MAGKS

Aachen · Siegen · Marburg · Gießen · Kassel · Göttingen

Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen Kassel · Marburg · Siegen

ISSN 1867-3678

No. 06-2017

Alfred Garloff and Duncan Roth

Regional population structure and young workers' wages

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

Regional population structure and young workers' wages

Alfred Garloff[1] - Duncan Roth[2]

Abstract

This paper estimates the effect that changes in the size of the youth population have on the wages of young workers. Assuming that differently aged workers are only imperfectly substitutable, economic theory predicts that individuals in larger age groups earn lower wages. We test this hypothesis for a sample of young, male, full-time employees in Western Germany during the period 1999-2010. In contrast to other studies, functional rather than administrative spatial entities are used as they provide a more accurate measure of the youth population in an actual labour market. Based on instrumental variables estimation, we show that an increase in the youth share by one percentage point is predicted to decrease a young worker's wages by 3%. Our results also suggest that a substantial part of this effect is due to members of larger age groups being more likely to be employed in lower-paying occupations.

JEL classification J21, J31, R23

Key words Population structure, wages, youth share, labour-market regions, instrumental variables, occupational selection

[1] Institute for Employment Research (IAB), Saonestraße 2-4, 60528 Frankfurt am Main, Germany

[2] Institute for Employment Research (IAB), Josef-Gockeln-Straße 7, 40474 Düsseldorf, Germany; Phone: +49 (0)211 4306 108, Fax: +49 (0)211 4306 910468, Email: duncan.roth@iab.de

1 Introduction

Germany is in the middle of a demographic transition. The size of its population was on the decline between 2003 – when positive net immigration started falling short of the natural population decrease – and 2010 and is projected to continue shrinking over the coming decades, falling by 11% between 2010 and 2040 (Statistisches Bundesamt, 2009).[1] However, this transition also has a second dimension: during the second half of the twentieth century fertility rates declined permanently and eventually fell below replacement level. Coupled with increases in life expectancy, these processes are having a substantial effect on the age structure of Germany's population as evidenced by the ongoing increases in the size of older age groups at the expense of younger ones.

Between 1990 and 2010 the ratio of the working-age to the total population fell by over three percentage points, a downward trend that is expected to be exacerbated by the entry into retirement of the large post-World War II birth cohorts. Moreover, demographic change has affected the age composition of the working-age population: while the share of individuals aged 15-24 in the working-age population increased between 2000 and 2010, this development is expected to reverse in the near future with the youth share projected to fall by 2.5 percentage points between 2010 and 2025. The implications of these changes – the combination of a shrinking and ageing population – for the future standard of living constitutes a widely discussed area of research (see Börsch-Supan, 2013). In this context, the question of how labour productivity will be affected by the changes in the population-age structure will be of prime importance (see Bloom and Sousa-Poza, 2013). Likewise, the sustainability of health care and public pension systems in light of demographic pressure has received considerable attention (see Arnds and Bonin, 2002; Jimeno et al., 2008).

The objective of this paper is to empirically analyse the impact of changes in the size of the youth population within regional labour markets on the wages of young workers. In the light of the projected population developments, this type of analysis is relevant as it provides a basis for evaluating how demographic processes can be expected to affect the wages of future cohorts of young workers. Given its focus, this paper belongs to a larger body of literature that

[1] To ensure comparability with the empirical analysis of this paper, the reported numbers refer to Western Germany (excluding West Berlin). With the availability of the 2011 census, the basis for estimating population variables has changed. As the population measures in this paper are based on pre-census data, we also use the population projections that are derived from this data rather than the recently released projections that make use of the 2011 census.

analyses the effects of changes in the age structure on labour-market outcomes. In addition to wage adjustments, a considerable amount of research has addressed the impact on age-specific (un-)employment (Zimmermann, 1991; Shimer, 2001; Skans, 2005; Biagi and Lucafora, 2008; Ochsen, 2009; Garloff et al., 2013; Moffat and Roth, 2016b) and educational attainment (Connelly, 1986; Stapleton and Young, 1988; Fertig et al., 2009).

While wage differences and wage trends between different cohorts in Western Germany are documented in Fitzenberger (1999), his analysis does not focus on the consequences of changes in the age structure, which is the concern of this paper. In a world with a single type of labour input, an increase in the size of the labour force will lead to an outward shift of the labour supply curve. If the labour market works in a way that the wage rate adjusts so as to equate the demand for and the supply of labour and diminishing marginal productivity implies a downward-sloping labour demand curve, the effect of an increase in the labour force will be a lower equilibrium wage rate. If instead labour inputs are not homogenous but rather only imperfectly substitutable across age groups, the effects of a change in age-specific labour supply will – depending on the degree of substitutability – be concentrated on the members of that age group. Within such a framework, an increase in the share of young individuals should be accompanied by a decrease in their wages.

Our contribution is threefold. First, our assessment of the relationship between the youth share and young workers' wages in Western Germany addresses the lack of recent empirical evidence on this topic. Second, we use functional entities in order to identify the size of the youth population within an actual labour market rather than within an administrative unit as is done by earlier studies, which reduces the potential for measurement error in this variable. Third, we assess the channels through which changes in the size of the youth population affect young workers' wages by controlling for industrial and occupational up- or downgrading. Gertler and Trigari (2009) argue that individuals have a better chance of moving into higher-paying industries, firms or jobs during boom periods than during recessions. We propose that a similar argument can be made with respect to age-group size, as increased competition may lead individuals to take up positions in lower-paying industries or occupations than they would have done, had they been part of a smaller age group. In order to distinguish between the direct and the selection-related, indirect effect of belonging to a larger age group, we compare the estimated wage effect of the youth share from models that exclude or include detailed information about an individual's industrial and occupational affiliation.

In our model the effect that the regional youth share has on the wages of young workers is identified solely through the within-variation of this variable. However, as the relative size of the youth population within a labour market is potentially endogenous due to migration into high-wage areas, an instrumental variables (IV) identification strategy is employed: within a given region the instrument is defined as the share of individuals that are fifteen years younger and that are observed fifteen years earlier than the age group of the endogenous regressor. We find that the youth share has a statistically significant negative effect on the wages of young workers. Specifically, an increase by one percentage point is predicted to decrease wages by 3% in our baseline model. When using a district-based measure of the youth-share variable, the estimated coefficients are smaller by between 13% and 48%. Finally, we find that controlling for an individual's industry and, particularly, occupation reduces the estimated wage decrease from 3% to 2%, which suggests that a substantial part of the negative effect of age-group size is the result of individuals in larger age groups being more likely to be employed in lower-paying occupations. According to these results, future generations of young workers can expect to benefit from demographic developments. Specifically, a decrease in the youth share by 2.5 percentage points, as projected to occur between 2010 and 2025, would be predicted to lead to an increase in young workers' wages of about 5%, ceteris paribus.

The remainder of the paper is structured as follows. Section 2 addresses the relationship between age structure and wage outcomes and reviews the relevant theoretical and empirical literature. Section 3 provides descriptive statistics on the youth population in Germany. The empirical analysis is the topic of Section 4, while Section 5 discusses the regression results. Section 6 presents the conclusion.

2 Population structure and wages

Differently aged workers are not perfectly substitutable. Age can be expected to be correlated with a worker's set of skills, which in turn affects his suitability for different tasks. First, age is a good predictor for work experience, and, ceteris paribus, more experienced workers will usually have more firm-specific, occupation-specific, industry-specific or general human capital. If this type of knowledge is relevant for on-the-job performance, differently aged workers can be expected to be only imperfectly substitutable. Indeed, Welch's (1979) career-phase model can be interpreted as an example of a model in which imperfect substitutability arises from differences in firm-specific human capital. Second, jobs vary with respect to the tasks that they contain and therefore also concerning the abilities that workers are required to have in order to perform these tasks. Older workers may be less easily substitutable for younger workers in occupations requiring physical or certain types of cognitive skills (Mazzonna and Perracchi, 2012). As a consequence of imperfect substitutability a change in the relative size of an age group will mainly affect the labour market outcomes of the members of that group.

As a starting point to analysing the effects of a change in the size of a specific age group on the wages of its members, it is useful to assume a production function with differently aged workers as distinct factors of production (see Card and Lemieux, 2001; Fitzenberger and Kohn, 2006). In the benchmark case of a perfectly competitive labour market, in which each factor of production is paid the monetary value of his marginal product, a change in the supply of a specific production factor will cause the wage to adjust in a way that the market is again cleared. In the case of each factor of production exhibiting diminishing marginal productivity, an increase in the size of an age group will reduce the wages paid to its members. Labour markets, however, do not necessarily clear. The existence of minimum or efficiency wages as well as collective wage bargaining are possible sources that can prevent the wage rate from fully adjusting in response to a change in labour supply, while the coexistence of unemployment and vacancies provides evidence against the existence of a market-clearing equilibrium as predicted by the benchmark model of a competitive labour market. Existing theoretical models, however, suggest that even in the absence of clearing labour markets, changes in the relative supply of an age group will have an effect on age-specific wages (Michaelis and Debus, 2011).

The extant empirical literature, though differing with respect to the time periods and countries (or regions) under study, the model specification and identification strategy, provides evidence that increases in the size of an age group are associated with depressed wage outcomes for the members of that group.[2] Early studies using US data estimate a negative

[2] Notable exceptions can be found in the migration literature where many studies conclude that natives' wages are not negatively affected by age-specific immigration (Ottaviano and Peri, 2012). A possible explanation for this finding is that migrants are complements rather than substitutes for native labour.

relationship between the relative size of an age group and the average wages that are earned by individuals within that group for different levels of educational qualification (Welch, 1979; Berger, 1985). Alternatively, Freeman (1979) finds a negative effect of the young-to-old population ratio on the average wages of young workers relative to those of old workers. The existence of a negative effect of age-group size is also supported by evidence from Sapozhnikov and Triest (2007). Most recently, Morin (2015) exploits an exogenous shock to the supply of high-school graduates in Canada due to a reform of the secondary schooling system and finds negative cohort-size effects on wages. Empirical evidence from Europe is scarcer but also supports the hypothesis that wages earned in larger age groups are depressed compared to those of smaller age groups (see Wright, 1991, for the UK and Brunello, 2010, as well as Moffat and Roth, 2016a, for a sample of European countries).

