-2012, this paper finds that the estimated effect of cohort size on employment and unemployment outcomes is very sensitive to the age range of the sample. We argue that this is because the identification strategy commonly used in this literature is unable to eliminate the bias caused by measurement error in the cohort-size variable. The latter arises because large shares of the young choose to acquire education and consequently the size of an age group provides a poor measure of age-specific labour supply. In our view older age groups provide a more suitable sample to test the implications of cohort crowding since the former will have largely entered the labour market. Using a sample aged 25-29, which has relatively low rates of participation in education, we find robust evidence that an increase in cohort size increases employment and reduces unemployment.

JEL classification J10, J21, R23

Keywords Cohort size, cohort crowding, unemployment, employment, measurement error, EU-SILC

[1] Durham University, Department of Economics and Finance

[2] Institute for Employment Research (IAB), Josef-Gockeln-Straße 7, 40747 Düsseldorf, Germany; Phone: +49 (0)211 4306 108, Fax: +49 (0)211 4306 910468, Email: duncan.roth@iab.de

1 Introduction

The effect of the size of the youth population upon its labour-market prospects is of critical importance, particularly in light of demographic trends which will cause the youth share of the population to fall in most countries in coming decades (United Nations, 2015). The cohort-crowding hypothesis suggests that this will be beneficial for young individuals (Easterlin, 1961; Welch, 1979). By contrast, the model of Shimer (2001) implies that smaller youth cohorts will have a detrimental impact as firms create fewer jobs in areas with smaller youth shares. While the bulk of the empirical literature has focused on earnings and generally found negative effects of cohort size (e.g. Welch, 1979; Wright, 1991; Brunello, 2010; Moffat and Roth, 2016; Garloff and Roth, 2016), the effect on unemployment and employment has received less attention and the empirical evidence is so far mixed (Korenman and Neumark, 2000; Shimer, 2001; Skans, 2005; Foote, 2007; Biagi and Lucifora, 2008; Garloff et al., 2013).

In this paper, we propose that the standard identification strategy that has been used in the cohort-size literature does not allow for consistent estimation of the effect of cohort crowding for young age groups. There are two reasons for this, both of which are based on the observation that, due to high rates of participation in education, the relative size of an age group represents a poor measure of age-specific labour supply among the young, the latter being the relevant variable for age-specific employment and unemployment outcomes. First, since the proportion of young people that choose to defer entry to the labour market in order to acquire education may be influenced by cohort size (Fertig et al., 2009), this complicates the interpretation of estimated effects of cohort size since they reflect effects on participation and, conditional on participation, on (un-)employment. More importantly, the use of the number of individuals in an age group as the basis for the cohort-size variable creates measurement error that the standard instrumental variables (IV) approach to estimating the effects of cohort-size is unable to overcome.

We assess this argument by estimating the effect of cohort size on employment and unemployment shares using data from the longitudinal European Union Statistics on Income and Living Conditions (EU-SILC) survey which provides us with data on 49 regions for the period 2005-2012. Our results show that the estimated cohort-size effects are very sensitive to the chosen age range of the sample. Our preferred results come from a sample of individuals aged 25-29 since most of that group has entered the labour market and therefore the decision to participate in the labour market as well as the degree of measurement error are less of a concern. Among this group, we find, in contradiction of the cohort-crowding hypothesis, a negative effect of cohort size on the unemployment share. These results are robust to a variety of changes in the sample and in the empirical specification. This finding is relevant because it casts doubt on the conclusions from previous studies, which have defined the youth population as individuals aged 15/16-24, regarding the relationship between the size of the youth population and its members' employment and unemployment outcomes.

Section 2 reviews the extant theoretical and empirical literature on the relationship between population structure and labour market outcomes. Section 3 discusses the dataset and empirical model. The results are presented in Section 4 and Section 5 concludes.

2 Literature review

Competing theoretical predictions and conflicting empirical evidence exist regarding the question of how changes in the size of an age group affect its (un-)employment prospects. The cohort-crowding hypothesis is based on the assumption that differently aged workers are only imperfectly substitutable due to differences in human capital (Welch, 1979) so that changes in the size of an age group have implications predominantly for members of that age group (see Moffat and Roth, 2016, for a more detailed discussion). In perfectly competitive labour markets, changes in age-group size would only be reflected in changes to age-specific wages. If labour markets are imperfectly competitive, however, wages need not be fully flexible and an increase in the size of an age group may lead to an increase in the unemployment rate of that group (a theoretical model of this relationship in imperfectly competitive markets is provided by Michaelis and Debus, 2011).

In line with the cohort-crowding hypothesis, Korenman and Neumark (2000) provide empirical evidence that large youth cohorts (measured as the ratio of individuals aged 15-24 to individuals aged 25-54) increase the youth unemployment rate. Their findings are robust to a number of specifications, including the use of lagged birth rates as an instrument for the potentially endogenous youth-share variable. Moreover, the use of cross-national variation in their dataset of Organisation for Economic Cooperation and Development countries allows the authors to separately identify the effects of changes in youth-cohort size from the effects of other macroeconomic developments and as such provides an improvement on earlier studies that relied solely on time-series variation (e.g. Zimmermann, 1991; Schmidt, 1993).

Rather different results are obtained by Shimer (2001). Using data on a panel of US states for the period 1970-1996, he finds that increases in the youth share – measured as the ratio of those aged 16-24 to those aged 16-64 – are associated with decreases in the state-level unemployment rate. This is surprising for two reasons: first, since the overall unemployment rate is the sum of age-specific unemployment rates weighted by the share of the respective age group in the labour force and the youth unemployment rate generally exceeds that of older individuals, the direct effect of an increase in the youth share should be to increase the overall rate. Second, according to the cohort-crowding hypothesis the indirect effect of an increase in the youth share should be to increase the youth unemployment rate, thereby reinforcing the direct effect. Shimer's (2001) empirical results, however, not only show a negative effect on the overall unemployment rate, but also that the youth share reduces the unemployment rate of youths as well as other age groups.

Shimer (2001) provides a theoretical foundation to his empirical findings in the form of a search and matching model with on-the-job search. Changes in the size of the youth population tend to be predictable, as evidenced by the explanatory power of lagged birth rates for the size of the current youth share. Moreover, young individuals are more often either without a job or less well matched than older individuals and are therefore, on average, more willing to take up or switch jobs. This makes it easier for firms to make a productive match with workers in markets with a large number of potential employees. They therefore react to an expected change in the youth share by creating vacancies, to the benefit of all age groups.

Aiming to explain the substantial differences between his own and Korenman and Neumark's (2000) empirical findings, Shimer (2001) points out that the former ignored the possibility of changes in the youth share having an effect on the unemployment rate of other age groups. Specifically, Korenman and Neumark's (2000) model includes the adult unemployment rate, alongside the youth share, as a regressor in the model of the youth unemployment rate. According to Shimer (2001), if changes in the youth share affect the unemployment rates of both age groups, the former's coefficient will be biased upwards and he is able to show this using his own dataset. However, applying his empirical model to the data of Korenman and

Neumark (2000) produces inconclusive results, which casts doubt on the applicability of his theoretical model to other countries and time periods.

The small number of studies that have since looked at the relationship between age structures and unemployment outcomes have yielded mixed results. Using data on Swedish labour markets for the years 1985-1999, Skans (2005) finds no evidence for an effect of the relative size of the group aged 16-24 on the total unemployment rate, but his results are otherwise in line with Shimer (2001) since they show that the youth unemployment rate falls when the size of young age groups increases. In contrast, Foote (2007) shows that when the time dimension of Shimer's (2001) dataset is extended to 2005 the negative effect of the youth share on the overall unemployment rate decreases considerably and becomes insignificant in most specifications. The empirical evidence of Biagi and Lucifora (2008) also contradicts the findings of Shimer (2001): their analysis of a dataset of European countries spanning the late 1970s to the early 2000s suggests that the share of individuals aged 15-24 has a positive effect on the unemployment rate of the young and is not statistically significant for the unemployment outcomes of prime-age individuals. Finally, Garloff et al. (2013), using data on West German labour-market regions for the years 1993-2008, find that increases in the share of individuals aged 15-24 years are associated with increases in the overall unemployment rate.

In light of the conflicting results produced by previous studies this analysis provides new evidence on the relationship between age-group size and age-specific unemployment outcomes. Our dataset is a longitudinal sample of European regions covering 2005-2012 which provides us with more heterogeneity to separate the effects of cohort size from other influences than has generally been available in the literature. However, the paper's main contribution is to consider the effect of the definition of the youth population on the estimates obtained. The previous literature has used the share of individuals aged either 15-24 or 16-24 as a definition of the youth share. Since a high proportion of this group will be in education and therefore potentially unavailable to the labour market, this will, as discussed in the introduction and in more detail below, have important implications for both the interpretation and econometric identification of the cohort-size effect.

3 Empirical analysis

3.1 Data

The major part of the dataset that is used in the empirical analysis is constructed by combining different longitudinal EU-SILC releases.[1] Appending data from different releases not only allows the extension of the sample period beyond the four years provided by a single longitudinal release, but also increases the number of observations within a given year. In order to match observations from different releases that refer to the same individual, a unique personal identifier is constructed.[2] This is then used to verify that there are very few individuals with inconsistencies in age and sex over time[3] (see Moffat and Roth, 2016, for further details on the process of appending the different datasets and Berger and Schaffner, 2015, for general information about EU-SILC).

Individuals in EU-SILC are not randomly sampled and weights are therefore provided so that unbiased population estimates may be calculated. We use these to construct two new weighting variables: the first of these variables corrects the initial weights for the number of rotational groups within a country-year combination that change as a result of appending data from different releases (see Moffat and Roth, 2016). The second weighting variable also re-scales the weights so that the size of the estimated population within a region-year-age-sex cell is identical to the statistics reported by Eurostat.[4]

The so-constructed dataset contains 2.76 million observations on just over 1 million individuals and covers the years 2004-2013. In addition to the country that an individual resides in, EU-SILC provides information about the region of residence at the first level of the Nomenclature of Territorial Units in Statistics (NUTS). Availability of this information allows us

[1] The longitudinal releases are: 2013 (version 1 from 01-08-2015), 2012 (version 3 from 01-08-2015), 2011 (version 4 from 01-03-2015), 2010 (version 5 from 01-08-2014), 2009 (version 4 from 01-03-2013), 2008 (version 4 from 01-03-2012), 2007 (version 5 from 01-08-2011), 2006 (version 2 from 01-03-2009) and 2005 (version 1 from 15-09-07).
[2] This identifier is defined as a combination of an observation's identification number (which is not unique across countries), his country of residence and the rotational group to which he belongs.
[3] In total, there are 36 individuals (182 observations) with inconsistencies. All of these individuals are from France, Luxembourg or Norway (i.e. countries in which individuals can be followed for more than 4 years). For these individuals, the inconsistent observations are dropped. If there are only two observations per individual, both are dropped.
[4] Note that while the Eurostat statistics refer to 1 January of a given year, use of the variable *age at the end of the income reference period* ensures that the population sizes estimated from EU-SILC data refer to 31 December of the preceding year.

to construct the relevant variables at the regional rather than at the national level, which is attractive because estimates of functional labour markets have tended to show them to be defined at the sub-national level (see Moffat and Roth, 2016).

Rather than focussing on outcomes at the individual level, the empirical analysis in this paper is concerned with estimating the effect of age-specific cohort size on unemployment and employment outcomes at the level of the corresponding age group. For this reason, the dataset is aggregated to the level of region-year-age cells. The resulting dataset is further supplemented by variables taken from Eurostat's publicly available database[5]: the level of regional GDP and the size of relevant age groups between 1991 and 1998 which are used as instruments in the empirical analysis.[6]

Due to data limitations, observations from the following countries are dropped: Germany, the Netherlands and Portugal (information on NUTS1 regions is not provided); Croatia (lagged population data for the construction of the instrument is not available); Finland, Iceland and Slovenia (age-related variables are randomly perturbed to prevent disclosure); Ireland and the United Kingdom (the age variable is measured at a different time of year for these countries, see footnote 4). Moreover, we exclude observations from Bulgaria, Cyprus, Malta, Norway and Romania because the necessary variables are not available throughout the whole sample period. This leaves a panel of 49 NUTS1 regions from the following countries for which age groups can be observed from 2005-2012 (number of regions per country in parentheses): Austria (3), Belgium (3), Czech Republic (1), Denmark (1), Estonia (1), Greece (4), Spain (7), France (8), Hungary (3), Italy (5), Lithuania (1), Luxemburg (1), Latvia (1), Poland (6), Sweden (3), Slovakia (1).

3.2 Variables and sample

This section serves several purposes: first, it defines the main variables of the empirical model; second, it discusses the age range of the sample; finally, an illustration is provided of the variation in the cohort-size variable that is used for identification.

[5] The data can be obtained through the following link: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

[6] Due to a change in delineation lagged population data is not available before the year 2003 for the two regions ITH (Northeast Italy) and ITI (Central Italy). Since these changes are minor compared to the total size of the regions we instead use lagged age-group size based on the predecessor regions ITD and ITE, which we obtain from the homepage of the Italian Statistical Office (www.istat.it).

The analysis separately estimates the effect of changes in cohort size on the share of individuals in age group *j*, region *r* and year *t* that are unemployed ($unemp_{jrt}$) and employed (emp_{jrt}). As discussed in the previous section, these shares are derived from individual-level data. Specifically, the weighted sum of male individuals who report to be (un-)employed in a given region-year-age group is calculated and divided by the total male population in that cell. Female observations are excluded in order to avoid the results being affected by selected labour-market participation. As these variables are standardised on the population rather than the labour force, the outcome variables differ from the unemployment and the employment rate. An advantage of this specification is that any effects that changes in cohort size, if measured without error, might have on participation rates could be ignored in the interpretation of the results.

Figure A1 in the Appendix shows the development of the dependent variables $unemp_{jrt}$ and emp_{jrt} as well as of a similarly defined variable that shows the share of individuals reporting to be in education in a given age group ($educ_{jrt}$). These variables are plotted for the age group 18-29 in selected regions and years to illustrate the variation in age-specific labour-market outcomes across Europe. While there are differences in the slope of the profiles, a common feature of all region-year combinations is that the employment share tends to increase and the share of individuals in education decreases with age. In contrast, there is no obvious trend in the unemployment share. In order to understand the implications of the high share of young individuals in education, the empirical model is firstly estimated for overlapping five-year age groups (beginning with individuals aged 18-22 and ending with individuals aged 25-29). The reason for adopting this strategy is that for younger age groups the coefficients will capture the effect of cohort size on labour market participation and, conditional on participation, the effect on (un-)employment. If the decision to participate in the labour market is also affected by cohort size, the estimated effects on employment and unemployment would be confounded by the effect of cohort size on participation. Moreover, the existence of measurement error in the cohort-size variable among young age groups, as described further in Section 3.3, may also lead to biased estimates. We therefore focus on individuals aged 25-29 since the estimates for this group will be less susceptible to these problems since, as shown in Figure A1, the share of individuals in education has decreased substantially by that age.