A drawback with respect to identifying the effect of interest is that the size of an age group within a given spatial unit is arguably endogenous due to self-selection of individuals into high-wage areas. Korenman and Neumark (2000) proposed birth rates as an instrument, while other authors have since used the lagged relative size of age groups as exogenous predictors (Skans, 2005; Garloff et al., 2013; Moffat and Roth, 2016a and 2016b). Whereas cross-country migration might be deemed too small to influence the size of nationally defined age groups, endogeneity resulting from self-selection through migration becomes a larger concern when the spatial units that are used to construct the measure of population structure are defined at a sub-national level.

While many empirical studies in this field of research have used measures of population structure at the national level, it appears questionable whether a country indeed constitutes the appropriate delineation of a labour market. If individuals are restricted in their mobility or if awareness of job openings in other regions decreases with distance, a nationally defined youth-share variable groups together young individuals that are not active in the same labour market and that are hence not substitutable for one another. Such a variable would be subject to measurement error if labour markets existed at a sub-national level and the size of the youth population varied across them. And while more recent studies have made use of administrative units at a sub-national level, so-constructed youth-share variables may still be measured with error as administrative units are generally not delineated in a way as to coincide with actual labour markets, meaning that they would not necessarily capture the relative supply of young labour that is relevant for the determination of a young worker's wages. To address this issue, we employ the functional labour-market regions that are defined by Eckey et al. (2006). These regions consist of one or more districts (*Kreise*) and are constructed on the basis of observed commuting flows with a typical labour-market region combining an economic centre with the surrounding *Umland* from which people commute to work in the centre. They approximate self-contained local labour markets in as far as they aim to maximise the overlap between the population living and working within such a region. Functional units therefore provide a better measure of the size of the youth population in an actual labour market than administrative units. The self-contained nature of these units also reduces the need to consider the youth population in surrounding labour markets as a factor determining the wages of young workers in a given region.

It should be noted that changes in the age structure of the population do not necessarily imply changes in age-specific labour supply as participation rates as well as the number of hours worked could in principle adjust in a way as to completely counteract changes in age-group size. However, such a reaction seems unlikely as empirical evidence suggests that male labour supply is inelastic – at the extensive and the intensive margin – to changes in the wage rate (Blundell and MaCurdy, 1999). More specifically, Garloff et al. (2013) show that a counteracting development in participation rates has not taken place in Germany in response to changes in the age structure at the national level in recent years.

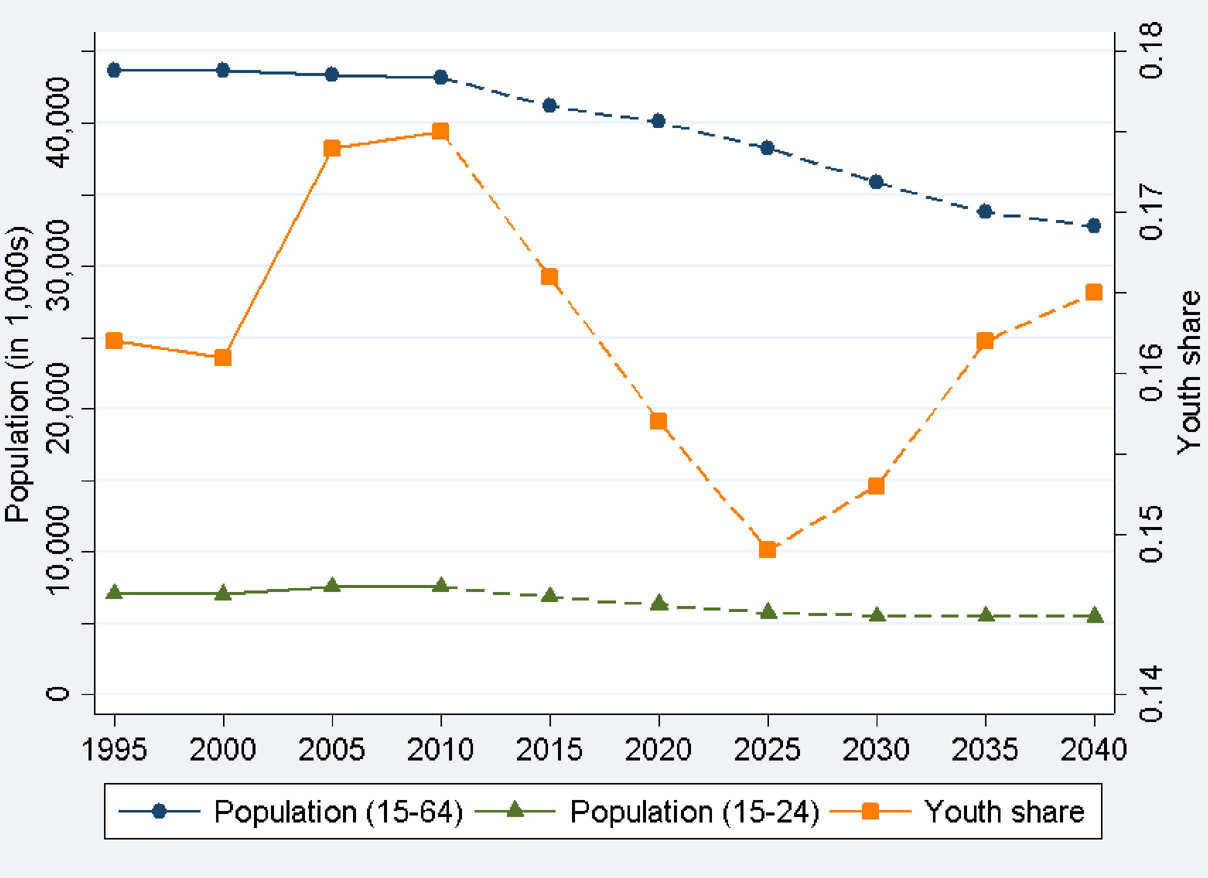
3 Youth-population structure in Western Germany

This section provides information about the development of the working-age (15-64) and the youth population (15-24) in Western Germany at the national level and at the level of the labour-market region. Figure 1 shows the absolute size of both populations at five-year intervals between 1995 and 2040. While the actual values are shown up to the year 2010[3], subsequent developments represent projections based on the variant *Untergrenze der mittleren Bevölkerung*, which assumes an annual net immigration of 100,000 individuals and

[3] Data comes from the Federal Statistical Office and has been obtained through the following link: https://www-genesis.destatis.de/genesis/online/link/tabellen/12411*

a fertility rate of 1.4 and which represents the lower bound of corridor within which population development is expected to take place (Statistisches Bundesamt, 2010).[4]

Figure 1: Development of the population and the youth share at the national level



Source: Federal Statistical Office

Except for a small increase between 1995 and 2000, the working-age population has been shrinking steadily and is projected to continue decreasing in size over the coming decades. By 2040 it will have fallen by almost 25% compared to its 2010 value, which reflects the effect of the large post-World War II birth cohorts reaching retirement age. In contrast, the number of young individuals grew by half a million between the years 2000 and 2010[5], but this development is expected to reverse in the near future with the size of the age group 15-24 projected to fall continuously until 2040. Reflecting changes in these two populations' relative rate of growth, the youth share, i.e. the size of the population aged 15-24 relative to the working-age population, displays a cyclical development: from 2000 to 2010 the share of

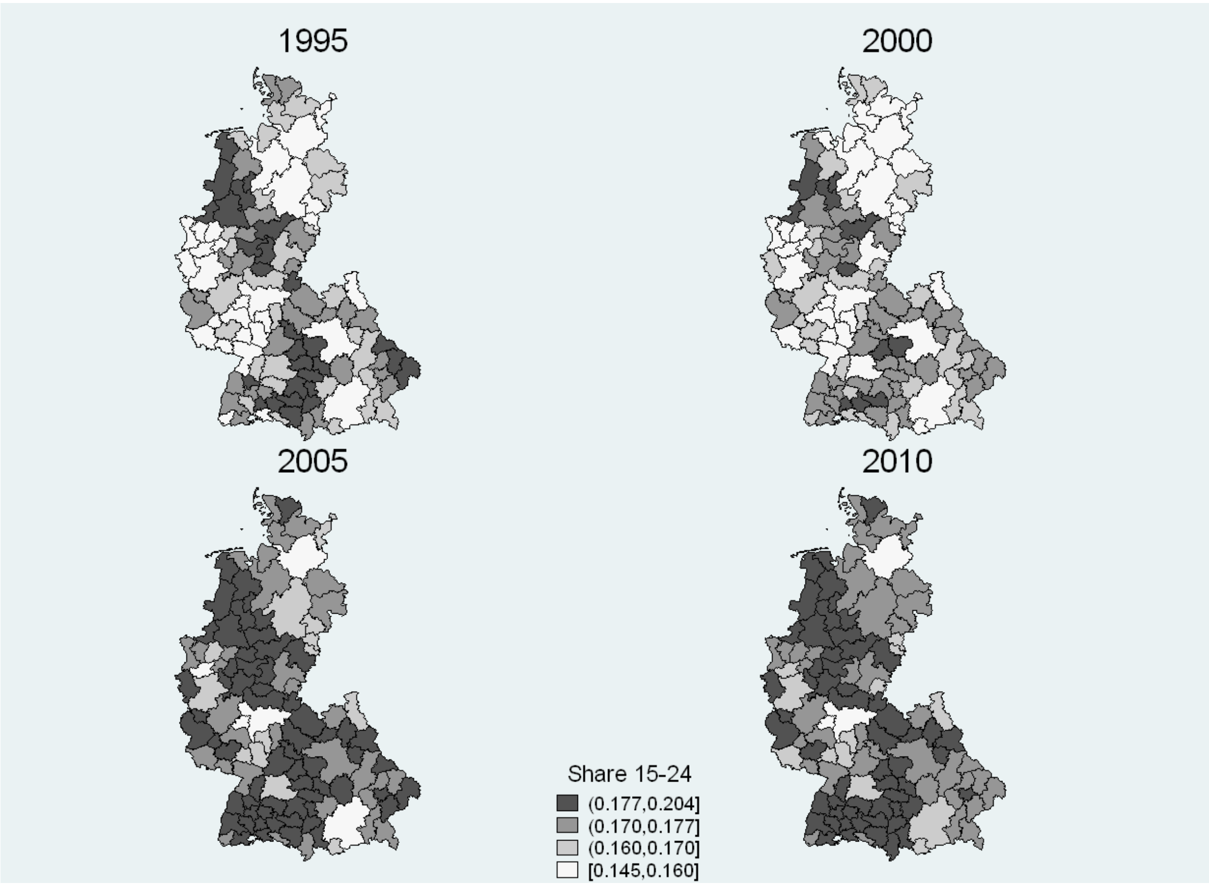
[4] The upper bound of this corridor (*Obergrenze der mittleren Bevölkerung*) differs by assuming that annual net immigration will increase steadily to 200,000 in the year 2020 before plateauing at that level. Despite this difference the projection for the youth share is very similar (the largest difference between both projections amounts to 0.25 percentage points in the year 2040).

[5] These age groups are the children of the large post-World War II birth cohorts. This increase therefore reflects the large size of the parental generation.

young individuals expanded by approximately one percentage point (equivalently, 7%). However, as the youth population is expected to decrease at a faster rate than the working-age population, its share is projected to fall by 2.5 percentage points (equivalently, 15%) between 2010 and 2025. At the national level, the increase in the youth share during most of the sample period therefore contrasts with its projected development in the immediate future, which implies that changing demographics may contribute positively towards the development of young workers' wages in the coming years.

Figure 2 illustrates the existing regional heterogeneity in the share of individuals aged between 15 and 24 in the working-age population by reporting the value of this variable for the West-German regional labour markets. The extent of cross-sectional variation in the youth-share variable is revealed for the year 1995 in the top left map, in which the labour-market regions are grouped into quartiles based on the size of the youth share. Compared to a value of about 16% at the national level, the regional youth share varies between 14% and 21%.

Figure 2: Variation in the youth share (15-24) at the regional level



Source: Federal Statistical Office (population data) and Federal Institute for Research on Building, Urban Affairs and Spatial Development (geodata)

The other maps show the cross-sectional variation in the youth-share variable for the years 2000, 2005 and 2010, respectively. Moreover, they reveal the within-region variation in this variable, i.e. its development over time (to allow for a comparison of the different years, the same intervals are chosen as for the year 1995). Reflecting the drop in the national youth share in the year 2000, the share has also generally fallen at the regional level as illustrated by a number of regions that were in the fourth or third quartile in 1995 now being in the third or second quartile, respectively. Likewise, an increasing number of regions are registered in higher quartiles in the years 2005 and 2010, reflecting the increase in the youth share at the national level.

4 Empirical analysis

The different steps of empirically analysing the relationship between the youth share and young workers' wages are the subject of this section: the relevant datasets are introduced in the first part, which is followed by a description of how the sample is constructed and how the model's main variables are defined. The final part discusses the empirical model and the identification strategy.

4.1 Data

Three data sources are used for the empirical analysis. The first source is population data for Germany on the regional level according to age groups which is used to construct the relative size of the youth population within a regional labour market. The information reported by the statistical offices refers to the end of the year (31 December). There is no information beyond age and sex in these data. Particularly, there is no information available on the educational composition. Corrections have to be made to account for changes in the delineations of municipalities and districts which results in a dataset that is spatially consistent over time back until 1978. However, the available age brackets differ for the time before and after 1985. Second, we use statistics from the Federal Employment Agency (FEA) to gather information on employment numbers and rates as well as unemployment rates. Employment numbers and rates can be obtained at the level of the labour-market regions starting in 1987 for employment at place of work and from 1999 for employment at place of residence. The data is available by single-age cohorts, sex and education and refers to the middle of the year (30 June), since those values are typically close to yearly averages. In order to better compare the results from a model using a youth-share variable based on an individual's place of residence with those derived from an individual's place of employment, the year 1999 is chosen as the start of the sample period.

The final source is the *Stichprobe der Integrierten Erwerbsbiografien* (SIAB), a large micro-dataset from the Institute for Employment Research (IAB), that includes information on a 2% random sample of all individuals in Germany that were employed, unemployed or participating in measures of active labour-market policy between 1975 and 2010 (civil servants and the self-employed are excluded). For employed individuals in the dataset we have information on their employment relationship on a daily basis. Moreover, it contains a wealth of additional information that we use in part as control variables. The data further contains information about an employee's place of residence and place of employment, though the former only becomes available in 1999. A detailed description of the dataset can be found in vom Berge et al. (2013).

4.2 Sample and descriptive statistics

The observations contained in SIAB refer to spells of an individual (e.g. an employment spell) with given start and end dates as well as characteristics of the spell (e.g. the average daily wage earned during this period). We use the setting-up routines by Eberle et al. (2013) to transform the structure of the data so that it contains data from a single spell per individual and year. In doing so, we choose 15 June as the annual reference date, which means that only those spells are retained that include the reference date in a given year. As employers are required to report the wages of their employees once a year and this is typically done on 31 December, the longest spells run from 1 January to 31 December in a given year. Using 15 June as the reference date implies that spells starting and ending before (or after) 15 June within a given year are not being considered. This specific reference date is chosen because June values of employment figures are usually close to annual averages, while the middle of the month is used to avoid any end-of-calendar-month effects. However, the results are robust to using 31 December as the reference date.[6]

The sample covers the period 1999-2010 and consists of regularly employed (*sozialversicherungspflichtig Beschäftigte*) males who are between 15 and 24 years old. Individuals in vocational training are excluded because the mechanisms determining their

[6] The results of this and all other robustness checks can be found in the Supplementary Material to this paper.

remuneration are considered to be different from the rest of the labour market. As there is no information about the number of hours worked in the data, the sample is further restricted to full-time employees. While 95% of the observations have one full-time job, some observations hold other jobs in addition to being full-time employed, e.g. 3% of observations are also in minor employment (*geringfügige Beschäftigung*). In such a case only information about the first full-time job is retained.[7] We do not restrict employment spells to have a minimum duration. However, the results are robust to keeping only observations with employment spells of at least 90 days in the sample.

The model's dependent variable is an individual's inflation-adjusted daily wage including social security contributions and taxes.[8] The reported wage is censored at the value of the corresponding year's upper social security threshold; but given that our sample is restricted to individuals aged between 15 and 24 only a small fraction of observations will have wages above the threshold, and since imputation procedures (see Gartner, 2005) suggest that in such a case the true wage values are close to the censoring value, we use the censored wage for these observations. At the other end of the spectrum, we also observe unrealistically low daily wages. To remove these observations we truncate the wage distribution at twice the value of the minor-employment threshold (*Geringfügigkeitsgrenze*) – an approach that has also been taken by other authors working with the same data source (e.g. Gürtzgen, 2016). This implies that observations with wages of less than 650 Euro per month (21.26 Euro per day or, alternatively, 2.57 Euro per hour, assuming an eight-hour working day) between 1999 and 2002 or less than 800 Euro per month (26.28 Euro per day or 3.29 Euro per hour) between 2003 and 2010 are dropped.[9]

[7] For individuals holding more than one job at the same time it would in principle be possible to use total earnings from all jobs rather than just the wage earned in one job as the relevant dependent variable. We abstain from doing so as our focus is on how the supply of young workers affects the wages earned in a particular job. Similar results to those shown in Table 1 are obtained when observations with more than a full-time job are removed from the sample.

[8] Inflation-adjustment is done using the consumer price index (base year: 2010). The data comes from the Federal Statistical Office and has been obtained through the following link: https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Preise/Verbraucherpreisindizes/Tabellen_/VerbraucherpreiseKategorien.html?cms_gtp=145110_slot%253D2&https=1

[9] If individuals with wages below the specified thresholds are not excluded from the analysis, the youth-share coefficients are smaller in size and less significant. Compared to the sample used in the empirical analysis of this paper, individuals below the threshold are more likely to have a lower secondary education without apprenticeship training (56% compared to 19%) and are employed in firms with on average a smaller number of employees (446 compared to 970), whereas the average size of the youth share is similar. In addition to measurement error in the wages, the decrease in the effect of the youth share might also be due to the wages of this group being less responsive to changes in the supply of young workers, possibly because they are downward-rigid due to institutional constraints (e.g. sector-specific minimum wages).

The main variable of interest is the youth share, which measures the number of individuals aged between 15 and 24 relative to the number of working-age individuals (ages 15-64) within a regional labour market as defined by Eckey et al. (2006).[10] Due to limitations pertaining to the availability of population data preceding re-unification, our empirical analysis is restricted to the 108 labour-market regions (313 districts) of Western Germany. This restriction is unfortunate: the demographic processes that have seen the youth share in Eastern Germany fall from 19% in 2004 to 14% in 2012 (Fuchs and Weyh, 2014) certainly warrant an analysis of the corresponding wage effects.

Using a sub-national variable allows us to identify the effect of the youth share on young workers' wages while also controlling for macroeconomic shocks at the national level in a flexible way. As discussed in Section 2, the main advantage of employing labour-market regions as opposed to administrative units is that they provide a more accurate measure of the size of the youth population in an actual labour market, thereby reducing the potential for measurement error. For comparative purposes, however, we also estimate a model using a youth-share measure that is based on districts, which represent administrative units at the third level of the Nomenclature of Territorial Units for Statistics (NUTS 3). Furthermore, we are able to define two versions of the youth-share variable that refer to either the relative size of the youth population within the labour-market region (or district) that an individual works in or within which he resides. Owing to the way in which labour-market regions are designed, the fraction of observations for which the region of residence and the region of employment are identical stands at 85%, whereas the value is considerably smaller in the case of districts (66%).