Means and standard deviations of the three dependent variables are shown in the first two columns of Table 1 for the age range 25-29. On average 78% of individuals in a region-year-age group cell are employed compared to 13% that are unemployed. The three remaining columns provide an insight into whether these variables tend to vary most across regions, years or age groups. This is done by regressing each of the dependent variables on a set of dummy variables for two of the aforementioned dimensions and then comparing the adjusted R^2. Dummies for years and age groups explain only 14% of the variation in the employment share but this value increases considerably once region dummies are included, which suggests that most of the variation in this variable exists between regions. While the explanatory power of the dummy variables is generally lower, the between-region variation also appears to be largest for the unemployment share.

Table 1: Descriptive statistics (employment and unemployment share)

	Mean	Standard deviation	Adjusted R^2 (year, age)	Adjusted R^2 (region, age)	Adjusted R^2 (region, year)
Emp_{jrt}	0.777	0.156	0.136	0.459	0.394
$Unemp_{jrt}$	0.126	0.109	0.063	0.281	0.333

Means and standard deviations are weighted by the weight-adjusted number of individuals per region-year-age group cell. Adjusted R^2 is derived from a regression of the dependent variables on dummies for the indicated variables; the regression is weighted by the weight-adjusted number of individuals per region-year-age group cell.

The main explanatory variable measures age-specific cohort size which refers to the number of individuals in age group *j*, region *r* and year *t*, N_{jrt}, relative to the size of the population aged between 16 and 65, $N_{16\text{-}65,rt}$. While most studies instead use a measure of the youth share, e.g. the relative size of the age group 16-24, we choose a specification that also varies across age to better capture the assumption of imperfect substitutability across age groups which has been posited in theoretical models (Card and Lemieux, 2001). Since it seems overly restrictive to assume that individuals only compete with individuals of the same age, we adopt another specification that has been previously used in this literature (Wright, 1991; Brunello, 2010).[7] This defines the cohort-size variable as a weighted sum that takes into account the size of the age groups that are up to two years older or younger than the reference group as shown in Equation 1:

$$CS_{jrt} = \frac{(1/9)N_{j-2,rt}+(2/9)N_{j-1,rt}+(3/9)N_{jrt}+(2/9)N_{j+1,rt}+(1/9)N_{j+2,rt}}{N_{16-65,rt}} \quad [1]$$

[7] We show in the Supplementary Material that alternative specifications of the cohort-size variable, including unweighted sums across three and five age groups, yield comparable results to those shown in Table 3.

These quantities are estimated from the EU-SILC dataset by computing the weighted sum of male and female observations in the corresponding region-year-age cells. As they are not available to the labour market, individuals reporting to be either in the military or disabled or unfit to work are omitted but individuals reporting that they are in education are included (the implications of this are discussed in Section 3.3).

The size of an age group in a given region and year is not necessarily exogenous because individuals might react to contemporaneous economic shocks by migrating into regions that offer better economic prospects. If such self-selection takes place, cohort-size would be endogenous to the share of individuals that are (un-)employed and estimation by ordinary least squares (OLS) would yield an inconsistent estimate of the cohort-size effect. We address this issue by employing an IV strategy in which the cohort size of the age group that is fourteen years younger than the reference group as observed fourteen years earlier serves as an instrument. Identification strategies based on time-lagged and age-lagged instruments or, as a special case of the former, birth rates are common in this literature (Korenman and Neumark, 2000; Shimer, 2001; Skans, 2005; Garloff et al., 2013; Moffat and Roth, 2016).[8] Instruments of this type are appealing because a cohort that was relatively large (small) in the past is likely to remain large (small) in the present despite migration and natural population changes[9]:

$$CS\_Ins_{jrt} = \frac{(1/9)N_{j-16,r,t-14}+(2/9)N_{j-15,r,t-14}+(3/9)N_{j-14r,t-14}+(2/9)N_{j-13,r,t-14}+(1/9)N_{j-12,r,t-14}}{N_{2-51,r,t-14}} \quad [2]$$

Table 2 contains descriptive statistics on the cohort-size variable and its instrument. On average, the five-year weighted sum of an age group in the range 25-29 accounts for about

[8] If cohort-size effects are heterogeneous across age, region and/or time, 2SLS estimates a local average treatment effect (LATE) (Imbens and Angrist, 1994). This estimate is the weighted average of the region-year-age cell-specific effects of cohort size with the largest weights attached to cells for which the relationship between the instrument and cohort-size is strongest (Angrist and Imbens, 1995). Since the strength of the relationship between the instrument and cohort-size will be mainly determined by net migration, greater weight will be attached to cells with low levels of net migration. If immigrants are less attractive to employers as a result of having less country-specific human capital (Kim and Park, 2013) than individuals that lived in the region fourteen years ago, this suggests that the LATE will be more positive (more negative) in the employment (unemployment) model than the average treatment effect (ATE). 2SLS estimates may then be larger than OLS estimates of the cohort-size effects if this effect outweighs that of self-selection bias, which would tend to cause OLS to overestimate the positive (negative) effect on employment (unemployment).

[9] Further information on the instrument can be found in Moffat and Roth (2016), while the validity of time- and age-lagged instruments is discussed in Garloff and Roth (2016).

2% of the population aged between 16 and 65, while the value is slightly smaller in the case of the instrument. For both variables, the larger part of the variation exists between regions.

Table 2: Descriptive statistics (cohort-size variable and instrument)

	Mean	Standard deviation	Adjusted R^2 (year, age)	Adjusted R^2 (region, age)	Adjusted R^2 (region, year)
CS_{jrt}	0.021	0.003	0.073	0.749	0.778
CS_Ins_{jrt}	0.020	0.003	0.080	0.780	0.826

Means and standard deviations are weighted by the weight-adjusted number of individuals per region-year-age group cell. Adjusted R^2 is derived from a regression of the dependent variables on dummies for the indicated variables; the regression is weighted by the weight-adjusted number of individuals per region-year-age group cell.

Figures A2 and A3 plot the dependent variables and the cohort-size variable (depicted as the fitted value from a weighted regression on the instrument) across time and age groups, respectively, for the same set of regions as in Figure A1 and thereby illustrate the variation from which cohort-size effects can be identified. Variation over time for given combinations of regions and age groups can be seen in Figure A2; the chosen regions are representative of the larger parts of Europe to which they belong: in Western and Northern Europe (represented by regions BE2 and SE1), the cohort-size profiles are rather flat. In contrast, in region ES5 there is a clear decrease in cohort size over time which affects all age groups – similar profiles can be found in the remaining regions of Spain as well as in Greece and Italy. Finally, different types of profiles can be found in Eastern Europe: on the one hand, the decreasing trend in cohort size in region HU1 resembles the developments in Southern Europe, while on the other hand age groups have increased in size in the Baltic country Latvia. Figure A3 suggests that variation across age groups is less pronounced: older age groups tend to be larger in ES5 and HU1, but the differences become smaller in later years. The profiles in the remaining regions are comparatively flat. At the same time both figures also illustrate the variation in cohort size across regions for given years and age groups. For example, the share of older age groups is larger in regions ES5 and HU1 in earlier years, whereas younger cohorts are relatively big in LV0 at the end of the sample period. While the regression analysis in Section 4 makes use of variation across each of these dimensions, in the Appendix we show results that are obtained from a single source of variation.

3.3 Model

According to the theory outlined in the literature review, age-specific labour market outcomes are determined by the supply of age-specific labour. Therefore the effect of cohort size on the outcome variables is modelled as shown in Equation 3 where $share_{jrt}$ represents either the unemployment or employment share, CS^*_{jrt} represents measurement error-free cohort size (i.e. the size of the age cohort that is available to the labour market), x_{jrt} represents a vector of control variables and ε_{jrt} is an error term:

$$share_{jrt} = \alpha + \beta CS^*_{jrt} + \boldsymbol{x'_{jrt}\gamma} + \varepsilon_{jrt} \quad [3]$$

In addition to the problem of regional self-selection that is addressed by IV estimation, there is also a problem of measurement error. This has so far not been addressed in this literature. It arises because of the inclusion of individuals, many of whom will be in education, that are unavailable to the labour market in the cohort-size variable. Moreover, datasets usually do not allow distinguishing individuals that are committed to long-term educational programmes and therefore unavailable to the labour market from individuals in education that would enter the labour market if an attractive opportunity arose (Jones and Riddell, 2006; Moffat and Yoo, 2015). The existence of the latter group means that the alternative approach of excluding those in education from the cohort-size variable would not provide a solution to the measurement-error problem.[10] Formally, the relationship between the observable age-specific cohort-size variable CS_{jrt} and the unobservable measurement error-free variable can be represented as follows:

$$CS_{jrt} = CS^*_{jrt} + u_{jrt} \quad [4]$$

In Equation (4), u_{jrt} is the part of observed cohort size that is not available to the labour market (i.e. the measurement error). Rearranging and substituting Equation (4) into Equation (3) gives:

$$share_{jrt} = \alpha + \beta CS_{jrt} + \boldsymbol{x'_{jrt}\gamma} + \varepsilon_{jrt} - \beta u_{jrt} \quad [5]$$

If the measurement error is 'classical', there is no correlation between the error-free measure of cohort size and the measurement error and this leads to attenuation of the estimated effect of cohort size. However, empirical evidence suggests that members of large cohorts are less likely to acquire education (Fertig et al., 2009), which suggests the existence of a correlation

[10] In the Supplementary Material we provide the regression results from a model in which the numerator of the cohort-size variable is constructed from individuals reporting to be employed or unemployed. For the age group 25-29 the obtained results are very similar to those reported in Table 3. Using younger age groups produces a pattern of cohort-size coefficients which is close to the one in Figure 1 which suggests that exclusion of those reporting to be in education does not remove the problem of measurement error.

between the size of an age group CS_{jrt} and u_{jrt}. Arguably, the number of individuals who are available to the labour market is larger in larger age groups and therefore the correlation between the degree of measurement error and the observable cohort size also carries over to the latent variable CS^*_{jrt}, which measures the size of an age group that is available to the labour market. In this 'non-classical' case, it is not possible to state a priori the direction of bias since it will be dependent on the relative variances of CS^*_{jrt} and u_{jrt}, the size of the covariance of CS^*_{jrt} and u_{jrt} and the partial correlations between the measurement error and the dummy variables in the model (Bound et al., 2001).

A second reason for the existence of non-classical measurement error is given by the current demographic processes, as a result of which younger age groups tend to be smaller than older ones in a given region and year (support for this hypothesis is provided in the Supplementary Material). Moreover, given the assumption that the share of non-participants is larger in younger age groups – for which the substantially larger education shares in younger age groups provide some evidence – it is possible for the latent cohort-size variable and the degree of measurement error to be negatively correlated across age groups. This will be the case as long as the ratio of the non-participation share in younger and older groups exceeds the ratio of the size of older and younger groups (details on this argument are provided in the Supplementary Material).

While two-stage least squares (2SLS) estimation is one approach to tackling measurement error (Hausman, 2001), the instrument which is standard in the literature does not purge the correlation with u_{jrt}. The instrument is based on the size of the same cohort observed at an earlier point in time and since an age group that is relatively large in the present can be expected to have also been relatively large in the past, the instrument would also be correlated with the degree of measurement error. As a result, 2SLS will not provide a consistent estimate of the cohort-size effect.

For the sample of individuals aged 25-29, the empirical analysis is based on 1,959 region-year-age cells[11]. Two specifications of Equation 5 are estimated for each of the outcome variables. Analogously to the use of control variables in Shimer (2001), in the baseline specification

[11] In principle, 5 age groups (25-29) are observed in 49 regions for 8 years (2005-2012), but since there are no observations for age group 26 in region FR1 and year 2010 in the sample, the total number of observations is reduced by one.

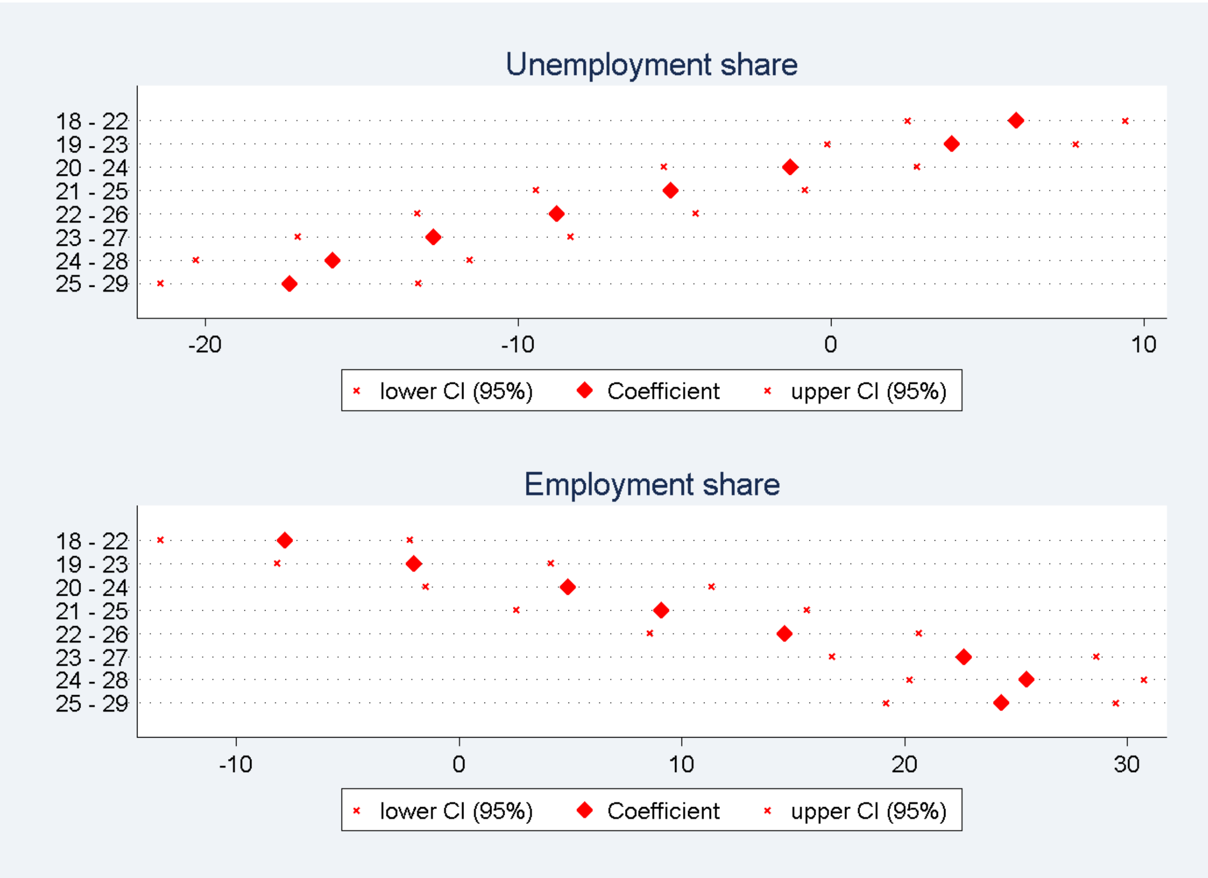
vector x_{jrt} only contains a constant and three sets of dummy variables for each of the three dimensions of the cohort-size variable: regions, years and age groups. In the second specification a set of control variables is added to the model (definitions and summary statistics are given in Table A1 in the Appendix). One part of these variables is assumed to affect the (un-)employment probability at the individual level and has therefore been aggregated in order to control for compositional differences between region-year-age cells. They include the share of individuals in such cells that a) belong to different educational groups according to the International Standard Classification of Education (ISCED), b) are married and c) reside in areas that differ with respect to their degree of urbanisation. Moreover, we add the level of regional GDP. While the use of year dummies accounts for shocks that are common to all region-age cells, this variable is useful in order to control for the region-specific economic environment in a given year. The inclusion of regional GDP therefore helps to avoid the estimated cohort-size effects being confounded by regional economic shocks.