A range of control variables are included in the model. At the individual level, SIAB contains information on age and labour-market experience as well as on an employee's level of education and his nationality. At the firm level, we use the size of the establishment and, in an extension to the baseline model, we also include two-digit indicators for an individual's occupation and industry which allows us to address the issue of industrial and occupational up- and downgrading (see Gertler and Trigari, 2009). In order to control for local macroeconomic effects we use the region-specific (district-specific) unemployment rate and,

[10] Similar results are obtained when we use an employment-based youth-share variable that is defined as the number of employed youths aged 15-24 relative to the workforce.

as the corresponding youth-unemployment rate is not available, the share of unemployed young individuals in the population. Descriptive statistics of the variables included in the baseline model are shown in Table A1 in the Appendix.

The average log real daily wage earnings are equal to 4.28 (approximately 72.24 Euro). The share of individuals aged between 15 and 24 in the working-age population is 17%. Since only employed individuals are included in the sample and individuals in vocational training are not considered, over 95% of observations are 20 years or older and a similar share has acquired up to four years of work experience. In terms of educational qualification the sample is rather homogenous as more than nine out of ten observations have lower secondary education and about three quarters of the cases also have a completed apprenticeship. The average firm size is slightly below 1,000 employees, while the regional unemployment rate has a mean value of about 8%, which is slightly higher than the share of unemployed youths in the population.

4.3 Empirical model and identification

In order to estimate the relationship between the wages of young workers and their relative supply, we specify an enhanced Mincer equation (Mincer, 1958) and regress the natural logarithm of an individual's inflation-adjusted daily wage earnings w_{irt} on the youth share y_{rt} and a set of control variables x_{irt} as formulated in Equation 1.[11] The indexes *i*, *r* and *t* denote individuals, spatial units and years, respectively. As described in the previous sub-section, separate models are estimated in which the spatial unit refers either to an individual's place of residence or to the place of employment. The variables δ_r and μ_t represent dummies for the spatial unit an individual resides or is employed in and for the sample year, respectively. Due to the inclusion of the region dummies it is only the within-region variation from which the coefficient of the youth share is identified. The error term ε_{irt} captures stochastic shocks as well as the effects of all other variables that are not explicitly controlled for[12]:

$$\log(w_{irt}) = \alpha + \beta y_{rt} + \boldsymbol{x_{irt}}'\boldsymbol{\gamma} + \delta_r + \boldsymbol{\mu_t} + \varepsilon_{irt} \quad [1]$$

[11] The specification of Equation 1 can also be interpreted as a special case of the model provided by Card and Lemieux (2001) in as far as our analysis also assumes imperfect substitutability across age groups but considers only the age group 15-24 in the empirical analysis.

[12] We abstain from estimating a model that includes fixed effects at the individual level. Since 44% of observations come from individuals that are included in the sample only once, estimation of such a model suffers from an insufficient degree of within-variation. Notice that for consistent estimation of the youth share's marginal effect, a fixed effects approach would only be required in the presence of unobserved, time-invariant heterogeneity at the individual level that is correlated with the youth share.

Consistent estimation of the effect that the youth share has on the wages of young workers by pooled ordinary least squares (OLS) requires that the regressor y_{rt} be conditionally uncorrelated with the error term. We argue that this requirement is unlikely to hold because individuals are able to self-select into regions where they can expect to earn higher wages, ceteris paribus, thereby turning the youth share into an endogenous variable. This endogeneity can be viewed as being the result of either omitted variables or reverse causality. The underlying mechanism is shown in the following set of equations:

$$\log(w_{irt}) = \alpha^a + \beta^a y_{rt} + \boldsymbol{x_{irt}}'\boldsymbol{\gamma^a} + \delta_r^a + \mu_t^a + \boldsymbol{\psi_{rt}}'\boldsymbol{\chi^a} + \varepsilon_{irt}^a \qquad [2a]$$

$$y_{rt} = \alpha^b + \beta^b w_{rt} + \boldsymbol{x_{rt}^b}'\boldsymbol{\gamma^b} + \delta_r^b + \mu_t^b + \boldsymbol{\psi_{rt}}'\boldsymbol{\chi^b} + \varepsilon_{rt}^b \qquad [2b]$$

First, there might be unobserved regional characteristics (e.g. regional industrial structure, regional labour-market conditions), ψ_{rt}, that jointly determine a young individual's wages (Equation 2a) as well as his decision to reside (work) in a specific region and hence the size of the youth share (Equation 2b). Assuming that individuals are likely to select into regions with characteristics that are favourable to their earnings ($\chi^a>0$ and $\chi^b>0$), pooled OLS estimates of the coefficient β in Equation 1 will be on average less negative than its true value (or even positive). The use of regional dummy variables, which capture unobserved time-invariant regional heterogeneity, and regional unemployment variables should help us to control for these characteristics. In an extension, we also fit a model with fixed effects for state-year combinations to further control for unobserved heterogeneity.

Second, even in the absence of omitted regional characteristics, endogeneity may yet arise from reverse causality. In Equation 2b the youth share is modelled as a function of the mean daily log-earnings of young workers in a given region, w_{rt}. As this variable is a linear function of the variable $\log(w_{irt})$, it follows that the youth share is correlated with the error term of Equation 1.[13] If the size of the youth share depends positively on the mean earnings of young workers in that region ($\beta^b>0$), the correlation between ε_{irt} and y_{rt} in Equation 1 will be positive. Under these assumptions and assuming further that the youth share has a negative effect on individual earnings ($\beta<0$), pooled OLS estimates of the corresponding coefficient would be expected to be less negative compared to the true value (or even positive). What is therefore

[13] Specifically, $w_{rt} = \frac{1}{N}\sum_{i=1}^{N}\log(w_{irt})$.

required to identify the true relationship between individual wages and the youth share is a source of exogenous variation in the latter variable.

To consistently estimate the causal effect that changes in the youth share have on the earnings of young workers we employ an IV strategy. Our instrument is the variable that has also been used by Skans (2005), Garloff et al. (2013) and Moffat and Roth (2016a, 2016b). This variable is defined as the relative size of the group of individuals who are 15 years younger than the age group on which the youth-share variable is based and who are observed 15 years earlier, i.e. we instrument the current share of those aged 15-24 (relative to the age group 15-64) with the share of those aged 0-9 (relative to the age group 0-49) 15 years earlier. The strength of the instrument derives from the fact that in the absence of migration and natural population changes the instrument and the youth-share variable would be based on the same group of individuals and both variables would actually be identical. We argue that migration and natural changes do not purge the association between the instrument and the endogenous regressor, meaning that if an age group in a given spatial unit was comparatively large (relative to the size of the same age group in other spatial units and years), the group of individuals in the same region who are 15 years older will still be relatively large in the present. This argument is supported by the results of the first-stage statistics, which show the instrument to have a high degree of explanatory power.

The identifying assumption is that individuals in the age group 0-9 do not choose where to reside based on the anticipation of their earnings 15 years in the future. If this condition is satisfied, the causal effect of the relative supply of young individuals in a given spatial unit on young workers' earnings can be identified by using the two-stage least squares (2SLS) estimator with the time-lagged and age-lagged population variable as an instrument. An argument that can be brought forward against the validity of the instrument is that the relative size of the age group 0-9 will depend on the locational choices of their parents. If parents, and thus their children, self-selected into high-wage areas and their wages were correlated with the wages of their children fifteen years in the future, this would lead the proposed identification strategy to fail. Notice, however, that if the parental generation's choice of location and the correlation between their own and their children's wages are due to time-invariant factors, these will be accounted for by the region dummies of Equation 1.

Another source of endogeneity due to omitted variables is the fact that we only observe daily but not hourly wages. If the number of hours worked varies systematically with the youth share, pooled OLS estimation will again produce inconsistent results, but as long as the supply of hours is uncorrelated with the proposed instrument, 2SLS estimation will be consistent. Finally, a feature of the model in Equation 1 is that the explanatory variable of interest, y_{rt}, is defined at a higher level of aggregation than the dependent variable, which also varies across individuals.[14] To account for this feature we cluster at the level of the spatial unit in order to avoid biased standard errors (see Moulton, 1990).

5 Results

Table 1 shows the results of estimating the baseline specification of Equation 1 (i.e. excluding indicators for an individual's industry and occupation). In the first two columns labour-market regions refer to an individual's place of residence, while the results for the place of employment are shown in the third and the fourth column. In both cases, the model is estimated by OLS as well as by 2SLS.

In line with the prediction that, ceteris paribus, members of larger age groups earn lower wages, the youth-share variables draw negative and statistically significant coefficients when estimated by 2SLS. Measured at an individual's place of residence, a decrease in the youth share by one percentage point is predicted to increase a young worker's wages by 3.2%. The corresponding figure for the place of employment is slightly smaller at 2.9%. The fact that these effects are similarly sized is not surprising given the way in which functional labour markets are constructed (see Section 2) and the large share of observations for which the region of residence and the region of employment are identical (see Section 4.2). We also calculate the marginal effect of a change in the youth share by one standard deviation, which is reported at the bottom of the table. The estimated effects are -4.1% at the place of residence and -3.7% at the place of employment. In terms of magnitude these changes are comparable to the average return to an additional year of experience during the first four years of a worker's career.

[14] Comparable results are obtained when all variables are averaged across the individuals in a region-year cell and the regression is weighted by the number of observations per cell (see Angrist and Pischke, 2009).

The first-stage coefficients of the instrument are positive and highly significant in both specifications. There is, however, no one-to-one relationship between the current and the lagged value of the youth share, which suggests that within the 15 years over which the instrument is lagged the size of the youth share is affected by natural population changes and migration; instead an increase in the instrument by 1 percentage point is associated with an increase in the current youth-size variable by 0.46 percentage points. The first-stage F-statistics, which measure the significance of the excluded instruments, are considerably larger than the rule-of-thumb value 10 (Staiger and Stock, 1997), and the instrument's explanatory power is further evidenced by the value of Shea's partial R^2. Identification does therefore not appear to be hampered by the presence of weak instruments.[15]

Table 1: Baseline model

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.46 (0.63)*	-3.22 (0.97)***	-0.85 (0.73)	-2.89 (1.22)*
Age	0.21 (0.02)***	0.21 (0.02)***	0.21 (0.02)***	0.21 (0.02)***
Age^2	-0.00 (0.00)***	-0.00 (0.00)***	-0.00 (0.00)***	-0.00 (0.00)***
Experience	0.05 (0.00)***	0.05 (0.00)***	0.05 (0.00)***	0.05 (0.01)***
Experience^2	-0.00 (0.00)***	-0.00 (0.00)***	-0.00 (0.00)***	-0.00 (0.00)***
Education				
Lower secondary (with apprenticeship)	0.22 (0.01)***	0.22 (0.01)***	0.22 (0.01)***	0.22 (0.01)***
Upper secondary (without apprenticeship)	0.05 (0.01)***	0.05 (0.01)**	0.04 (0.02)**	0.04 (0.02)**
Upper secondary (with apprenticeship)	0.30 (0.01)***	0.30 (0.01)***	0.30 (0.01)***	0.30 (0.01)***
Tertiary (University of Applied Sciences)	0.31 (0.02)***	0.31 (0.02)***	0.31 (0.02)***	0.31 (0.02)***
Tertiary (University)	0.46 (0.03)***	0.46 (0.03)***	0.46 (0.03)***	0.46 (0.03)***
Nationality				
Turkey	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Switzerland/Austria	-0.02 (0.04)	-0.01 (0.04)	-0.02 (0.04)	-0.02 (0.04)
Western Europe	-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)
Northern Europe	-0.07 (0.08)	-0.07 (0.08)	-0.07 (0.07)	-0.07 (0.07)
Central Europe	-0.03 (0.02)†	-0.03 (0.02)†	-0.03 (0.02)*	-0.03 (0.02)*
Eastern Europe	-0.08 (0.03)**	-0.08 (0.03)**	-0.08 (0.03)***	-0.08 (0.03)***
South-East Europe	-0.02 (0.01)†	-0.02 (0.01)*	-0.02 (0.01)†	-0.02 (0.01)†
Southern Europe	-0.07 (0.01)***	-0.07 (0.01)***	-0.07 (0.01)***	-0.07 (0.01)***
Africa	-0.11 (0.02)***	-0.11 (0.02)***	-0.11 (0.02)***	-0.11 (0.02)***
Asia	-0.12 (0.02)***	-0.12 (0.02)***	-0.12 (0.02)***	-0.12 (0.02)***
America/Oceania	0.08 (0.05)	0.08 (0.05)	0.08 (0.06)	0.08 (0.06)
Firm size (in 1,000s)	0.20 (0.02)***	0.20 (0.02)***	0.21 (0.03)***	0.21 (0.03)***
Unemployment rate	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
Youth unemployment rate	0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Constant	1.78 (0.28)***	2.06 (0.29)***	1.69 (0.29)***	2.02 (0.32)***
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market region	Yes	Yes	Yes	Yes

15 A number of studies find that the magnitude of cohort-size effects differs across educational groupings (e.g. Brunello, 2010). The results of Table 1 are not affected by excluding either those observations with tertiary education or all observations with either tertiary or upper secondary education.

First-stage regression				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage test statistics				
F-statistic	-	136.60***	-	131.80***
Shea's partial R^2	-	0.32	-	0.32
Observations				
Individuals	107,351	107,352	107,351	107,351
Labour-market region-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24
ME (stdev)	-1.84%*	-4.05%***	-1.09%	-3.67%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

The OLS point estimates, while still negative, are considerably smaller than their 2SLS counterparts and their values lie outside the formers' 90% confidence interval. This finding is in line with the discussion of Section 4.3: if the value of the youth share is influenced by individuals migrating into economically attractive regions, OLS estimation can be expected to produce coefficients that are on average less negative than the true value of the youth share's marginal effect. The coefficients of the control variables display a large degree of similarity across the four different specifications of Table 1. Wages are predicted to increase at a decreasing rate in age and experience – the latter being suggestive of the widely documented concave experience-earnings profile (Polachek, 2008).[16] Higher levels of schooling and professional qualification are associated with higher earnings.

Nationals from Eastern European, South-East and Southern European countries are predicted to earn significantly less than Germans, while the largest difference is found for Africans and Asians with earnings lower by more than 10%. Individuals who are employed in firms with larger workforces are found to have higher earnings, which is in line with evidence by Lehmer and Möller (2010). Finally, the effects of the unemployment rate and the share of young unemployed individuals are small. The youth-unemployment variable draws a negative coefficient in the 2SLS estimations, but in contrast to findings by Baltagi and Blien (1998) its effect is not statistically significant.

Related studies have used administrative units at the sub-national level as the basis for constructing population variables. As discussed in Section 2, the drawback of such an approach is that these units do not necessarily represent actual labour markets and that, consequently,

[16] We have also estimated Equation 1 using mutually exclusive sets of age and experience dummies. Changing the specification in this way has no effect on the estimated youth-share coefficients.

the size of age groups within a given labour market is potentially measured with error (see the Supplementary Material for a discussion). We assess the effect of using administrative rather than functional units by estimating Equation 1 on the basis of a district-specific youth-share variable. The results are shown in Table 2.

The 2SLS coefficients of the youth-share variables remain negative and larger in absolute value than the corresponding OLS estimates, but only the specification referring to an individual's place of residence produces statistically significant results. However, compared to the results of Table 1, using districts rather than labour-market regions leads to an underestimation of the youth share's negative effect: the point estimates referring to the place of residence are smaller by 13%, while the size of the coefficient for the place of employment drops by almost 50%.[17] The increased discrepancy between the youth-share coefficients of these specifications reflects the fact that individuals are more likely to live and work in different districts than is the case for labour-market regions.

Table 2: District-based youth share variable

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.31 (0.45)***	-2.79 (0.81)***	-0.12 (0.42)	-1.50 (0.92)
Dummies				
Year	Yes	Yes	Yes	Yes
District	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage regression				
Instrument	-	0.44 (0.00)***	-	0.43 (0.00)***
First-stage test statistics				
F-statistic	-	300.92***	-	181.95***
Shea's partial R^2	-	0.27	-	0.22
Observations				
Individuals	107,351	107,351	107,351	107,351
District-year cells	3,756	3,756	3,756	3,756
Districts (clusters)	313	313	313	313
R^2	0.25	0.25	0.26	0.26
ME(stdev)	-1.80%***	-3.84%***	-0.17%	-2.18%

Cluster-robust standard errors in parentheses (clustered at the district level). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

It can be shown that the negative and significant youth-share coefficients of Table 2 are driven by those districts in which individuals in the sample are more likely to live than to work. At the same time, the average absolute difference between the district-based youth-share variable and its value at the corresponding labour-market region is smaller for these districts, which

[17] Due to the higher variance of the district-based youth-share variable the proportional changes in the marginal effects for a change of one standard deviation are less pronounced.

suggests that measurement error in the size of the youth-share variable is less pronounced. The fact that these districts account for a larger fraction of observations in the place-of-residence specification suggests that size and significance of the youth-share coefficients will be less affected in that specification. It turns out that individuals in the sample are more likely to work in cities and to live in rural areas. The rationale behind the above argument could therefore be that the youth share within a city-district provides only an inaccurate measure of the size of the youth population that is relevant for the determination of wage outcomes as cities will also draw workers from surrounding districts.

As discussed in Section 1, the size of an individual's age group could have an effect on the conditions of his employment. Specifically, if young workers in larger age groups are more likely to be in positions in lower-paying occupations or industries, the estimated wage effect of the youth share in Table 1 would be confounded by these types of selection effects. In particular, the negative effect would be overestimated. To address this issue, we successively add indicator variables to the model of Equation 1 which are derived from two-digit codes referring to an individual's industry and occupation. The results are shown in Table 3 for the place-of-residence specification and in Table 4 for the place of employment.

Table 3: Industry and occupation indicators (place of residence)

Dependent variable: log real daily earnings	Baseline	+industry	+occupation	+industry +occupation
Youth share (2SLS)	-3.22 (0.97)***	-2.81 (0.92)***	-1.88 (0.91)*	-1.89 (0.86)*
Youth share (OLS)	-1.46 (0.63)*	-1.36 (0.57)*	-0.90 (0.60)	-0.97 (0.53)†
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market region	Yes	Yes	Yes	Yes
Industry	No	Yes	No	Yes
Occupation	No	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage regression				
Instrument	0.46 (0.00)***	0.46 (0.00)***	0.46 (0.00)***	0.46 (0.00)***
First-stage test statistics				
F-statistic	136.60***	137.17***	137.32***	137.70***
Shea's partial R^2	0.32	0.32	0.32	0.32
Observations				
Individuals	107,351	107,351	107,351	107,351
Labour-market region-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2 (2SLS)	0.24	0.46	0.40	0.51
R^2 (OLS)	0.24	0.46	0.40	0.51
ME(stdev, 2SLS)	-4.05%***	-3.54%***	-2.36%*	-2.39%*
ME(stdev, OLS)	-1.84%*	-1.71%*	-1.14%	-1.22%†

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

In both cases we find that adding industry and, especially, occupation indicators has a sizeable impact on the estimated youth-share effects compared to the baseline specification: the inclusion of industry indicators decreases the size of the 2SLS coefficients by 12% (place of residence) and 4% (place of employment) compared to the results of the baseline model, while the reduction resulting from adding occupation indicators is considerably larger at 40% and 30%, respectively. Similar results are obtained when both sets of indicator variables are used. Moreover, it can be seen that when dummies for industry or occupation are added the difference in the size of the 2SLS coefficients between the place of residence and the place of employment decrease in magnitude. This supports the argument that once labour-market regions are used as the spatial entities from which the youth-share variable is constructed both types of places produce similar results.