4 Results

Figure 1 shows the estimated coefficients and confidence intervals on the cohort-size variable using overlapping samples of differently aged individuals when the dependent variable is the unemployment and employment share, respectively. For both outcome variables, the effect of cohort size varies substantially across age groups. When the dependent variable is the unemployment share, the effects are positive and statistically significant for individuals aged 18-22 but are negative and statistically significant for older groups. The effect appears to converge to between -10 and -20 for the older groups. The shift in sign and magnitude of the coefficients coincides with a decrease in the share of individuals reporting to be in education (see Figure A1 in the Appendix). In the employment model, cohort-size effects are significant and negative for individuals aged 18-22 but positive and significant for older age groups, converging to a value of approximately 25.

The results for the younger age groups appear to be supportive of the cohort-crowding hypothesis. However, our view is that the estimated effects for younger age groups cannot be regarded as a direct test of this hypothesis since they capture both the effect of cohort size on labour-market participation and the effect on (un-)employment. For example, the finding that cohort size reduces the employment share of individuals aged 18-22 may indicate either that large cohorts lead young individuals to acquire education and thereby defer entry to the labour market or that young individuals in the labour market are disadvantaged by belonging to a large age group. In addition to this problem of interpretation, the change in the coefficients may be driven by measurement error in the cohort-size variable. As discussed above, this variable is supposed to measure the availability of similarly aged individuals on the labour market, but in light of the large share of young individuals in education, some of whom will be committed to long-term programmes, it is less suitable as a measure of labour-market availability in younger than in older groups.

Figure 1: Cohort-size coefficients for different age groups



Source: EU-SILC (authors' calculations). Coefficients are obtained from weighted 2SLS estimation of a model containing dummy variables for regions, years and age groups. Robust standard errors are used.

In order to mitigate this problem, the remainder of this section focuses on individuals aged 25-29. As can be seen from Figure A1, the share of individuals in education is considerably smaller for those age groups. In this age range, the cohort-size variable should therefore present a better measure of the degree of labour-market crowding, while any confounding effects resulting from the preceding decision to enter the labour market or to acquire further education will be less relevant. Table 3 contains OLS and 2SLS estimation results for each of the two specifications discussed in Section 3.3 using a sample of individuals aged 25-29 (full results including the coefficients of the control variables can be found in Tables A2 and A3 in the Appendix and the results of the first-stage regressions are shown in Table A4).

Table 3: OLS and 2SLS regression results

Panel A: Unemployment share	OLS	2SLS	OLS	2SLS
Cohort size	-10.32*** (1.70)	-17.30*** (2.10)	-7.98*** (1.73)	-15.06*** (2.05)
Dummies				
Region	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes
Observations				
Region-year-age cells	1,959	1,959	1,959	1,959
R^2	0.38	0.37	0.41	0.40
F-stat	-	1,540.67***	-	1,642.59***
ME(std)	-0.03***	-0.05***	-0.02***	-0.05***
Panel B: Employment share	**OLS**	**2SLS**	**OLS**	**2SLS**
Cohort size	14.39*** (2.03)	24.32*** (2.64)	11.91*** (2.02)	22.07*** (2.52)
Dummies				
Region	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes
Observations				
Region-year-age cells	1,959	1,959	1,959	1,959
R^2	0.53	0.52	0.56	0.55
F-stat	-	1,540.67***	-	1,642.59***
ME(std)	0.04***	0.08***	0.04***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

The first two columns of panel A show that in the baseline model an increase in cohort size is predicted to decrease the share of individuals in the corresponding age group that are unemployed. OLS and 2SLS estimates have the same sign and are statistically significant at the 1% level. The finding that the latter are larger (in absolute terms) was also obtained by Shimer (2001) in some specifications and is consistent with the argument (see footnote 8) that cohort-size effects are heterogeneous across region-year-age cells and that immigrants are less attractive to employers than individuals that have lived in the region for 14 years. The third and fourth columns show that when the set of control variables, described in Section 3.3, are added to the model, the cohort-size coefficients decrease somewhat in magnitude. To give a better impression of the size of the coefficients, marginal effects for changes in cohort size of one standard deviation are shown at the bottom of panel A. Such an increase is predicted to reduce the share of unemployed in an age group by 5 percentage points, which is a sizeable effect given that the average unemployment share is 13% (see Table 1). Finally, the size of the F-statistics suggests that the excluded instrument has predictive power for the endogenous cohort-size variable with values considerably larger than the threshold value of 10 (Staiger and Stock, 1997). The results for the employment model are shown in panel B. The cohort-size variable is found to have a statistically significant and positive effect on the employment share. Adding control variables slightly reduces the size of the coefficients. For 2SLS estimation, an increase in cohort size by one standard deviation is predicted to increase the employment share by between 7 and 8 percentage points. In light of an average employment share of 77% this change is comparatively smaller than the corresponding effect on the unemployment share.

As discussed in Section 3.2, the above results use variation across regions, years and age groups. Table A5 shows cohort-size coefficients that are obtained when the identifying variation is restricted to a single source. This is accomplished by adding dummy variables for interactions between regions and age groups (identification is based on variation over time), between years and age groups (variation across regions only) or between regions and years (variation across age groups only). Except for an increase in the marginal effect of cohort-size on the unemployment share when only variation over time is used, the key results are not materially affected in the first two cases. By contrast, the cohort-size variable is not statistically significant in the unemployment model when region-year dummies are included. This is unsurprising since there is relatively little variation in cohort size across age within the sample. The results of various sensitivity analyses are available in the Supplementary Material.

The signs of the estimated coefficients suggest that members of large cohorts do not fare worse in terms of unemployment and employment outcomes. As such the results of this paper contradict the cohort-crowding hypothesis that increases in the size of an age group lead to increased unemployment within that group. Our findings rather provide evidence in support of Shimer (2001) that young individuals benefit from being part of large cohorts. However, even though increases in cohort size are found to increase the share of employed individuals in the corresponding age group, these results do not provide any evidence regarding the type and conditions of employment. Indeed results by Moffat and Roth (2016) that are also based on EU-SILC data show that individuals with completed secondary education command lower wages when they are part of a larger cohort. Similarly, using German microdata Garloff and Roth (2016) find that an increase in the share of youths in the population reduces young workers' wages; moreover, their analysis provides evidence that belonging to a larger youth cohort increases the likelihood of being employed in occupations and industries that pay lower wages.

5 Conclusion

A prominent research question of the cohort-size literature concerns the effect that the size of an age group has on its members' employment and unemployment outcomes. Based on the assumption of imperfect substitutability of differently aged workers, these outcomes should be determined by the size of an age group that is available to the labour market. As this quantity is typically not observable, the common approach has been to use the size of an age group as a proxy for age-specific labour supply instead. However, this ignores the fact that among the young the size of an age group will only be a poor measure of the size of the group that is available to the labour market because of the large share of individuals who participate in education.

This gives rise to two problems. First, for young age groups the estimated effect of cohort size on (un-)employment will be confounded by the former's effect on the decision to participate in the labour market in the first place. Second, using the size of an age group induces a problem of measurement error that the standard IV approach is unable to solve. For these reasons, the standard identification strategy is unsuited to produce informative insights into the effects of cohort crowding for young age groups regardless of whether an age-specific cohort-size variable is used that also varies across age or, as in other papers, a youth-share variable is employed.

To illustrate this, we estimate the effect of cohort size on age-specific employment and unemployment outcomes using data comprising information on 49 regions covering the period 2005-2012. In a first step we show that the estimated effects of cohort size are indeed highly sensitive to the chosen age range. In particular, we find that the sign of the coefficient changes as successively younger age groups are used. In a second step we apply these models to the age group 25-29 for which the above-mentioned problems should be less of a concern because participation rates in education are considerably lower. The results of this analysis suggest that an increase in cohort size reduces the unemployment share in an age group and increases the employment share, which is consistent with the mechanism between the youth share and (un-)employment outcomes that is described in Shimer (2001).

Acknowledgements

The authors would like to thank Georgios Marios Chrysanthou, Bernd Hayo and Michael Kirk as well as the participants of IAB's regional research network meeting, the 63rd Annual Meeting of the French Economic Association (AFSE), the 7th Summer Conference in Regional Science, the 4th World Meeting of the Society of Labor Economists (SOLE) and the European Association of Labour Economists (EALE) and the 2014 Congress of the European Regional Science Association (ERSA) for valuable comments. This paper uses data from the European Union Statistics on Income and Living Conditions (EU-SILC). The results and conclusions are those of the authors and not those of Eurostat, the European Commission or any of the national statistical authorities whose data have been used.

References

Angrist, J. and Imbens, G. (1995) Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity, *Journal of the American Statistical Association*, **90**, 431-42.

Berger, M. and Schaffner, S. (2015) A Note on How to Realize the Full Potential of the EU-SILC Data, ZEW Discussion Paper 15-005, Centre for European Economic Research, Mannheim.

Biagi, F. and Lucifora, C. (2008) Demographic and education effects on unemployment in Europe, *Labour Economics,* **15**, 1076-101.

Bound, J., Brown, C. and Mathiowetz, N. (2001) Measurement Error in Survey Data, in J. Heckman and E. Leamer (eds.) *Handbook of Econometrics*, North-Holland, Amsterdam.

Brunello, G. (2010) The effects of cohort size on European earnings, *Journal of Population Economics*, **23**, 273-90.

Card, D. and Lemieux, T. (2001) Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis, *Quarterly Journal of Economics*, **116**, 705-46.

Easterlin, R. (1961) The American baby boom in historical perspective, *American Economic Review*, **51**, 869-911.

Fertig, M., Schmidt, C. and Sinning, M. (2009) The Impact of Demographic Change on Human Capital Accumulation, *Labour Economics,* **16**, 659-68.

Foote, C. (2007) Space and time in macroeconomic panel data: young workers and state-level unemployment revisited, Research Department Working Paper No. 07-10, Federal Reserve Bank of Boston, Boston.

Garloff, A., Pohl, C. and Schanne, N. (2013) Do small labor market entry cohorts reduce unemployment?, *Demographic Research*, **29**, 379-406.

Garloff, A. and Roth, D. (2016) Regional population structure and young workers' wages, IAB Discussion Paper No. 06/2016, Institute for Employment Research, Nuremberg.

Hausman, J. (2001) Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left, *Journal of Economic Perspectives,* **15**, 57-67.

Imbens, G. and Angrist, J. (1994) Identification and Estimation of Local Average Treatment Effects, *Econometrica*, **62**, 467-75.

Jones, S. and Riddell, C. (2006) Unemployment and Nonemployment: Heterogeneities in Labor Market States, *Review of Economics and Statistics*, **88**, 314-23.

Kim, J. and Park, J. (2013) Foreign Direct Investment and Country-Specific Human Capital, *Economic Inquiry*, **51**, 198-210.

Korenman, S. and Neumark, D. (2000) Cohort Crowding and Youth Labor Markets: A Cross-National Analysis, in D.G. Blanchflower and R.B. Freeman (eds.) *Youth Employment and Joblessness in Advanced Countries*, University of Chicago Press, Chicago.

Michaelis, J. and Debus, M. (2011) Wage and (un-)employment effects of an ageing workforce, *Journal of Population Economics*, **24**, 1493-511.

Moffat, J. and Roth, D. (2016) The Cohort Size-Wage Relationship in Europe, *Labour*, forthcoming.

Moffat, J. and Yoo, H.I. (2015) Who are the unemployed? Evidence from the United Kingdom, *Economics Letters*, **132**, 61-4.

Schmidt, C. (1993) Ageing and unemployment, in P. Johnson and K. Zimmermann (eds.) *Labour markets in an ageing Europe*, Cambridge University Press, Cambridge.

Shimer, R. (2001) The Impact of Young Workers on the Aggregate Labor Market, *The Quarterly Journal of Economics*, **116**, 969-1007.

Skans, O.N. (2005) Age effects in Swedish local labor markets, *Economics Letters,* **86**, 419-26.

Staiger, D. and Stock, J.H. (1997) Instrumental Variables Regression with Weak Instruments, *Econometrica,* **65**, 557-86.

United Nations (2015) World Population Prospects: The 2015 Revision, Key Findings and Advance Tables, Working Paper No. ESA/P/WP.241, Department of Economics and Social Affairs, Population Division, New York.