Table 4: Industry and occupation indicators (place of employment)

Dependent variable: log real daily earnings	Baseline	+industry	+occupation	+industry +occupation
Youth share (2SLS)	-2.89 (1.22)*	-2.77 (1.06)**	-1.96 (1.08)†	-2.11 (1.01)*
Youth share (OLS)	-0.85 (0.73)	-0.89 (0.62)	-0.52 (0.66)	-0.62 (0.57)
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market region	Yes	Yes	Yes	Yes
Industry	No	Yes	No	Yes
Occupation	No	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage regression				
Instrument	0.46 (0.00)***	0.46 (0.00)***	0.46 (0.00)***	0.46 (0.00)***
First-stage test statistics				
F-statistic	131.80***	132.31***	132.33***	132.67***
Shea's partial R^2	0.32	0.32	0.32	0.32
Observations				
Individuals	107,351	107,351	107,351	107,351
Labour-market region-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2 (2SLS)	0.24	0.46	0.41	0.51
R^2 (OLS)	0.24	0.46	0.41	0.51
ME(stdev, 2SLS)	-3.67%*	-3.52%**	-2.49%†	-2.68%*
ME(stdev, OLS)	-1.09%	-1.13%	-0.66%	-0.79%

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

While in the baseline model an increase in the size of the youth-share variable by one percentage point was predicted to decrease an individual's wages by about 3%, ceteris paribus, the size of this effect is reduced once an individual's industrial and, in particular, occupational affiliation are controlled for. This finding suggests that the estimated youth-share coefficients of the baseline specification were indeed confounded by the positive association between young workers being in larger age groups and being employed in lower-paying industries and occupations. We conclude that in addition to the direct negative effect of the size of the youth share, there is an indirect effect driven by selection into specific industries and occupations. A possible explanation for this finding is that, ceteris paribus, a larger supply of young individuals increases competition for higher-quality jobs, forcing some individuals to take up employment in lower-paying occupations.

This interpretation is in line with recent results pertaining to the wage effects of labour market conditions. Kahn (2010) and Brunner and Kuhn (2014) find that adverse labour market conditions (measured by the unemployment rate at the time of labour-market entry) depress wages and increase the probability of employment in lower-quality occupations. Morin (2015) studies the wage effects of the increase in labour supply due to the double cohort of high-school graduates in Ontario and provides evidence that part of the negative wage effect is due to selection into lower-paying occupations. Alternatively, higher-quality jobs may require a specific type of qualification. If the supply of training positions does not adjust to the supply of young individuals, the number of individuals barred from entering higher-paying occupations will increase in larger age groups. The effect of age-group size on selection into industries and occupations certainly warrants further research.

The Supplementary Material contains the corresponding output tables for the case in which districts provide the basis for the construction of the youth-share variable. These show that region-specific and district-specific variables continue to produce different results once industry and occupation dummies have been added and also illustrate that the difference between the results of the place-of-residence and the place-of-employment specifications are more pronounced at the district level.

To assess to what extent the results of Table 1 merely reflect unobserved heterogeneity at the federal state-year level, we add dummy variables for the interaction between federal states and years to the model of Equation 1. Doing so allows us to control for annual shocks that affect states differently and that are relevant for the determination of individual wages, e.g. the effects of macroeconomic shocks may vary between states due to differences in industrial structure. The results displayed in Table 5 suggest that, at least for the place-of-residence specification, the estimated effects of the youth-share variable in the baseline specification are not driven by unobserved heterogeneity at the state-year level.

Table 5: State-by-year interactions

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.01 (0.63)	-2.58 (1.27)*	-0.46 (0.74)	-2.35 (1.56)
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market region	Yes	Yes	Yes	Yes
Federal state-by-year	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage regression				
Instrument	-	0.43 (0.00)***	-	0.43 (0.00)***
First-stage test statistics				
F-statistic	-	85.39***	-	84.74***
Shea's partial R^2	-	0.26	-	0.26
Observations				
Individuals	107,351	107,351	107,351	107,351
Labour-market region-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24
ME(stdev)	-1.28%	-3.25%*	-0.58%	-2.99%

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

The 2SLS point estimates fall by approximately 20% as part of the explained variation in the earnings variable is now picked up by the additional dummies. The standard errors increase presumably because parts of the variation in the youth-share variable are now explained by the additional dummy variables, which results in less precise estimates. For a similar reason there is a drop in the values of the first-stage F-statistic and the partial R^2 of the excluded instrument: the explanatory power of the instrument is reduced as a consequence of including the state-by-year dummies in the first-stage equation.

6 Conclusion

This paper empirically analyses how changes in the size of the youth population affect the wages of young workers. Under the assumption that differently aged individuals are only imperfectly substitutable because of differences in firm-specific, occupation-specific, industry-specific or general human capital, economic theory predicts that an increase in the size of an age group reduces the earnings of the members of that group. This hypothesis is tested using a sample of young male employees from Western Germany. The demographic forces that are currently changing the age-structure of the German population illustrate the relevance of this analysis. Specifically, the share of young individuals is projected to fall by 2.5 percentage points (equivalently, by 15%) at the national level over the period 2010-2025 following a period of an increasing youth share.

Besides providing an analysis of this relationship using recent data from administrative records, this paper makes two additional contributions. First, functional labour-market regions rather than administrative units are used as the spatial entities within which the size of the youth population is measured. These units provide a better measure of the number of young individuals in an actual labour market than administrative units, which are usually not delineated according to economic criteria, and hence of the supply of young labour that is relevant for the determination of young workers' wages. Use of a youth-share variable based on labour-market regions therefore reduces the potential for measurement error in this variable. Second, we address the channels through which an increase in the supply of young individuals affects their wages by controlling for industrial and occupational upgrading, i.e. for the possibility that changes in the size of the youth population affect the chances of finding employment in higher-paying industries or occupations.

The empirical analysis employs an IV approach in order to account for the possibility that the youth share is endogenous due to young individuals migrating into high-wage areas. In line with the hypothesis that increases in age-group size reduce the wages of the members of that group, the 2SLS coefficients are negative and significant: an increase in the youth share by one percentage point is predicted to decrease young workers' wages by 3%. Consistent with the argument that migration into high-wage regions induces endogeneity, the corresponding OLS estimates are less negative. Estimating our model using a youth-share variable that is based on districts rather than labour-market regions reduces the size of the 2SLS coefficients by either 13% (place of residence) or 48% (place of employment), which is consistent with the hypothesis that the use of administrative units induces measurement error in the youth-share variable. Finally, adding indicators for an individual's occupation and industry reduces the size of the youth-share coefficients from -3% to -2%. We interpret this result as providing evidence for the hypothesis that belonging to a larger age group increases the likelihood of being employed in lower-paying occupations or industries.

What are the implications of these findings for the wages of the coming cohorts of young workers in light of Western Germany's changing demographics? As the youth share is projected to decrease over the coming years, demographic processes appear to be favourable to the development of the wages that young workers can expect in the future. But as the development of population structures is likely to differ between regions, regional variation in the extent to which young workers stand to benefit is to be expected. Finally, it should be borne in mind that these results come from a specific sample consisting of young, male, full-time employees with a few years of work experience and, predominantly, lower secondary education. Whether the relationship between the youth share and young workers' wages is similar for other groups, such as females or the highly educated, remains a topic for future research.

Acknowledgements

The authors would like to thank Stefan Fuchs, Bernd Hayo, John Moffat and Norbert Schanne for their advice and are grateful for comments from the participants of IAB's regional research network meeting in Aalen, the 54th Conference of the European Regional Science Association (ERSA), the 5th ifo Workshop Arbeitsmarkt und Sozialpolitik, the joint IAB Regional Science Academy workshop in Amsterdam and the 28th Conference of the European Association of Labour Economists (EALE). The Federal Institute for Research on Building, Urban Affairs and Spatial Development kindly provided the shapefile of the German labour-market regions. Our thanks also go to Annie Roth for proof-reading an earlier version of this paper.

References

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

Arnds, P. and Bonin, H. (2002) Arbeitsmarkteffekte und finanzpolitische Folgen der demographischen Alterung in Deutschland, IZA Discussion Paper No. 667, Institute of Labor Economics, Bonn.

Baltagi, B.H. and Blien, U. (1998) The German wage curve: evidence from the IAB employment sample, *Economics Letters*, **61**, 135-42.

Berger, M.C. (1985) The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence, *Journal of Political Economy*, **93**, 561-73.

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

Bloom, D.E. and Sousa-Poza, A. (2013) Aging and productivity: Introduction, *Labour Economics*, **22**, 1-4.

Blundell, R. and MaCurdy, T. (1999) Labour Supply: A Review of Alternative Approaches, in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics, Vol. 3*., North-Holland, Amsterdam.

Börsch-Supan, A. (2013) Myths, Scientific Evidence and Economic Policy in an Aging World, *Journal of the Economics of Ageing,* **1-2**, 3-15.

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

Brunner, B. and Kuhn, A. (2014) The impact of labor market entry conditions on initial job assignments and wages, *Journal of Population Economics,* **27**, 705-38.

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.

Connelly, R. (1986) A Framework for Analyzing the Impact of Cohort Size on Education and Labor Earning, *Journal of Human Resources,* **21**, 543-62.

Eberle, J., Schmucker, A. and Seth, S. (2013) Programmierbeispiele zur Datenaufbereitung der Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) in Stata. Generierung von Querschnittsdaten und biografischen Daten, FDZ Methodenreport No. 04/2013, Institute for Employment Research, Nuremberg.

Eckey, H.-F., Kosfeld, R. and Türck, M. (2006) Abgrenzung deutscher Arbeitsmarktregionen, *Raumforschung und Raumordnung*, **64**, 299-309.

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

Fitzenberger, B. (1999) *Wages and employment across skill groups: An analysis for West* Germany, Physica-Verlag: Heidelberg.

Fitzenberger, B. and Kohn, K. (2006) Skill Wage Premia, Employment, and Cohort Effects: Are Workers in Germany All of the Same Type?, IZA Discussion Paper No. 2185, Institute of Labor Economics, Bonn.

Freeman, R.B. (1979) The effect of demographic factors on age-earnings profiles, *Journal of Human Resources*, **14**, 289-318.

Fuchs, M. and Weyh, A. (2014) Demography and unemployment in East Germany. How close are the ties?, IAB Discussion Paper No. 26/2014, Institute for Employment Research, Nuremberg.

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

Gartner, H. (2005) The imputation of wages above the contribution limit with the German IAB employment sample, FDZ Methodenreport No. 02/2005, Institute for Employment Research, Nuremberg.

Gertler, M. and Trigari, A. (2009) Unemployment Fluctuations with Staggered Nash Wage Bargaining, *Journal of Political Economy,* **117**, 38-86.

Gürtzgen, N. (2016) Estimating the Wage Premium of Collective Wage Contracts - Evidence from Longitudinal Linked Employer-Employee Data*, Industrial Relations,* **55**, 294-322.

Jimeno, J.F., Rojas, J.A. and Puente, S. (2008) Modelling the impact of aging on social security expenditures, *Economic Modelling*, **25**, 201-24.

Kahn, L.B. (2010) The long-term labor market consequences of graduating from college in a bad economy, *Labour Economics,* **17**, 303-16.

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.

Lehmer, F. and Möller, J. (2010) Interrelations between the urban wage premium and firm-size wage differentials: a microdata cohort analysis for Germany, *The Annals of Regional Science*, **45**, 31-53.

Mazzonna, F. and Peracchi, F. (2012) Ageing, cognitive abilities and retirement, *European Economic Review*, **56**, 691-710.

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

Mincer, J. (1958) Investment in Human Capital and the Personal Income Distribution, *Journal of Political Economy,* **66**, 281-302.

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

Moffat, J. and Roth, D. (2016b) Cohort size and youth labour-market outcomes: the role of measurement error, IAB Discussion Paper No. 37/2016, Institute for Employment Research, Nuremberg.

Morin, L.-P. (2015) Cohort size and youth earnings: Evidence from a quasi-experiment, *Labour Economics,* **32**, 99-111.

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.

Ochsen, C. (2009) Regional labor markets and aging in Germany, Thünen-Series of Applied Economic Theory Working Paper No. 102, Department of Economics, Faculty of Economics and Social Sciences, University of Rostock, Rostock.

Ottaviano, G. and Peri, G. (2012) Rethinking the effect of immigration on wages, *Journal of the European Economic Association,* **10**, 152-97.

Polachek, S.W. (2008) Earnings Over the Life Cycle: The Mincer Earnings Function and Its Applications, *Foundations and Trends in Microeconomics*, **4**, 165-272.

Sapozhnikov, M. and Triest, R.K. (2007) Population Aging, Labor Demand, and the Structure of Wages, Research Department Working Paper 07-8, Federal Reserve Bank of Boston, Boston.

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.

Stapleton, D.C. and Young, D.J. (1988) Educational Attainment and Cohort Size, *Journal of Labor Economics,* **6**, 330-61.

Statistisches Bundesamt (2009) *Bevölkerung Deutschlands bis 2060: 12. koordinierte Bevölkerungsvorausberechnung*, Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt (2010) *Bevölkerung und Erwerbstätigkeit: Bevölkerung in den Bundesländern, dem früheren Bundesgebiet und den neuen Ländern bis 2060: Ergebnisse der 12. koordinierten Bevölkerungsvorausberechnung*, Statistisches Bundesamt, Wiesbaden.

vom Berge, P., König, M. and Seth, S. (2013) Sample of Integrated Labour Market Biographies (SIAB) 1975 – 2010, FDZ Datenreport No. 01/2013, Institute for Employment Research, Nuremberg.

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

Table A1: Descriptive statistics

Variable	Mean	Standard deviation	Minimum	Maximum
Log daily earnings	4.28	0.31	3.18	6.21
Youth share				
Labour market region				
Population-based (place of residence)	0.17	0.01	0.14	0.21
Population-based (place of employment)	0.17	0.01	0.14	0.21
District				
Population-based (place of residence)	0.17	0.01	0.13	0.23
Population-based (place of employment)	0.17	0.01	0.13	0.23
Instrument				
Labour market region				
Population-based (place of residence)	0.16	0.02	0.12	0.20
Population-based (place of employment)	0.16	0.02	0.12	0.20
District				
Population-based (place of residence)	0.16	0.02	0.10	0.21
Population-based (place of employment)	0.16	0.02	0.10	0.21
Age	22.33	1.46	15	24
15	0.00	0.01	0	1
16	0.00	0.01	0	1
17	0.00	0.03	0	1
18	0.00	0.07	0	1
19	0.03	0.17	0	1
20	0.09	0.29	0	1
21	0.16	0.37	0	1
22	0.20	0.40	0	1
23	0.24	0.43	0	1
24	0.28	0.45	0	1
Experience	2.04	1.48	0	10
0	0.14	0.35	0	1
1	0.27	0.44	0	1
2	0.24	0.43	0	1
3	0.18	0.38	0	1
4	0.11	0.31	0	1
5	0.04	0.20	0	1
6	0.01	0.11	0	1
7	0.00	0.06	0	1
8	0.00	0.04	0	1
9	0.00	0.02	0	1
10	0.00	0.01	0	1
Education				
Lower secondary (without apprenticeship)*	0.19	0.39	0	1
Lower secondary (with apprenticeship)	0.76	0.42	0	1
Upper secondary (without apprenticeship)	0.01	0.12	0	1
Upper secondary (with apprenticeship)	0.03	0.17	0	1
Tertiary (University of Applied Sciences)	0.01	0.08	0	1
Tertiary (University)	0.00	0.05	0	1

Nationality				
Germany*	0.90	0.30	0	1
Turkey	0.04	0.20	0	1
Switzerland/Austria	0.00	0.03	0	1
Western Europe	0.00	0.04	0	1
Northern Europe	0.00	0.02	0	1
Central Europe	0.01	0.09	0	1
Eastern Europe	0.00	0.04	0	1
South-East Europe	0.02	0.14	0	1
Southern Europe	0.02	0.13	0	1
Africa	0.00	0.05	0	1
Asia	0.01	0.08	0	1
America/Oceania	0.00	0.02	0	1
Firm size	970.47	3,897.56	1	42,626
Unemployment rate				
Labour-market region (place of residence)	8.16	2.53	2.60	18.04
Labour-market region (place of employment)	8.13	2.51	2.60	18.04
District (place of residence)	8.01	2.94	1.90	25.59
District (place of employment)	8.28	3.01	1.90	25.59
Youth unemployment share				
Labour-market region (place of residence)	7.26	2.65	1.90	24.38
Labour-market region (place of employment)	7.23	2.63	1.90	24.38
District (place of residence)	7.20	2.95	1.70	24.92
District (place of employment)	7.36	2.99	1.70	24.92
Observations		107,351		

* Base category in the regression analysis

Supplementary Material

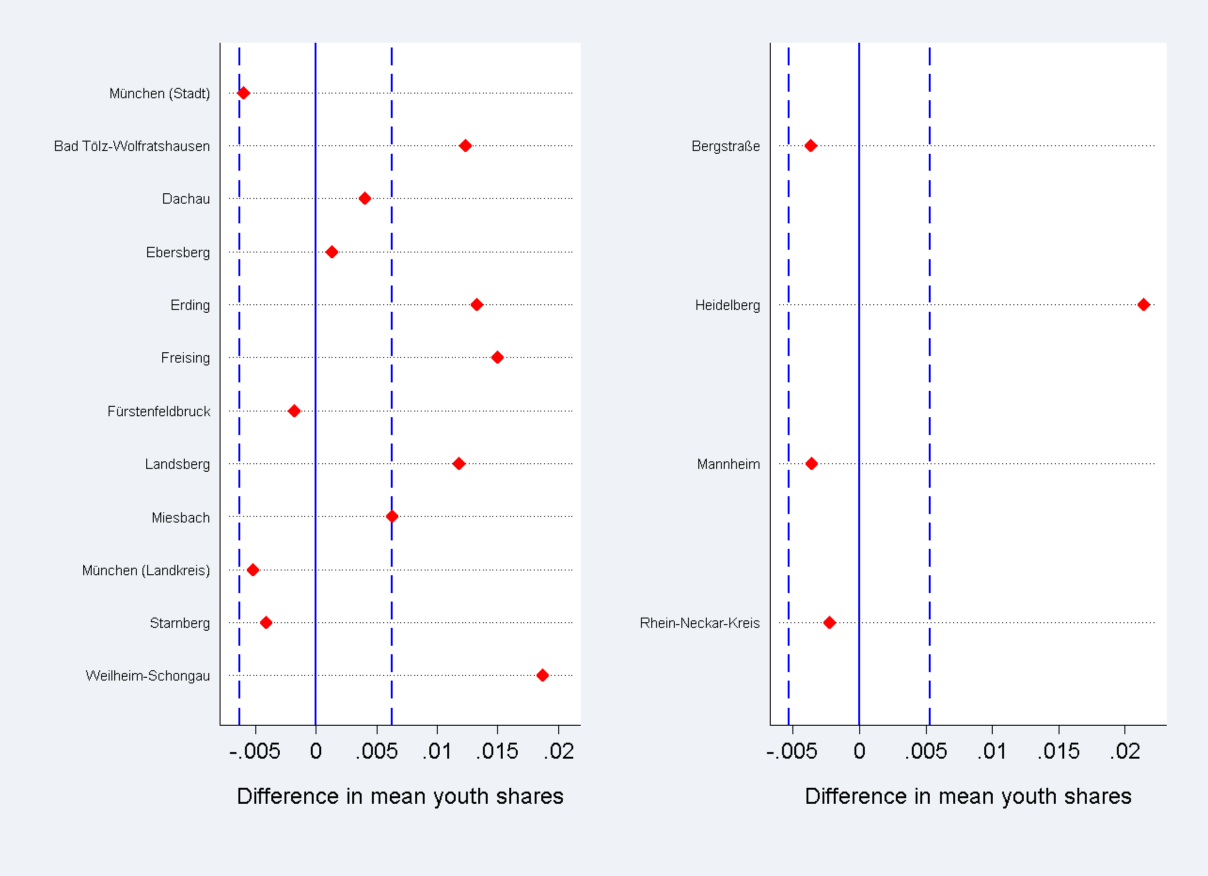
The supplementary material serves two purposes: first, it provides a discussion of the implications for the estimated coefficients when the youth share is constructed from districts rather than from labour-market regions. Since the former are administrative units, which typically do not correspond with local labour market, a district-based youth-share variable is potentially subject to measurement error in that it groups together individuals that are not active within the same labour market. It is, moreover, argued, that the identification strategy employed in the paper may not lead to a consistent estimate of the youth-share coefficient when the former variable is district-specific. The second part addresses the robustness of the results of the empirical analysis and presents various sensitivity analyses.