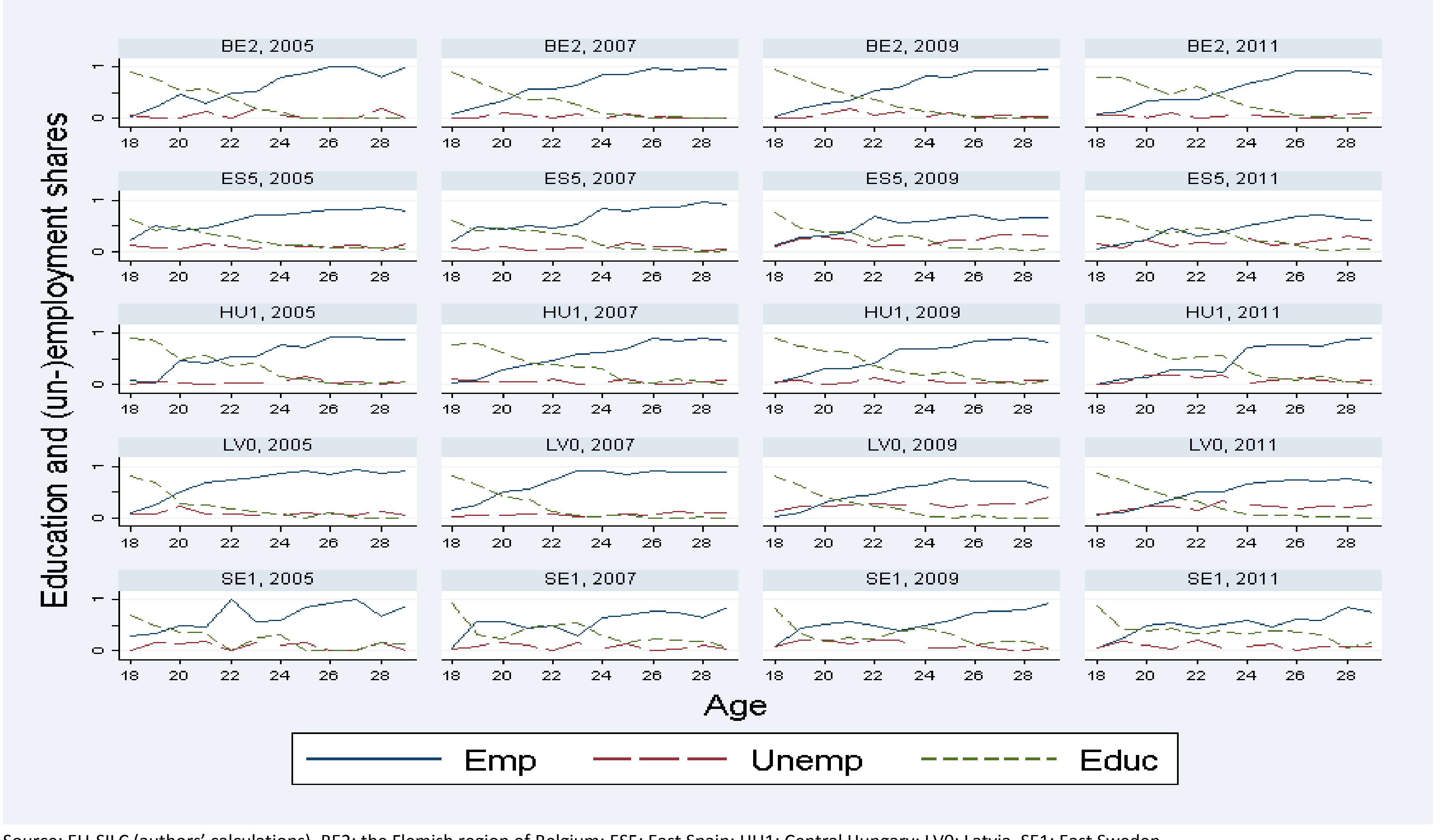
Welch, F. (1979) Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust, *Journal of Political Economy*, **87**, S65-S97.

Wright, R.E. (1991) Cohort size and earnings in Great Britain, *Journal of Population Economics*, **4**, 295-305.

Zimmermann, K. (1991) Ageing and the Labour Market. Age structure, cohort size and unemployment, *Journal of Population Economics,* **4**, 177-200.

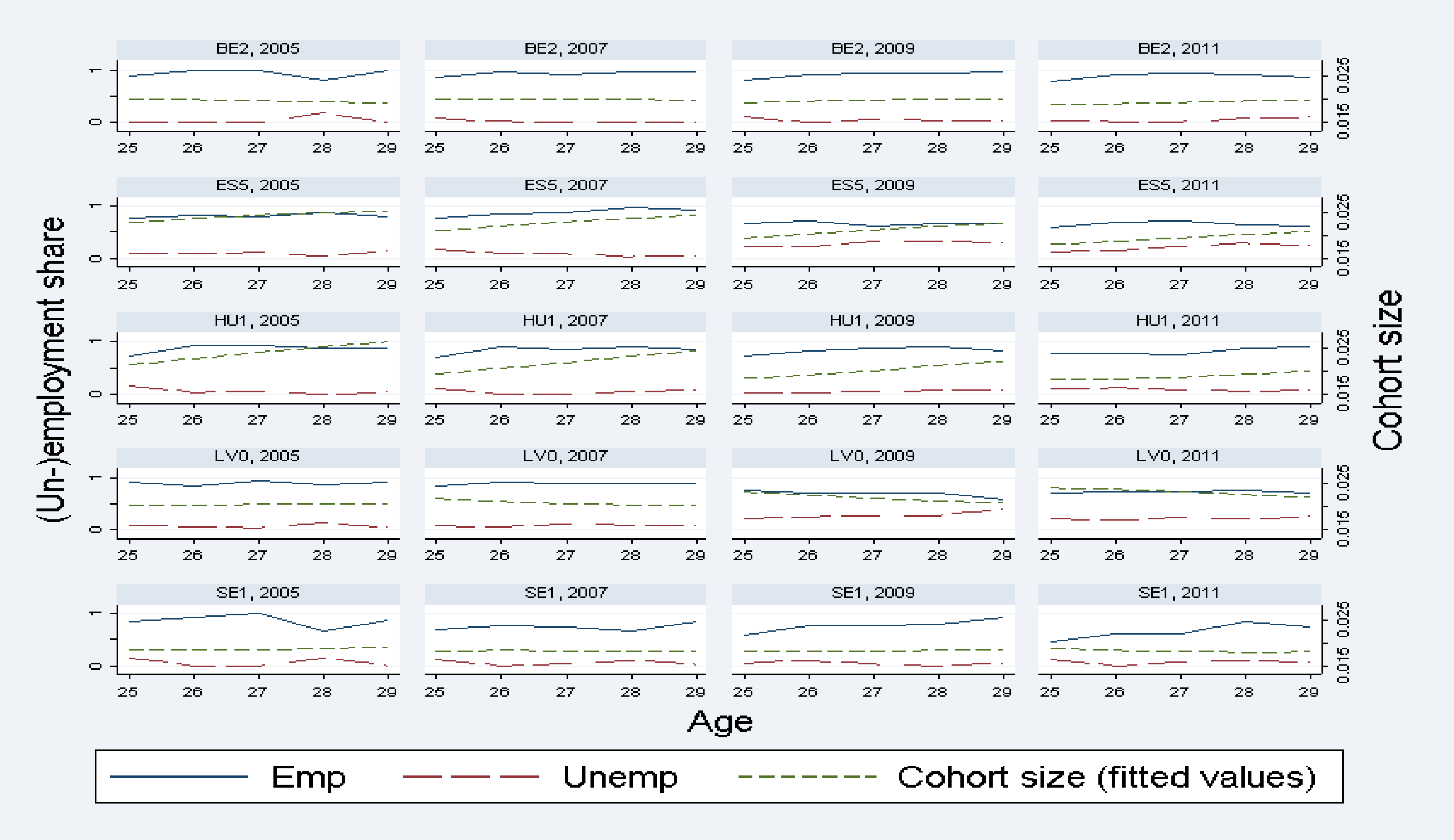
Appendix

Figure A1: Development of employment, unemployment and education shares across age groups



Source: EU-SILC (authors' calculations). BE2: the Flemish region of Belgium; ES5: East Spain; HU1: Central Hungary; LV0: Latvia, SE1: East Sweden.

Figure A2: Development of employment and unemployment shares and of fitted cohort-size variable over time



Source: EU-SILC (authors' calculations). BE2: the Flemish region of Belgium; ES5: East Spain; HU1: Central Hungary; LV0: Latvia, SE1: East Sweden.

Figure A3: Development of employment and unemployment shares and of fitted cohort-size variable over age groups

Source: EU-SILC (authors' calculations). BE2: the Flemish region of Belgium; ES5: East Spain; HU1: Central Hungary; LV0: Latvia, SE1: East Sweden.

Table A1: Definitions and descriptive statistics of control variables

Name	Definition	Source	Mean	Standard deviation
ISCED_0	Share of individuals in region-year-age cell with pre-primary education	EU-SILC	0.006	0.027
ISCED_1	Share of individuals in region-year-age cell with primary education	EU-SILC	0.040	0.060
ISCED_2	Share of individuals in region-year-age cell with lower secondary education	EU-SILC	0.136	0.133
ISCED_3	Share of individuals in region-year-age cell with upper secondary education	EU-SILC	0.479	0.187
ISCED_4	Share of individuals in region-year-age cell with post-secondary, non-tertiary education	EU-SILC	0.035	0.052
ISCED_5	Share of individuals in region-year-age cell with tertiary education (also includes category ISCED_6, i.e. individuals with second stage of tertiary education)	EU-SILC	0.304	0.168
Married	Share of individuals in region-year-age cell that are married	EU-SILC	0.195	0.153
Urban_1	Share of individuals in region-year-age cell living in densely populated areas (an area with a population density of more than 500 inhabitants per square kilometre (km) and a population of at least 50,000 inhabitants)	EU-SILC	0.461	0.216
Urban_2	Share of individuals in region-year-age cell living in intermediately populated areas (an area with a population density of more than 100 inhabitants per square km and either a population of at least 50,000 inhabitants or adjacent to a 'densely populated' area)	EU-SILC	0.248	0.170
Urban_3	Share of individuals in region-year-age cell living in thinly populated areas (an area with fewer than 100 inhabitants per square km and a population of less than 50,000 inhabitants)	EU-SILC	0.291	0.222
GDP	Gross domestic product at the NUTS1 level (in billion Euros, adjusted for purchasing-power-parity)	Eurostat	188.391	127.737

Means and standard deviations are weighted by the weight-adjusted number of individuals per region-year-age group cell.

Table A2: Full OLS and 2SLS regression results (Unemployment share)

Unemployment share	OLS	2SLS	OLS	2SLS
Cohort size	-10.32***	-17.30***	-7.98***	-15.06***
	(1.70)	(2.10)	(1.73)	(2.05)
Dummies				
Region	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Control variables				
ISCED_1	-	-	0.07	0.08
			(0.14)	(0.14)
ISCED_2	-	-	0.06	0.06
			(0.13)	(0.13)
ISCED_3	-	-	-0.01	-0.01
			(0.12)	(0.12)
ISCED_4	-	-	-0.11	-0.09
			(0.13)	(0.13)
ISCED_5	-	-	-0.08	-0.08
			(0.13)	(0.13)
Married	-	-	-0.10***	-0.09***
			(0.02)	(0.02)
Urban_2	-	-	-0.03	-0.03
			(0.03)	(0.03)
Urban_3	-	-	-0.03	-0.03
			(0.03)	(0.03)
GDP	-	-	-0.00***	-0.00***
			(0.00)	(0.00)
Observations				
Region-year-age cells	1,959	1,959	1,959	1,959
R^2	0.38	0.37	0.41	0.40
F-stat	-	1,540.67***	-	1,642.59***
ME(std)	-0.03***	-0.05***	-0.02***	-0.05***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table A3: Full OLS and 2SLS regression results (Employment share)

Employment share	OLS	2SLS	OLS	2SLS
Cohort size	14.39***	24.32***	11.91***	22.07***
	(2.03)	(2.64)	(2.02)	(2.52)
Dummies				
Region	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Control variables				
ISCED_1	-	-	0.47**	0.46**
			(0.20)	(0.20)
ISCED_2	-	-	0.51**	0.51**
			(0.20)	(0.20)
ISCED_3	-	-	0.57***	0.58***
			(0.19)	(0.19)
ISCED_4	-	-	0.66***	0.64***
			(0.20)	(0.20)
ISCED_5	-	-	0.62***	0.62***
			(0.19)	(0.19)
Married	-	-	0.09***	0.09**
			(0.03)	(0.03)
Urban_2	-	-	0.06*	0.06*
			(0.03)	(0.03)
Urban_3	-	-	0.07*	0.07*
			(0.04)	(0.04)
GDP	-	-	0.00***	0.00***
			(0.00)	(0.00)
Observations				
Region-year-age cells	1,959	1,959	1,959	1,959
R^2	0.53	0.52	0.56	0.55
F-stat	-	1,540.67***	-	1,642.59***
ME(std)	0.04***	0.08***	0.04***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table A4: First-stage regression results

	Unemployment share		Employment share	
Instrument	0.93*** (0.02)	0.93*** (0.02)	0.93*** (0.02)	0.93*** (0.02)
Dummies				
Region	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Control variables	No	Yes	No	Yes
Observations				
Region-year-age cells	1,959	1,959	1,959	1,959
R^2	0.92	0.92	0.92	0.92
F-stat	1,540.67***	1,642.59***	1,540.67***	1,642.59***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables.

Table A5: OLS and 2SLS results

Panel A: Unemployment share	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-12.84*** (1.91)	-23.01*** (2.37)	-10.76*** (1.67)	-17.35*** (2.07)	-2.16 (2.15)	-1.46 (2.64)
Dummies						
Region	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	Yes	Yes	No	No	No	No
Year-by-age	No	No	Yes	Yes	No	No
Region-by-year	No	No	No	No	Yes	Yes
Control variables	No	No	No	No	No	No
Observations						
Region-year-age cells	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.43	0.41	0.39	0.38	0.57	0.57
F-stat	-	1,140.11***	-	1,582.57***	-	674.72***
ME(std)	-0.04***	-0.07***	-0.03***	-0.05***	-0.01	-0.00
Panel B: Employment share	**OLS**	**2SLS**	**OLS**	**2SLS**	**OLS**	**2SLS**
Cohort size	13.88*** (2.24)	26.55*** (2.89)	14.41*** (2.02)	24.40*** (2.59)	7.24*** (2.73)	11.15*** (3.37)
Dummies						
Region	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	Yes	Yes	No	No	No	No
Year-by-age	No	No	Yes	Yes	No	No
Region-by-year	No	No	No	No	Yes	Yes
Control variables	No	No	No	No	No	No
Observations						
Region-year-age cells	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.58	0.57	0.54	0.53	0.66	0.66
F-stat	-	1,140.11***	-	1,582.57	-	674.72***
ME(std)	0.04***	0.08***	0.04***	0.08***	0.02***	0.03***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Supplementary material

S1 Selection of the age range and measurement error

The paper's main finding is that the estimated effect of cohort size on the (un-)employment share is sensitive to the selected age range of the sample (see Figure 1 in the paper). We propose two explanations for the observed pattern of the coefficients and in both cases the core of the argument is that for young age groups the cohort-size variable can be a poor measure of the age-specific supply of labour: first, a population-based cohort-size variable will include a substantial number of individuals that are not on the labour market, primarily because they are acquiring education; second, given the large share of non-participants among young age groups the estimated effect of cohort-size on the (un-)employment share will be confounded by the former's effect on the decision to participate in the labour market. In the following, we provide further detail on the former point.

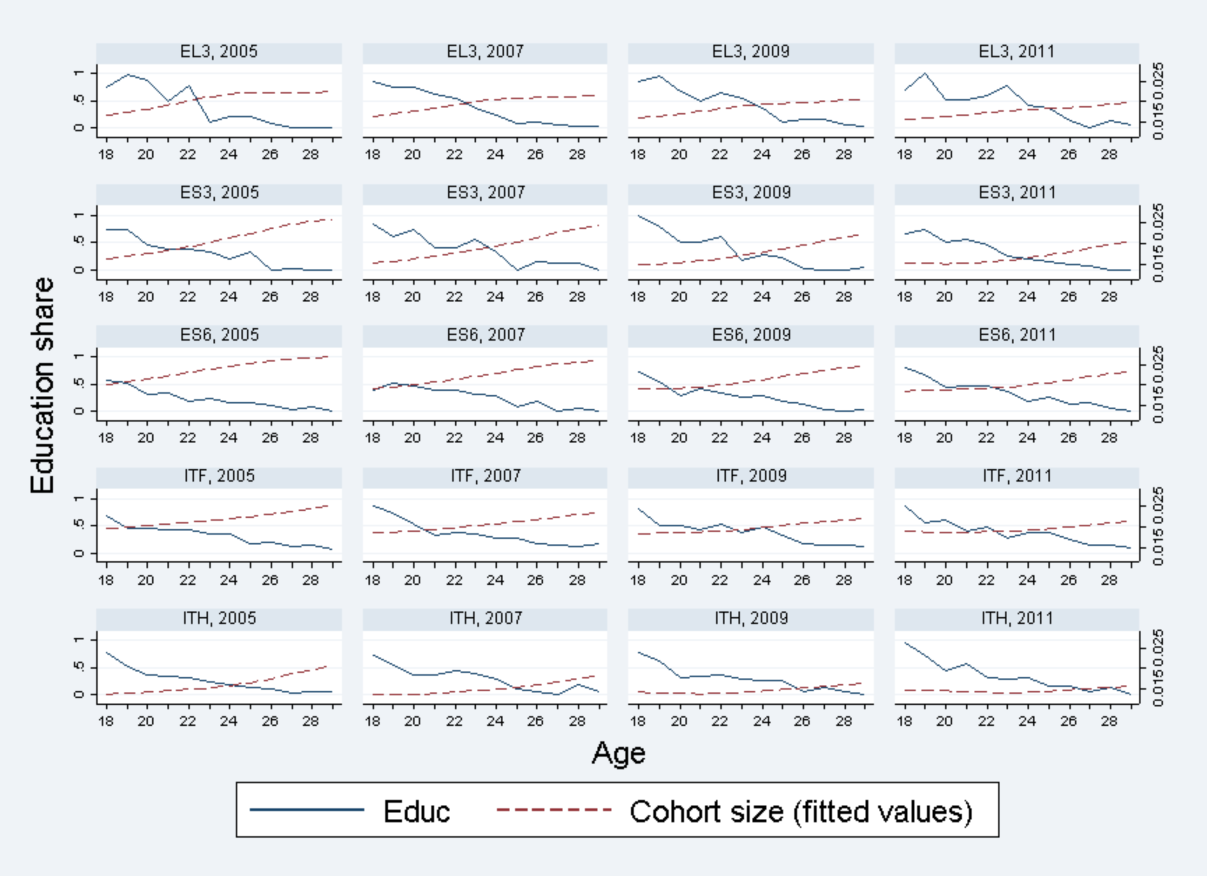
Figures S1 and S2 plot the share of individuals reporting to be in education against age for different region-year combinations.[1] As can be seen, the education share can be close to 100% at age 18 and usually is in excess of 50% at age 20, whereas the share is considerably smaller in the age range 25-29, which is used in the empirical analysis of this paper.[2] This observation provides support for the hypothesis that the share of individuals that are included in a population-based cohort-size variable but that are not on the labour market can be substantial, especially among young age groups. However, it is important to note that simply excluding those individuals that report to be in education from the construction of the cohort-size variable does not necessarily lead to a better measure of age-specific labour supply. First, a part of the group of individuals reporting to be in education may be enticed to enter the labour market depending on the conditions of employment and as such should be treated as being available to the labour market, whereas participants in lengthy degree programmes are less likely to do so (these groups cannot be separated in the data); second, switching between

[1] The regions are EL3 (Attica), ES3 (Madrid), ES6 (Andalusia), ITF (Southern Italy) and ITH (Northeast Italy), CZ0 (Czech Republic), DKO (Denmark), FR1 (Île de France), LT0 (Lithuania) and PL1 (Central Poland).

[2] The main exception is Denmark where the education share takes longer to decrease and can be large at later ages (e.g. age 26 in the year 2011). However, we are able to show in Figures S3 and S4 that the exclusion of Denmark from the sample has virtually no effect on the size of the coefficient in the unemployment and the employment model, respectively, while allowing the sample to start at age 26 instead of 25 also yields comparable coefficients in both models (see Figure S6).

periods of participation and non-participation is more likely to occur among young individuals compared to older age groups whose members tend to be more established in the labour market.

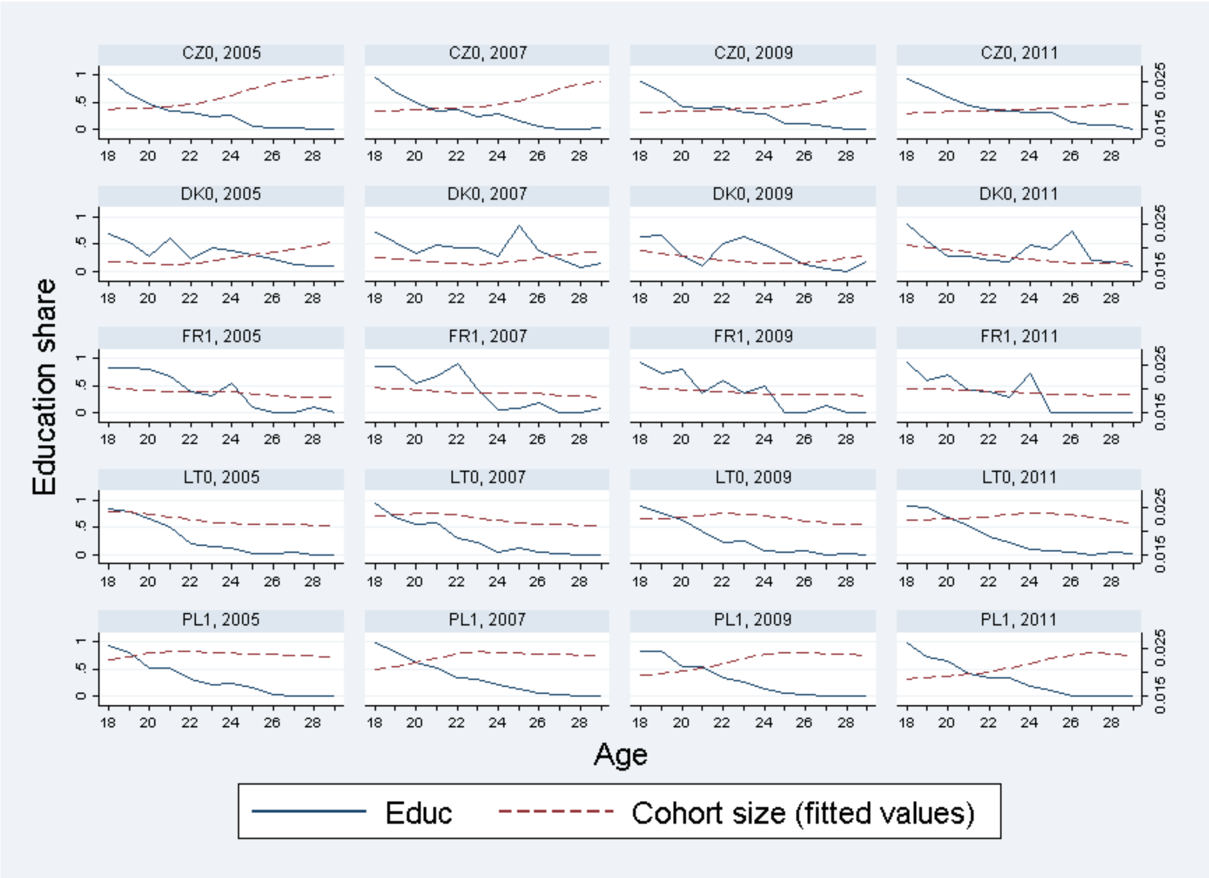
Figure S1: Development of education share and fitted cohort-size variable (set 1)



Source: EU-SILC (authors' calculations).

The case of non-classical measurement error arises when the degree of measurement error is correlated with the measurement error-free cohort-size variable which in this case is given by the size of an age group that is also available to the labour market. One reason why such a correlation might arise is that the degree of measurement error is larger in younger age groups (as shown in Figures S1 and S2, the share of individuals in education is considerably higher among younger age groups), while at the same time younger age groups tend to be smaller than older ones (for given regions and years). This implies a (negative) correlation between observed cohort size and the degree of measurement error.

Figure S2: Development of education share and fitted cohort-size variable (set 2)



Source: EU-SILC (authors' calculations).

This hypothesis is supported by Figures S1 and S2 which also show the development of the fitted value of the cohort-size variable (obtained from a regression on the instrument). From Figure S1 it can be seen that in the Southern European regions of Spain (ES3, ES6), Italy (ITF, ITH) and Greece (EL3) younger cohorts are indeed smaller than older ones, especially in earlier years. Figure S2 illustrates that similar patterns can be found in the Czech Republic (CZ0) and Central Poland (PL1). In contrast, younger cohorts are larger than older ones in Lithuania (LT0). The profiles of most Western European regions tend to be flat, as exemplified by the Île de France (FR1); an exception is given by Denmark (DK0) where older age groups also tend to belong to larger cohorts than younger ones.

It is argued in the paper that under certain conditions there will be a negative correlation between the latent cohort-size variable, which measures age-specific labour supply, and the degree of measurement error. According to Equation 4 in the paper the observed cohort-size variable can be expressed as the sum of age-specific labour supply and measurement error:

$$CS_{jrt} = CS^{*}_{jrt} + u_{jrt} \qquad \text{[S1]}$$

This condition can be re-written in form of the size of the age-group *j* in region *r* at time *t*, N_{jrt}, the number of individuals in that age group that are available to the labour market, N^*_{jrt}, those that are not available, N^{out}_{jrt}, and the overall population, N_{rt}:

$$\frac{N_{jrt}}{N_{rt}} = \frac{N^*_{jrt}}{N_{rt}} + \frac{N^{out}_{jrt}}{N_{rt}} \qquad [S2]$$

The degree of measurement error can be expressed in terms of the share of non-participants, N^{out}_{jrt}, in an age group, N_{jrt}:

$$\alpha_{jrt} = \frac{N^{out}_{jrt}}{N_{jrt}} \qquad [S3]$$

Since for a given region and year, the denominators are identical for different age groups, it is sufficient to focus on the numerators. There will be a negative correlation between the latent cohort-size variable and the degree of measurement error across age groups, if the number of participants, N^*_{jrt}, increases in older age groups while the number of non-participants, N^{out}_{jrt}, becomes smaller. This can be formalised in terms of two age groups *k* and *l* ($k<l$):

$$N^*_{krt} < N^*_{lrt} \qquad [S4]$$

$$N^{out}_{krt} > N^{out}_{lrt} \qquad [S5]$$

Substituting Equation S3 into S5 and re-formulating yields the condition that the ratio of the non-participation shares in younger and older age groups exceeds the ratio of the size of the older and the younger age group (since older age groups are typically larger than younger ones, the condition in Equation S4 will hold if condition S5 is satisfied):

$$\frac{\alpha_{krt}}{\alpha_{lrt}} > \frac{N_{lrt}}{N_{krt}} \qquad [S6]$$

If the size of the education share is used as a proxy for the degree of measurement error, Figures S1 and S2 suggest that the above condition is not unreasonable since the difference in cohort size between age groups often appears less pronounced than the difference between education shares.

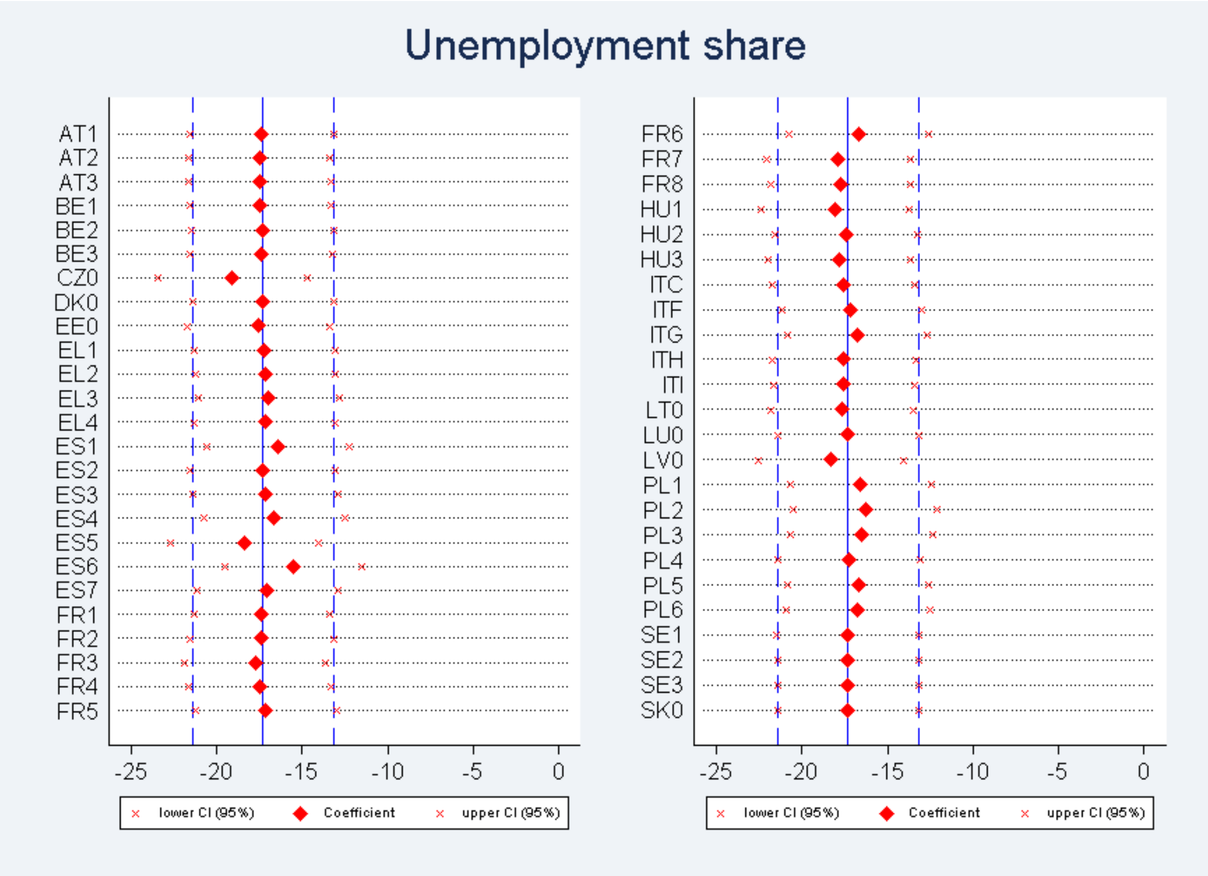
S2 Robustness of the empirical results

This section addresses the robustness of the estimated cohort-size coefficients to a variety of changes in the empirical model and in the underlying sample.

S2.1Robustness to the exclusion of individual regions, year and age groups

This part starts by assessing the sensitivity of the results to dropping individual regions, years and age groups. The cohort-size coefficients and their 95% confidence interval that are estimated from the reduced sample using an analogue of the specification that includes region, year and age dummies are shown in Figures S3 to S6. For better comparability these figures also contain the cohort-size coefficient and confidence interval from the full sample.

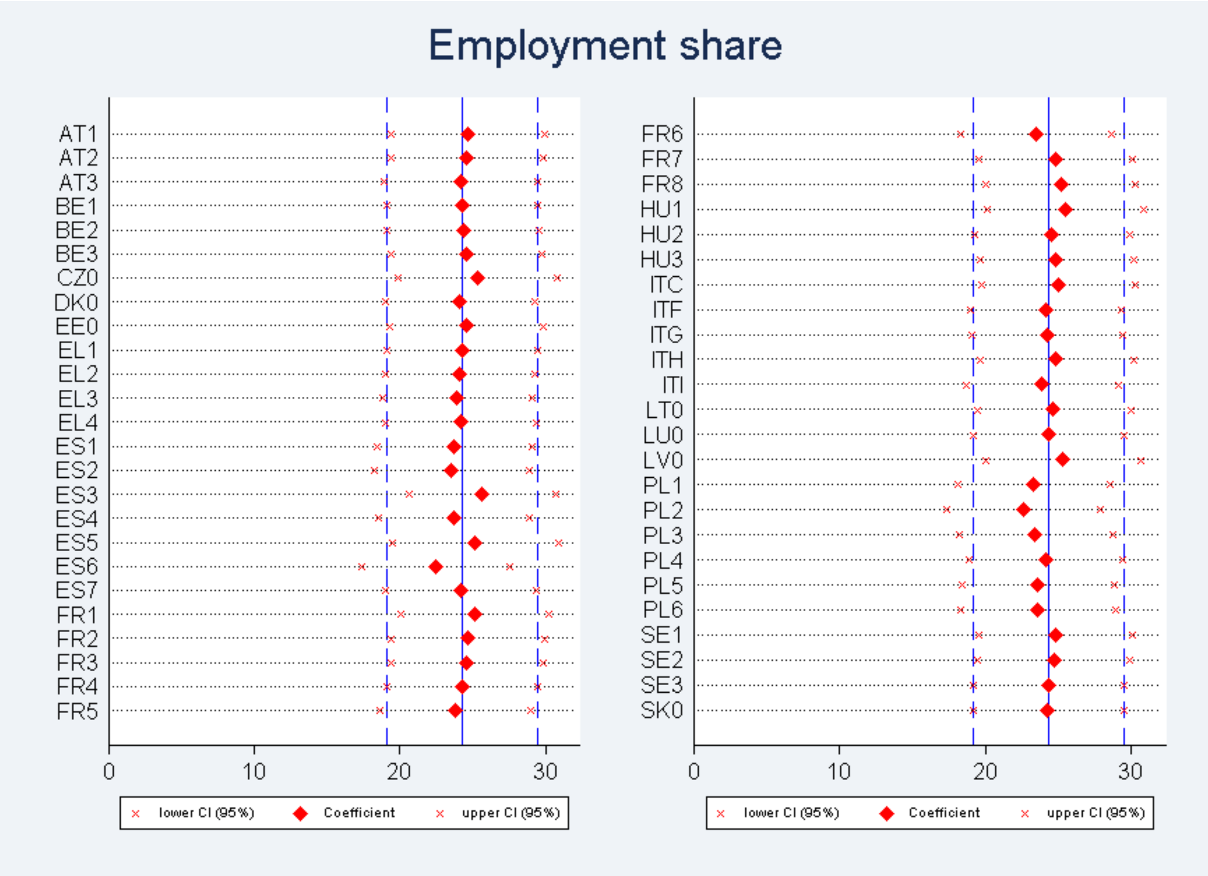
Figure S3: Exclusion of single regions (Unemployment share)



Source: EU-SILC (authors' calculations). Cohort-size coefficients are estimated as described in Section 3; the estimated model also includes region, year and age dummies; the blue solid line represents the cohort-size coefficients from the full model, the blue dashed lines the corresponding 95% confidence interval.

As illustrated by Figure S3, for most regions it is the case that their exclusion does not have a large effect on the cohort-size coefficient as can be seen by the former's closeness to the solid blue line. Some regions, however, do affect the size of the coefficient if they are excluded: in the unemployment model dropping the Czech Republic (CZ0) or Latvia (LV0) increases the magnitude of the coefficient, while exclusion of the Spanish region Andalusia (ES6) or the Polish regions PL1-PL3 leads to a decrease. The resulting estimates do, however, remain well within the 95% confidence interval of the full sample's cohort-size coefficient (given by the dashed blue lines). Those regions that, when excluded, decrease or increase the magnitude of the cohort-size coefficient tend to have the same effect in the employment model, while there are also some additional regions that now have a larger effect on the size of the coefficient (ES3, ES5, FR8), as shown in Figure S4. As with the unemployment share, the estimates always lie within the confidence interval of the full sample's coefficient. An increase in the magnitude of the cohort-size coefficient implies that in the specific sub-sample labour-market shares are more responsive to changes in cohort size: the decreasing effect on the unemployment share as well as the increasing effect on the employment share both become larger –and vice-versa for a decrease in the magnitude of the coefficient. However, when interpreting the change in the coefficients it should be borne in mind that omission of a certain region (or year or age group) will also have an effect on the distribution of the cohort-size variable in the sample. The effect of an increase (decrease) in the coefficient's magnitude can be mitigated if the change in the underlying sample reduces (increases) the standard deviation of the cohort-size variable.

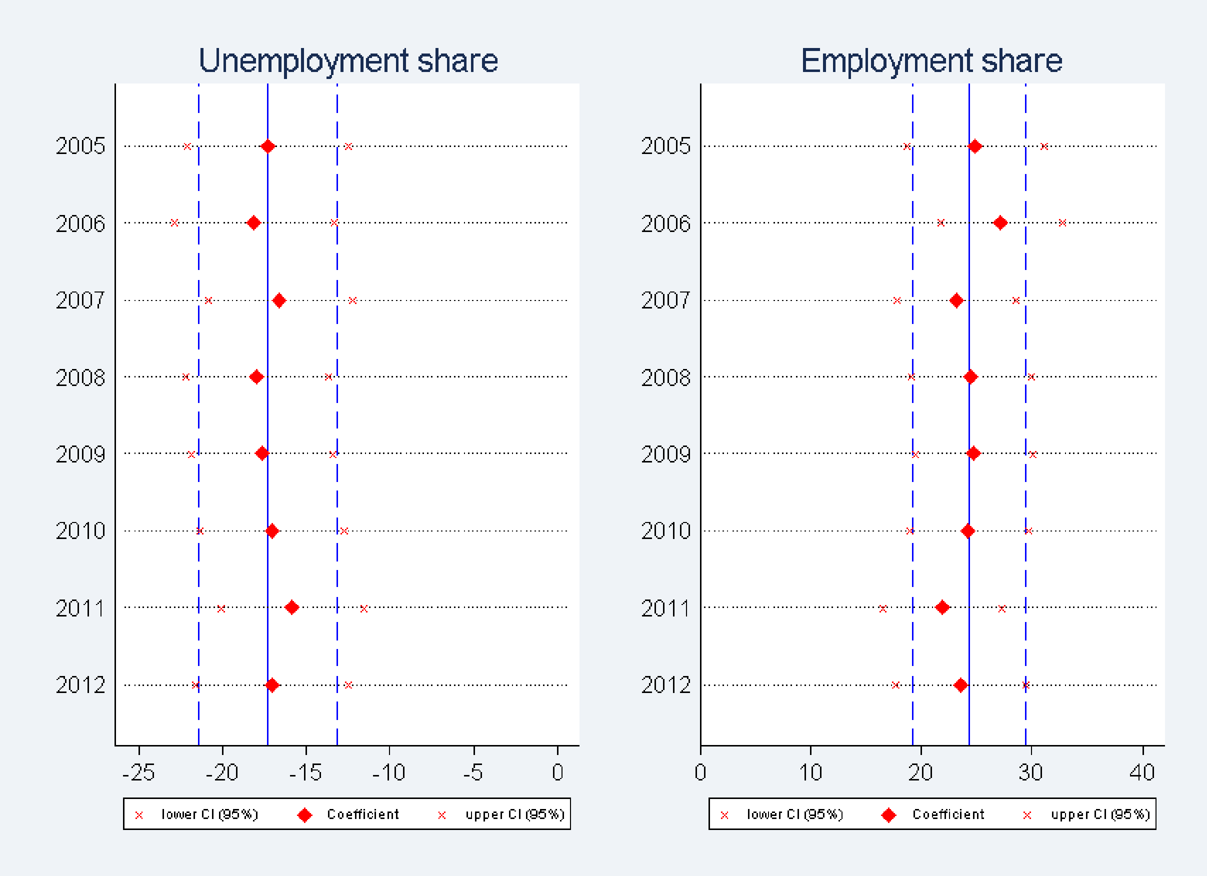
Figure S4: Exclusion of single regions (Employment share)



Source: EU-SILC (authors' calculations). Cohort-size coefficients are estimated as described in Section 3; the estimated model also includes region, year and age dummies; the blue solid line represents the cohort-size coefficients from the full model, the blue dashed lines the corresponding 95% confidence interval.

Figure S5 provides an overview of the effect that the exclusion of individual years has on the estimated cohort-size coefficients. While there are changes in the estimates in some cases, the former always remain within the confidence interval of the full sample's coefficients. Comparing the unemployment and the employment model, the coefficients appear to change in a symmetric manner, e.g. omission of the year 2006 increases the magnitude of the coefficients in both model, while dropping observations from the year 2011 leads to a decrease in size.

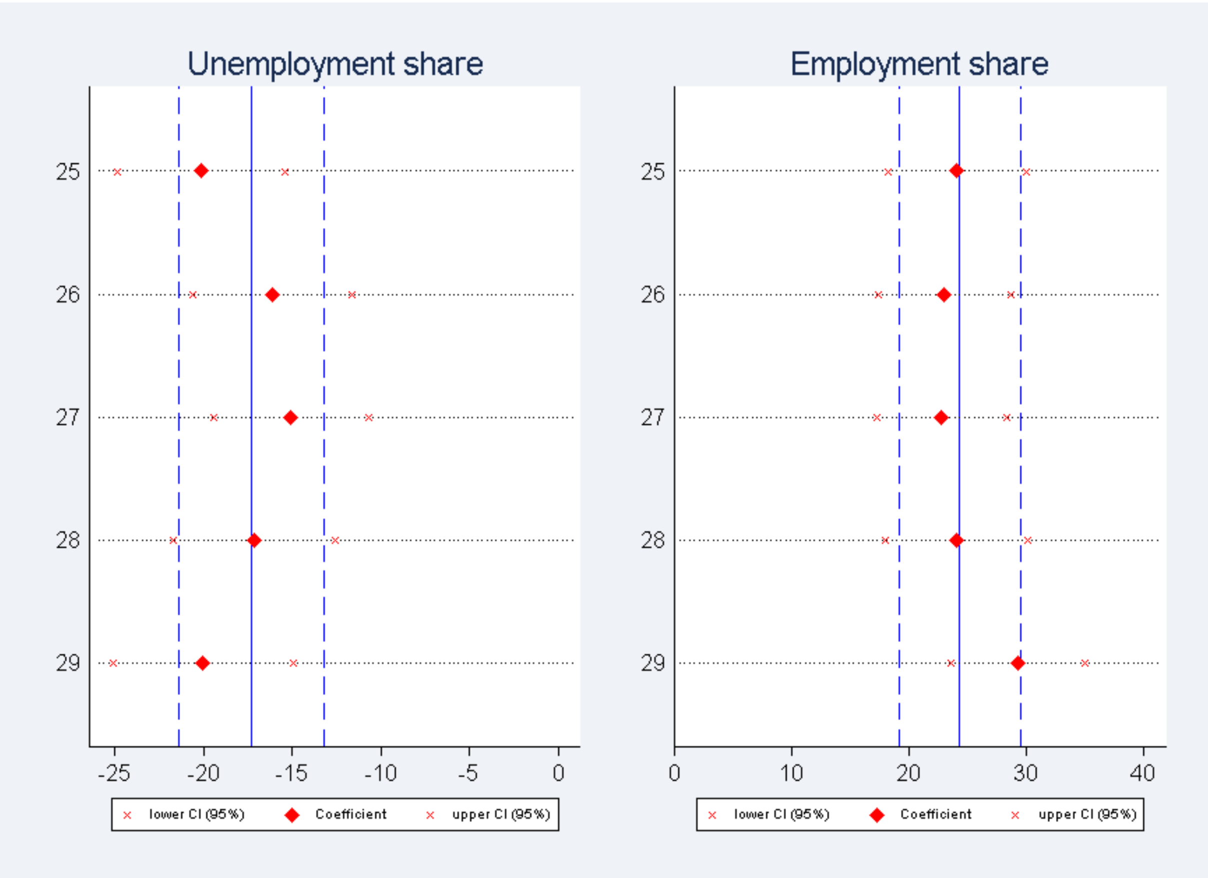
Figure S5: Exclusion of single years



Source: EU-SILC (authors' calculations). Cohort-size coefficients are estimated as described in Section 3; the estimated model also includes region, year and age dummies; the blue solid line represents the cohort-size coefficients from the full model, the blue dashed lines the corresponding 95% confidence interval.

The effects of excluding individual age groups from the sample are illustrated in Figure S6. The largest change in the coefficient can be observed when the age group 29 is dropped, in which case the magnitude of the coefficient increases in both models and almost moves outside of the full sample's confidence interval in the employment model. The responsiveness of labour-market shares to changes in cohort size therefore appears less pronounced for this age group. Unfortunately, the unavailability of lagged population data prevents the inclusion of older age groups in the sample and thus the possibility to check whether a further decrease in the strength of the relationship between cohort size and labour-market shares could be found at older ages. Such a development would be in line with the underlying mechanism that is proposed by Shimer (2001): firms create vacancies in areas where the share of young individuals is large because the former are usually not well matched to their jobs and a large pool of such individuals makes it easier for firms to find good matches for these vacancies. However, if the degree to which individuals are matched to their job increases with age, larger older age groups would not necessarily induce the same reaction on the firms' side because members of those age groups would not be as easily enticed to engage in on-the-job search as younger individuals, thereby reducing the incentive to firms to create vacancies. In addition, dropping age 25 also increases the magnitude of the coefficient in the unemployment model but has no sizeable effect in the employment model.

Figure S6: Exclusion of single age groups



Source: EU-SILC (authors' calculations). Cohort-size coefficients are estimated as described in Section 3; the estimated model also includes region, year and age dummies; the blue solid line represents the cohort-size coefficients from the full model, the blue dashed lines the corresponding 95% confidence interval.

To further assess to what extent the estimated cohort-size effects vary between different groups of regions, we estimate Equation 3 separately for regions from three parts of Europe: Southern Europe (16 regions from Greece, Italy and Spain), Eastern Europe (14 regions from the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland and Slovakia) and a combination of Northern and Western Europe (19 regions from Austria, Belgium, Denmark, France, Luxembourg and Sweden).Tables S1-S3 show the results of the baseline model as well as the coefficients from the model containing region-by-age dummies (with and without control variables).

Estimating separate models for each of the three regions reduces the degrees of freedom compared to the pooled sample, which is reflected in higher standard errors. Moreover, the explanatory power of the instrument appears to be lower as evidenced by a reduction in the first-stage F-statistics. Nevertheless, in many specifications the 2SLS coefficients remain negative and significant in the unemployment model and positive and significant in the employment model when the Southern European regions are used. All of the coefficients have the expected sign and are significant at the 1% level for the sample of Eastern European regions. While there are no significant effects for the remaining regions of Northern and Western Europe, this need not imply that the relationship between cohort size and labour-market outcomes is structurally different in this part of Europe, but may rather be a reflection of the limited variation in the cohort-size variable as could already be seen in Figures A2 and A3.

Table S1: OLS and 2SLS regression results (Southern European regions)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-5.54 (3.58)	-11.16** (5.13)	-3.51 (3.68)	-6.63 (5.36)	-8.93** (3.86)	-16.87*** (5.44)	-6.56* (3.95)	-12.23** (5.59)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	640	640	640	640	640	640	640	640
R^2	0.53	0.52	0.55	0.55	0.57	0.57	0.59	0.59
F-stat	-	338.11***	-	323.28***	-	340.42***	-	300.82***
ME(std)	-0.02	-0.04**	-0.01	-0.02	-0.03**	-0.06***	-0.02*	-0.04**

Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	7.55* (4.05)	8.97 (6.58)	6.07 (4.11)	5.92 (7.27)	11.76** (4.61)	14.73** (6.78)	10.05** (4.69)	11.23 (7.24)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	640	640	640	640	640	640	640	640
R^2	0.65	0.65	0.66	0.66	0.69	0.69	0.70	0.70
F-stat	-	338.11***	-	323.28***	-	340.42***	-	300.82***
ME(std)	0.03*	0.03	0.02	0.02	0.04**	0.05**	0.03**	0.04

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S2: OLS and 2SLS regression results (Eastern European regions)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-9.21*** (1.87)	-8.94*** (2.06)	-5.78*** (2.00)	-5.35*** (2.20)	-9.90*** (2.18)	-10.56*** (2.30)	-5.96*** (2.29)	-6.33*** (2.47)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	560	560	560	560	560	560	560	560
R^2	0.32	0.32	0.40	0.40	0.36	0.36	0.44	0.44
F-stat	-	843.57***	-	899.35***	-	705.96***	-	732.99***
ME(std)	-0.02***	-0.02***	-0.01***	-0.01***	-0.02***	-0.02***	-0.01***	-0.01***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	15.91*** (2.33)	20.28*** (2.63)	11.99*** (2.49)	16.06*** (2.80)	14.74*** (2.64)	19.99*** (2.94)	9.87*** (2.84)	14.71*** (3.33)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	560	560	560	560	560	560	560	560
R^2	0.41	0.41	0.48	0.48	0.46	0.45	0.53	0.52
F-stat	-	843.57***	-	899.35***	-	705.96***	-	732.99***
ME(std)	0.03***	0.04***	0.02***	0.03***	0.03***	0.04***	0.02***	0.03***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S3: OLS and 2SLS regression results (Northern and Western European regions)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-0.23 (3.93)	5.73 (7.88)	1.74 (4.07)	5.08 (7.51)	0.44 (4.22)	1.18 (7.51)	2.38 (4.46)	1.18 (6.99)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	759	759	759	759	759	759	759	759
R^2	0.13	0.12	0.17	0.17	0.19	0.19	0.22	0.22
F-stat	-	83.20***	-	94.78***	-	67.78***	-	77.30***
ME(std)	-0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-1.55 (5.45)	-0.23 (11.00)	-2.74 (5.31)	4.75 (10.36)	-6.66 (5.39)	-3.28 (9.58)	-7.75 (5.36)	1.20 (9.07)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	759	759	759	759	759	759	759	759
R^2	0.24	0.24	0.30	0.29	0.32	0.32	0.37	0.37
F-stat	-	83.20***	-	94.78***	-	67.78***	-	77.30***
ME(std)	-0.00	-0.00	-0.01	0.01	-0.01	-0.01	-0.02	0.00

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

S2.2 Robustness to changes in the model specification and the sample

This part assesses the robustness of the estimated relationship between cohort-size and the unemployment and the employment shares to a variety of changes in the specification of the empirical model or the underlying sample.

In Table S4 we first show that the paper's results also hold when instead of aggregating the dependent variable to the level of the region-year-age group the underlying microdata is used (Angrist and Pischke, 2009). In this case the dependent variable is defined as a binary variable that indicates whether an individual i in age group j, region r and year t is unemployed ($unemp_{ijrt}$) or employed (emp_{ijrt}). In light of the strong assumptions that have to be made to ensure consistency in a binary dependent variable model with endogenous regressors (Cameron and Trivedi, 2009) and since the focus of the analysis is on estimating marginal effects rather than on making predictions, a linear probability model is used to which we apply the same IV estimation strategy that is outlined in Section 3. As the cohort-size variable is defined at a higher level of aggregation than the dependent variable, which now may also vary across individuals in the same region-year-age group, standard errors are clustered at the level of the region-age group cell (Moulton, 1990). Observations are weighted by the individual-level weights which have been provided as part of the EU-SILC data and which have then been calibrated so that the estimated size of a region-year-age-sex cell matches the population size as reported by Eurostat (see Section 2). The size of the standard errors increases compared to the aggregate-level analysis but all coefficients remain statistically significant at the 1% level.

Table S4: OLS and 2SLS regression results (individual-level analysis)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-10.32*** (1.97)	-17.30*** (2.55)	-8.38*** (1.95)	-15.50*** (2.44)	-12.84*** (2.35)	-23.01*** (3.13)	-10.17*** (2.34)	-20.47*** (3.00)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations								
Individual-level	64,387	64,387	64,387	64,387	64,387	64,387	64,387	64,387
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
Region-age	243	243	243	243	243	243	243	243
R^2	0.04	0.04	0.06	0.06	0.05	0.04	0.07	0.07
F-stat	-	1,352.68***	-	1,568.95***	-	1,024.29***	-	1,385.52***
ME (std)	-0.03***	-0.05***	-0.03***	-0.05***	-0.04***	-0.07***	-0.03***	-0.06***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	14.39*** (2.29)	24.32*** (3.05)	12.20*** (2.32)	22.06*** (3.02)	13.88*** (2.61)	26.55*** (3.71)	10.73*** (2.59)	23.15*** (3.63)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations								
Individual-level	64,387	64,387	64,387	64,387	64,387	64,387	64,387	64,387
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
Region-age	243	243	243	243	243	243	243	243
R^2	0.07	0.07	0.10	0.10	0.08	0.08	0.10	0.10
F-stat	-	1,352.68***	-	1,568.95***	-	1,024.29	-	1,385.52
ME(std)	0.04***	0.08***	0.04***	0.07***	0.04***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Standard errors that are clustered at the level of the region-age group cell are shown in parentheses. The regression is weighted using calibrated individual-level weights. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

In this paper a specific form of the cohort-size variable is used which, first, includes age groups that are up to two years younger and older and, second, assigns lower weights to age groups that are further away from the reference group. This specification is chosen to incorporate the assumption that members of an age group also compete with individuals that are slightly younger and older, but that substitutability decreases with the age difference. However, Wright (1991) already notes that this specific formulation is arbitrary. We therefore show that the results are robust to using a weighted cohort-size variable that only includes age groups that are up to one year younger or older (Equation S7), the relative size of the own-age group which does not consider any other age groups (Equation S8) as well as a three-year sum (Equation S9) and a five-year sum (Equation S10) in which each group receives an equal weight. Tables S5 to S8 show that the cohort-size coefficients retain their sign and significance. Since the distribution of these variables differ, it is useful to look at the marginal effects of a change in the corresponding cohort-size variable by one standard deviation instead of the cohort-size coefficients in order to compare the magnitude of the effects across the different specifications.

$$CS_{jrt} = \frac{(1/4)N_{j-1,rt} + (1/2)N_{jrt} + (1/4)N_{j+1,rt}}{N_{16-64,rt}} \quad [S7]$$

$$CS_{jrt} = \frac{N_{jrt}}{N_{16-64,rt}} \quad [S8]$$

$$CS_{jrt} = \frac{N_{j-1,rt} + N_{jrt} + N_{j+1,rt}}{N_{16-64,rt}} \quad [S9]$$

$$CS_{jrt} = \frac{N_{j-1,rt} + N_{jrt} + N_{j+1,rt}}{N_{16-64,rt}} \quad [S10]$$

Table S5: OLS and 2SLS regression results (3-year weighted cohort-size variable)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-8.58*** (1.54)	-16.30*** (2.00)	-6.72*** (1.57)	-14.20*** (1.94)	-10.55*** (1.71)	-21.97*** (2.29)	-8.04*** (1.76)	-18.98*** (2.16)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.38	0.36	0.41	0.40	0.42	0.40	0.45	0.43
F-stat	-	1,216.01***	-	1,242.94***	-	885.58***	-	972.80***
ME(std)	-0.03***	-0.05***	-0.02***	-0.05***	-0.03***	-0.07***	-0.03***	-0.06***

Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	11.27*** (1.91)	23.02*** (2.51)	9.47*** (1.87)	21.04*** (2.39)	10.66*** (2.06)	25.23*** (2.78)	8.30*** (2.06)	22.40*** (2.59)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.53	0.51	0.56	0.54	0.57	0.56	0.60	0.59
F-stat	-	1,216.01***	-	1,242.94***	-	885.58***	-	972.80***
ME(std)	0.04***	0.07***	0.03***	0.07***	0.03***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S6: OLS and 2SLS regression results (own-age cohort-size variable)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-3.07*** (0.95)	-14.72*** (1.93)	-1.99** (0.93)	-12.96*** (1.88)	-3.41*** (1.02)	-20.02*** (2.32)	-2.11*** (1.02)	-17.62*** (2.18)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.37	0.29	0.40	0.33	0.41	0.26	0.45	0.32
F-stat	-	326.62***	-	337.13***	-	226.31***	-	249.71***
ME(std)	-0.01***	-0.06***	-0.01**	-0.05***	-0.01***	-0.08***	-0.01***	-0.07***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	3.01** (1.23)	20.84*** (2.55)	1.96* (1.19)	19.29*** (2.41)	2.45* (1.26)	22.92*** (2.84)	1.21 (1.23)	20.73*** (2.64)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.52	0.42	0.55	0.46	0.57	0.46	0.60	0.50
F-stat	-	326.62***	-	337.13***	-	226.31***	-	249.71***
ME(std)	0.01**	0.08***	0.01*	0.08***	0.01*	0.09***	0.00	0.08***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S7: OLS and 2SLS regression results (3-year non-weighted cohort-size variable)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-2.99***	-5.56***	-2.46***	-4.83***	-3.69***	-7.45***	-2.96***	-6.39***
	(0.54)	(0.68)	(0.55)	(0.66)	(0.61)	(0.78)	(0.62)	(0.73)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.38	0.37	0.41	0.40	0.42	0.40	0.46	0.44
F-stat	-	1,356.30***	-	1,410.33***	-	1,029.92***	-	1,118.76***
ME(std)	-0.03***	-0.05***	-0.02***	-0.05***	-0.04***	-0.07***	-0.03***	-0.06***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	4.30***	7.85***	3.77***	7.14***	4.12***	8.57***	3.43***	7.55***
	(0.67)	(0.84)	(0.66)	(0.81)	(0.74)	(0.94)	(0.72)	(0.87)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.53	0.52	0.56	0.55	0.58	0.56	0.61	0.60
F-stat	-	1,356.30***	-	1,410.33***	-	1,029.92***	-	1,118.76***
ME(std)	0.04***	0.08***	0.04***	0.07***	0.04***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S8: OLS and 2SLS regression results (5-year non-weighted cohort-size variable)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-2.03***	-3.60***	-1.54***	-3.13***	-2.50***	-4.72***	-1.85***	-4.08***
	(0.34)	(0.44)	(0.35)	(0.43)	(0.39)	(0.49)	(0.40)	(0.47)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.38	0.37	0.41	0.40	0.42	0.41	0.45	0.44
F-stat	-	1,483.63***	-	1,605.32***	-	1,096.43***	-	1,202.69***
ME(std)	-0.03***	-0.06***	-0.02***	-0.05***	-0.04***	-0.07***	-0.03***	-0.06***

Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	2.99*** (0.41)	5.04*** (0.55)	2.43*** (0.41)	4.55*** (0.53)	2.86*** (0.46)	5.47*** (0.60)	2.14*** (0.46)	4.80*** (0.56)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.53	0.52	0.56	0.55	0.58	0.57	0.61	0.60
F-stat	-	1,483.63***	-	1,605.32***	-	1,096.43***	-	1,202.69***
ME(std)	0.05***	0.08***	0.04***	0.07***	0.04***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

In the paper, individuals reporting to be in education are not excluded from the construction of the cohort-size variable. This is done because a part of these individuals may be willing to join the labour market if an attractive opportunity became available, while others are unlikely to do so because they are enrolled in long-term degree programmes. Crucially, distinguishing between these groups is not possible and consequently both approaches – including or excluding individuals in education – lead to measurement error in the cohort-size variable. However, Table S9 shows the results when a cohort-size variable is constructed in which the numerator is derived only from individuals who are either employed or unemployed (to ensure a better comparison with the results in the paper, the construction of the denominator is left unchanged). In the unemployment model the 2SLS coefficients and corresponding marginal effects are similar in size to those reported in Table 3, while there is a decrease in the magnitude of the OLS estimates. In the employment model, there is a pronounced increase in the magnitude of the OLS coefficients and marginal effects, while the 2SLS effects are only slightly larger.

Table S9: OLS and 2SLS regression results (cohort-size variable from employed and unemployed individuals)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-6.15*** (1.63)	-18.13*** (2.30)	-3.72** (1.65)	-15.91*** (2.26)	-7.77*** (1.83)	-23.84*** (2.57)	-4.44** (1.86)	-20.73*** (2.44)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.37	0.34	0.40	0.38	0.41	0.38	0.45	0.41
F-stat	-	892.61***	-	920.50***	-	740.37***	-	857.18***
ME(std)	-0.02***	-0.06***	-0.01**	-0.05***	-0.03***	-0.08***	-0.02**	-0.07***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	21.50*** (1.99)	25.47*** (2.64)	18.93*** (2.04)	23.31*** (2.59)	21.48*** (2.38)	27.51*** (2.88)	18.04*** (2.46)	24.43*** (2.74)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.56	0.55	0.58	0.58	0.60	0.59	0.62	0.62
F-stat	-	892.61***	-	920.50***	-	740.37***	-	857.18***
ME(std)	0.07***	0.09***	0.06***	0.08***	0.07***	0.09***	0.06***	0.08***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Next we report aggregate-level results which are not derived from weights that have been modified so that the weighted sum of observations per region-year-age-sex cell matches the corresponding population values reported by Eurostat. Instead this analysis is based on the weights provided as part of the EU-SILC dataset which have only been modified to take account of the change in the number of rotational groups per year by appending different longitudinal releases (see Section 3.1 and Moffat and Roth, 2016). Table S10 shows that using calibrated weights does not affect sign and significance of the cohort-size coefficients, though it increases the size of the marginal effects.

Table S10: OLS and 2SLS regression results (non-calibrated weights)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-4.02*** (1.23)	-23.84*** (3.20)	-2.67** (1.19)	-20.87*** (3.09)	-4.19*** (1.39)	-28.36*** (3.45)	-2.62** (1.32)	-24.32*** (3.30)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.40	0.30	0.44	0.36	0.45	0.31	0.48	0.38
F-stat	-	242.66***	-	245.20***	-	215.47***	-	209.95***
ME(std)	-0.02***	-0.09***	-0.01**	-0.08***	-0.02***	-0.11***	-0.01**	-0.09***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	7.70*** (1.50)	35.48*** (3.95)	6.13*** (1.50)	32.08*** (3.82)	6.99*** (1.67)	34.52*** (4.09)	5.19*** (1.67)	30.08*** (3.91)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.57	0.47	0.60	0.51	0.62	0.53	0.64	0.57
F-stat	-	242.66***	-	245.20***	-	215.47***	-	209.95***
ME(std)	0.03***	0.13***	0.02***	0.12***	0.03***	0.13***	0.02***	0.11***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S11 shows the results when standard errors are estimated that are clustered at the level of the region-age cell, as is done in the individual-level analysis, instead of standard errors that are merely robust against heteroscedasticity. Despite the increase in the size of the standard errors, the cohort-size coefficients remain significant at the 1% level.

Table S11: OLS and 2SLS regression results (clustered standard errors)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-10.32*** (2.00)	-17.30*** (2.55)	-7.98*** (1.98)	-15.06*** (2.41)	-12.84*** (2.52)	-23.01*** (3.13)	-9.70*** (2.54)	-19.88*** (2.95)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
Region-age	243	243	243	243	243	243	243	243
R^2	0.38	0.37	0.41	0.40	0.43	0.41	0.45	0.44
F-stat	-	1,312.45***	-	1,561.99**	-	895.96***	-	1,226.07***
ME(std)	-0.03***	-0.05***	-0.02***	-0.05***	-0.04***	-0.07***	-0.03***	-0.06***

Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	14.39*** (2.32)	24.32*** (3.05)	11.91*** (2.39)	22.07*** (2.93)	13.88*** (2.79)	26.55*** (3.71)	10.63*** (2.82)	23.43*** (3.56)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
Region-age	243	243	243	243	243	243	243	243
R^2	0.53	0.52	0.56	0.55	0.58	0.57	0.61	0.60
F-stat	-	1,312.45***	-	1,561.99***	-	895.96***	-	1,226.07***
ME(std)	0.04***	0.08***	0.04***	0.07***	0.04***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Standard errors that are clustered at the level of the region-age group cell are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

In the paper, the empirical analysis is conducted for the age range 25-29 in order to avoid the estimated effects of cohort size on the unemployment and the employment share being confounded by the decision to enter the labour market or to acquire education. If indeed only a small share of individuals participates in education in this age range, we would expect to obtain similar results when the empirical analysis is restricted to the regression sample of individuals who are either employed or unemployed (notice that this restriction does not affect the construction of the cohort-size variable, which is population-based and therefore independent of the distribution of individuals across different labour-market states). The results in Table S12 show that if this restriction is imposed, the marginal effects for a change of one standard deviation increase slightly in the unemployment model (notice that since there are only two labour-market states in the sample, the coefficients in the employment model have the same magnitude but opposite sign compared to those in the unemployment model).

Table S12: OLS and 2SLS regression results (data aggregated from unemployed and employed individuals only)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-12.40*** (1.84)	-21.12*** (2.28)	-9.51*** (1.86)	-18.23*** (2.21)	-14.37*** (2.08)	-26.00*** (2.58)	-10.72*** (2.12)	-22.39*** (2.47)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.43	0.42	0.46	0.45	0.47	0.46	0.50	0.49
F-stat	-	1,499.69***	-	1,619.54***	-	1,104.85***	-	1,216.07***
ME(std)	-0.04***	-0.07***	-0.03***	-0.06***	-0.04***	-0.08***	-0.03***	-0.07***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	12.40*** (1.84)	21.12*** (2.28)	9.51*** (1.86)	18.23*** (2.21)	14.37*** (2.08)	26.00*** (2.58)	10.72*** (2.12)	22.39*** (2.47)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,959	1,959	1,959	1,959	1,959	1,959	1,959	1,959
R^2	0.43	0.42	0.46	0.45	0.47	0.46	0.50	0.49
F-stat	-	1,499.69***	-	1,619.54***	-	1,104.85***	-	1,216.07***
ME(std)	0.04***	0.07***	0.03***	0.06***	0.04***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Standard errors that are clustered at the level of the region-age group cell are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

As can be seen from Figures A2 and A3 in the Appendix there is fluctuation in the share of individuals in a particular labour-market state across age groups and over time for a given region. While these fluctuations may reflect 'true' variation in the dependent variables, it is also possible that they are the result of labour-market shares being derived from small cell sizes: if the number of observations per region-year-age cell is small, the estimated shares may no longer be representative of the actual distribution of labour-market status in the population. While measurement error in the dependent variable generally reduces estimation precision, estimates may also be biased if the fluctuations vary systematically with the cohort-size variable. In order to assess the sensitivity of the results we drop cells containing less than 3, less than 5 and less than 10 observations. As shown in Tables S13 to S15, the resulting coefficients and marginal effects are close to those reported in the paper.

Table S13: OLS and 2SLS regression results (cells with less than three observations are excluded)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-10.25*** (1.70)	-17.32*** (2.10)	-7.91*** (1.73)	-15.08*** (2.05)	-12.75*** (1.91)	-23.04*** (2.38)	-9.61*** (1.96)	-19.90*** (2.25)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,953	1,953	1,953	1,953	1,953	1,953	1,953	1,953
R^2	0.38	0.37	0.41	0.40	0.43	0.41	0.45	0.44
F-stat	-	1,538.49***	-	1,639.52***	-	1,138.38***	-	1,267.89***
ME(std)	-0.03***	-0.05***	-0.02***	-0.05***	-0.04***	-0.07***	-0.03***	-0.06***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	14.19*** (2.03)	24.15*** (2.64)	11.66*** (2.02)	21.88*** (2.52)	13.62*** (2.24)	26.39*** (2.89)	10.33*** (2.25)	23.26*** (2.70)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,953	1,953	1,953	1,953	1,953	1,953	1,953	1,953
R^2	0.53	0.53	0.56	0.55	0.58	0.57	0.61	0.60
F-stat	-	1,538.49***	-	1,639.52***	-	1,138.38***	-	1,267.89***
ME(std)	0.04***	0.08***	0.04***	0.07***	0.04***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S14: OLS and 2SLS regression results (cells with less than five observations are excluded)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-10.55*** (1.71)	-17.58*** (2.11)	-8.15*** (1.74)	-15.27*** (2.05)	-13.43*** (1.91)	-23.56*** (2.39)	-10.29*** (1.96)	-20.45*** (2.26)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937
R^2	0.38	0.38	0.42	0.41	0.43	0.42	0.46	0.45
F-stat	-	1,531.02***	-	1,637.04***	-	1,134.46***	-	1,272.82***
ME(std)	-0.03***	-0.05***	-0.03***	-0.05***	-0.04***	-0.07***	-0.03***	-0.06***

Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	14.48*** (2.03)	24.43*** (2.64)	11.86*** (2.02)	22.03*** (2.51)	14.36*** (2.22)	27.02*** (2.90)	11.03*** (2.23)	23.81*** (2.69)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937
R^2	0.54	0.53	0.57	0.56	0.58	0.58	0.61	0.61
F-stat	-	1,531.02***	-	1,637.04***	-	1,134.46***	-	1,272.82***
ME(std)	0.05***	0.08***	0.04***	0.07***	0.04***	0.08***	0.03***	0.07***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Table S15: OLS and 2SLS regression results (cells with less than ten observations are excluded)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-10.94*** (1.70)	-17.26*** (2.09)	-8.40*** (1.72)	-14.82*** (2.03)	-14.09*** (1.92)	-23.23*** (2.36)	-10.77*** (1.98)	-19.94*** (2.25)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,836	1,836	1,836	1,836	1,836	1,836	1,836	1,836
R^2	0.41	0.40	0.44	0.43	0.46	0.45	0.49	0.48
F-stat	-	1,512.88***	-	1,600.87***	-	1,106.95***	-	1,224.71***
ME(std)	-0.03***	-0.05***	-0.03***	-0.05***	-0.04***	-0.07***	-0.03***	-0.06***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	15.70*** (1.99)	25.50*** (2.48)	12.87*** (1.97)	22.85*** (2.35)	15.40*** (2.21)	27.82*** (2.78)	11.76*** (2.24)	24.21*** (2.61)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	1,836	1,836	1,836	1,836	1,836	1,836	1,836	1,836
R^2	0.57	0.56	0.60	0.59	0.62	0.61	0.65	0.64
F-stat	-	1,512.88***	-	1,600.87***	-	1,106.95***	-	1,224.71***
ME(std)	0.05***	0.08***	0.04***	0.07***	0.05***	0.09***	0.04***	0.08***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

The empirical analysis of the paper is based on a balanced panel of regions which can be observed throughout the whole sample period 2005-2012. Table S16 shows the cohort-size coefficients which are obtained if the following regions are not excluded from the analysis: 2

regions from Bulgaria (2006-2012), 1 region from Cyprus (2007-2012; due to unavailability of the instrumental variable, age group 25 can only be included from 2009 onwards), 1 region from Malta (2006-2012), 1 region from Norway (2008-2012) and 4 regions from Romania (2007-2012). The marginal effects are slightly smaller than those shown in the paper, but retain their sign and significance at the 1% level.

Table S16: OLS and 2SLS regression results (inclusion of all available regions)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-9.47*** (1.59)	-16.32*** (2.03)	-7.33*** (1.62)	-13.99*** (1.97)	-11.54*** (1.78)	-21.39*** (2.26)	-8.77*** (1.82)	-18.39*** (2.14)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	2,236	2,236	2,236	2,236	2,236	2,236	2,236	2,236
R^2	0.38	0.38	0.41	0.41	0.43	0.42	0.46	0.44
F-stat	-	1,531.47***	-	1,656.76***	-	1,124.69***	-	1,249.18***
ME(std)	-0.03***	-0.05***	-0.02***	-0.04***	-0.03***	-0.06***	-0.03***	-0.06***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	13.27*** (1.91)	22.91*** (2.55)	10.96*** (1.90)	20.42*** (2.43)	12.57*** (2.09)	24.71*** (2.76)	9.61*** (2.10)	21.39*** (2.58)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	2,236	2,236	2,236	2,236	2,236	2,236	2,236	2,236
R^2	0.53	0.52	0.56	0.55	0.57	0.57	0.60	0.59
F-stat	-	1,531.47***	-	1,656.76***	-	1,124.69***	-	1,249.18***
ME(std)	0.04***	0.07***	0.03***	0.06***	0.04***	0.07***	0.03***	0.06***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

Since the EU-SILC dataset also contains observations from some regions for the years 2004 and 2013, the sample period can in principle be extended by another two years, though this is only possible for the regions from the following set of countries: Austria (3 regions, 2004, 2013), Belgium (3 regions, 2004), Bulgaria (2 regions, 2013), Cyprus (1 region, 2013), Czech Republic (1 region, 2013), Denmark (1 region, 2004, 2013), Estonia (1 region, 2004, 2013), Greece (4 regions, 2004), Spain (7 regions, 2004, 2013), France (8 regions, 2004, 2013), Hungary (3 regions, 2013), Italy (3 regions, 2004, 2013), Italy (2 regions, 2013), Lithuania (1 region, 2013), Luxembourg (1 region, 2004, 2013), Latvia (1 region, 2013), Malta (1 region, 2013), Poland (6 regions, 2013) and Slovakia (1 region, 2013). As can be seen from Table S17, the cohort-size coefficients retain their sign and continue to be significant at the 1% level. In the unemployment model the size of the marginal effects is slightly smaller than in Table S16, whereas the size of the marginal effects in the employment model stays the same.

Table S17: OLS and 2SLS regression results (inclusion of all available regions and years)

Panel A: Unemployment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	-8.31*** (1.54)	-13.70*** (2.07)	-6.73*** (1.53)	-12.20*** (1.97)	-9.66*** (1.68)	16.86*** (2.26)	-7.67*** (1.69)	-14.82*** (2.16)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	2,606	2,606	2,606	2,606	2,606	2,606	2,606	2,606
R^2	0.38	0.37	0.40	0.40	0.42	0.41	0.44	0.43
F-stat	-	2,104.78***	-	2,309.70***	-	1,674.20***	-	1,899.38***
ME(std)	-0.03***	-0.04***	-0.02***	-0.04***	-0.03***	-0.05***	-0.02***	-0.05***
Panel B: Employment	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Cohort size	12.78*** (1.76)	21.60*** (2.36)	10.99*** (1.75)	19.63*** (2.26)	12.21*** (1.90)	22.34*** (2.52)	10.10*** (1.91)	19.67*** (2.41)
Dummies								
Region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-by-age	No	No	No	No	Yes	Yes	Yes	Yes
Control variables	No	No	Yes	Yes	No	No	Yes	Yes
Observations (cells)								
Region-year-age	2,606	2,606	2,606	2,606	2,606	2,606	2,606	2,606
R^2	0.54	0.53	0.56	0.55	0.57	0.57	0.60	0.59
F-stat	-	2,104.78***	-	2,309.70***	-	1,674.20***	-	1,899.38***
ME(std)	0.04***	0.07***	0.03***	0.06***	0.04***	0.07***	0.03***	0.06***

***/**/* indicate significance at the 1%/5%/10% level, respectively. Robust standard errors are shown in parentheses. The regression is weighted by the estimated number of male observations in a region-year-age cell. *F-stat* represents the first-stage F-statistic from a regression of the endogenous cohort-size variable on the instrument and control variables. *ME(std)* shows the change in the dependent variable if the cohort-size variable increases by one standard deviation.

References

Angrist, J. and Pischke, J.-S. (2009) *Mostly harmless econometrics: an empiricist's companian,* Princeton University Press, Princeton.

Cameron, A.C. and Trivedi, P.K. (2009) *Microeconometrics Using STATA,* STATA Press, College Station.

Moffat, J. and Roth, D. (2016) The Cohort Size-Wage Relationship in Europe, *Labour*, forthcoming.

Moulton, B.R. (1990) An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units, *Review of Economics and Statistics,* **72**, 334-38.

Shimer, R. (2001) The Impact of Young Workers on the Aggregate Labor Market, *The Quarterly Journal of Economics*, **116**, 969-1007.

Wright, R.E. (1991) Cohort size and earnings in Great Britain, *Journal of Population Economics*, **4**, 295-305.