S1 Measurement error

The paper uses the functional labour-market regions defined by Eckey et al. (2006) as the spatial units from which the empirical model's main variable, the share of the population aged 15-24 relative to the population aged 15-64, is constructed. As these entities are designed to approximate regional labour markets, their use provides a measure of the potential supply of young workers within an actual labour market. In contrast, administrative units, such as districts, are not delineated accordingly and therefore only provide an incorrect measure of the size of the youth share in the corresponding labour market. The aim of this section is to provide evidence for the existence of measurement error in a district-based youth-share variable and to discuss the implications for the estimation of a model that uses district-specific variables.

Figure S1 illustrates the potential for measurement error in a district-based youth-share variable through use of two exemplary labour-market regions. The labour-market region of Munich (left panel) consists of twelve districts and combines the city of Munich with the surrounding periphery, whereas the region Mannheim-Heidelberg (right panel) is an example of two cities sharing a joint labour market. The graphs show the difference between the average value of the youth-share variable at the level of the labour-market region and at the level of the districts it contains (values are averaged over the period 1999-2010). For some districts this difference can be substantial, exceeding the standard deviation of the average youth share at the level of the labour-market region (as indicated by the dashed lines). In these cases a district-based youth-share variable would appear to provide only an inaccurate measure of the size of the youth population within the corresponding labour-market region.

Figure S1: Difference between the district- and labour-market-based mean youth share



Source: Sample of Integrated Employment Biographies (authors' calculations). Dashed lines indicate the value of plus/minus one standard deviation in the mean value of the youth-share variable when measured at the level of the corresponding labour-market region (averaged over the period 1999-2010).

Under the assumption that the labour-market regions of Eckey et al. (2006) indeed represent the labour markets that individuals are active on, the relationship between an individual's (log) wage and the size of the youth share in his labour-market region can be specified according to the following model (which corresponds to Equation 1 in the paper):

$$\log(w_{irt}) = \alpha + \beta y_{rt} + \boldsymbol{x_{irt}}'\boldsymbol{\gamma} + \delta_r + \mu_t + \varepsilon_{irt} \qquad [S1]$$

We propose that a district-specific version of the youth-share variable, y_{rt}^{dis}, provides an incorrect measure of the youth share within the labour market that an individual belongs to which we specify in form of a district-level dummy variable, ψ_{rt}^{dis}, and an additive random measurement error ξ_{rt}^{dis}. The variable ψ_{rt}^{dis} allows for the possibility of the youth share in a specific district being permanently smaller or larger than its value in the corresponding labour market region:

$$y_{rt}^{dis} = y_{rt} + \psi_r^{dis} + \xi_{rt}^{dis} \quad [S2]$$

Substituting Equation S2 into the model of Equation S1 shows that a district-specific youth-share variable is correlated with the composite error term ε_{rt}^{dis}, which contains the measurement error component ξ_{rt}^{dis}:

$$\log(w_{irt}) = \alpha + \beta y_{rt}^{dis} + \boldsymbol{x_{irt}}'\boldsymbol{\gamma} + \eta_r^{dis} + \mu_t + \varepsilon_{irt}^{dis} \quad [S3]$$

$$\eta_r^{dis} = \delta_r - \beta \psi_r^{dis} \quad [S4]$$

$$\varepsilon_{irt}^{dis} = \varepsilon_{irt} - \beta \xi_{rt}^{dis} \quad [S5]$$

As labour-market regions are comprised of one or more districts, it can be shown that the youth share of a given labour-market region *k* is equal to the weighted sum of the youth shares in the districts *l* (l = 1,…, L) that are contained in region *k*, where the weights are given by the fraction of the population aged 15-64 in region *k*, $N_{15\text{-}64,kt}$, that can be ascribed to district *l*, $N_{15\text{-}64,lt}^{dis}$ (consequently, the weights add up to unity):

$$y_{kt} = \sum_{l=1}^{L} \omega_{lt}^{dis} y_{lt}^{dis} \quad [S6]$$

$$\omega_{lt}^{dis} = \frac{N_{15-64,lt}^{dis}}{N_{15-64,kt}} \quad [S7]$$

The relationship between the region-based and the district-based youth-share variables shown in Equation S2 implies that the measurement errors of those districts within a given labour-market region are linearly dependent:

$$\xi_{1t}^{dis} = -\psi_1^{dis} - \sum_{l=2}^{L} \frac{\omega_{lt}^{dis}}{\omega_{1t}^{dis}} \psi_l^{dis} - \sum_{l=2}^{L} \frac{\omega_{lt}^{dis}}{\omega_{1t}^{dis}} \xi_{lt}^{dis} \quad [S8]$$

Consequently, use of a district-based youth-share variable not only induces endogeneity due to measurement error, but also leads to the error terms of observations from different districts being correlated if they belong to the same labour-market region.

In principle, IV estimation can be used to obtain consistent estimates in the presence of measurement error (Hausman, 2001) if the instrument is uncorrelated with the composite error term of Equation S5. In this paper the instrument is defined as the ratio of the number of individuals up to the age of 9 and the number of individuals up to the age of 49 observed 15 years earlier. As in the case of the youth-share variable the instrument can be constructed from either districts or labour-market regions and, analogously to Equation S2, it is possible to interpret the district-based version of the instrument as an incorrect measure of the regional variable:

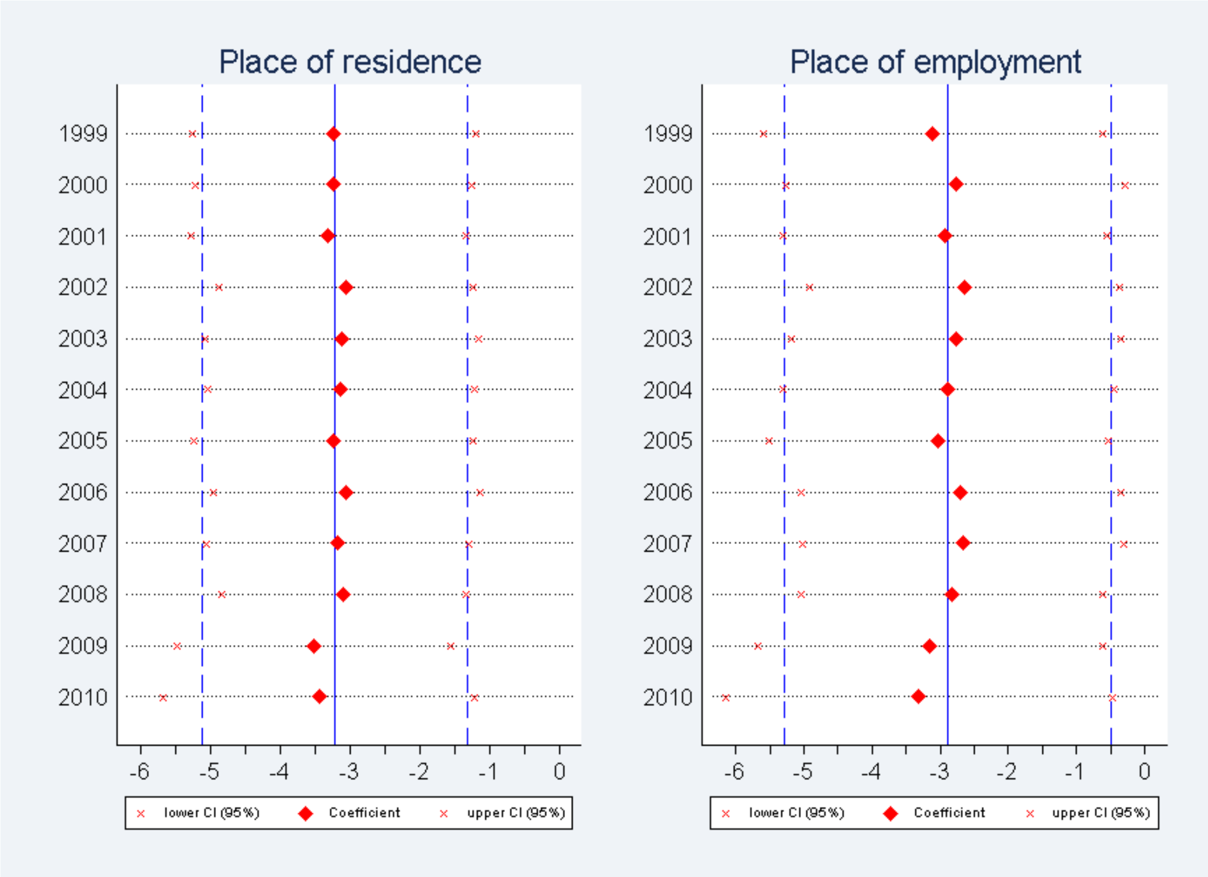
$$z_{r,t-15}^{dis} = z_{r,t-15} + \phi_r^{dis} + \nu_{r,t-15}^{dis} \qquad \text{[S9]}$$

Consistent estimation of a model with a district-based youth-share variable in combination with a district-specific instrument requires that the current and the lagged measurement errors are uncorrelated. This condition would not be satisfied if the extent of measurement error exhibited persistence over time, e.g. if a large difference between the district-based and the region-based instrument was associated with a large difference in the district-specific and region-specific youth-share variable. Under these circumstances, application of the identification strategy of the paper to a model with a district-based youth-share variable would not yield a consistent estimate of the former's effect on an individual's wages.

S2 Sensitivity analysis

In this section we perform various sensitivity analyses in order to assess the robustness of the paper's findings. First, we show that the 2SLS coefficients of the baseline model are not driven by individual regions or years. Figure S2 presents the youth-share coefficients that are obtained when observations from a single year are excluded from the sample: regardless whether an individual's region of residence or region of employment is used, the estimated coefficients are always very close to those of the full model and always lie within the formers' 95% confidence interval.

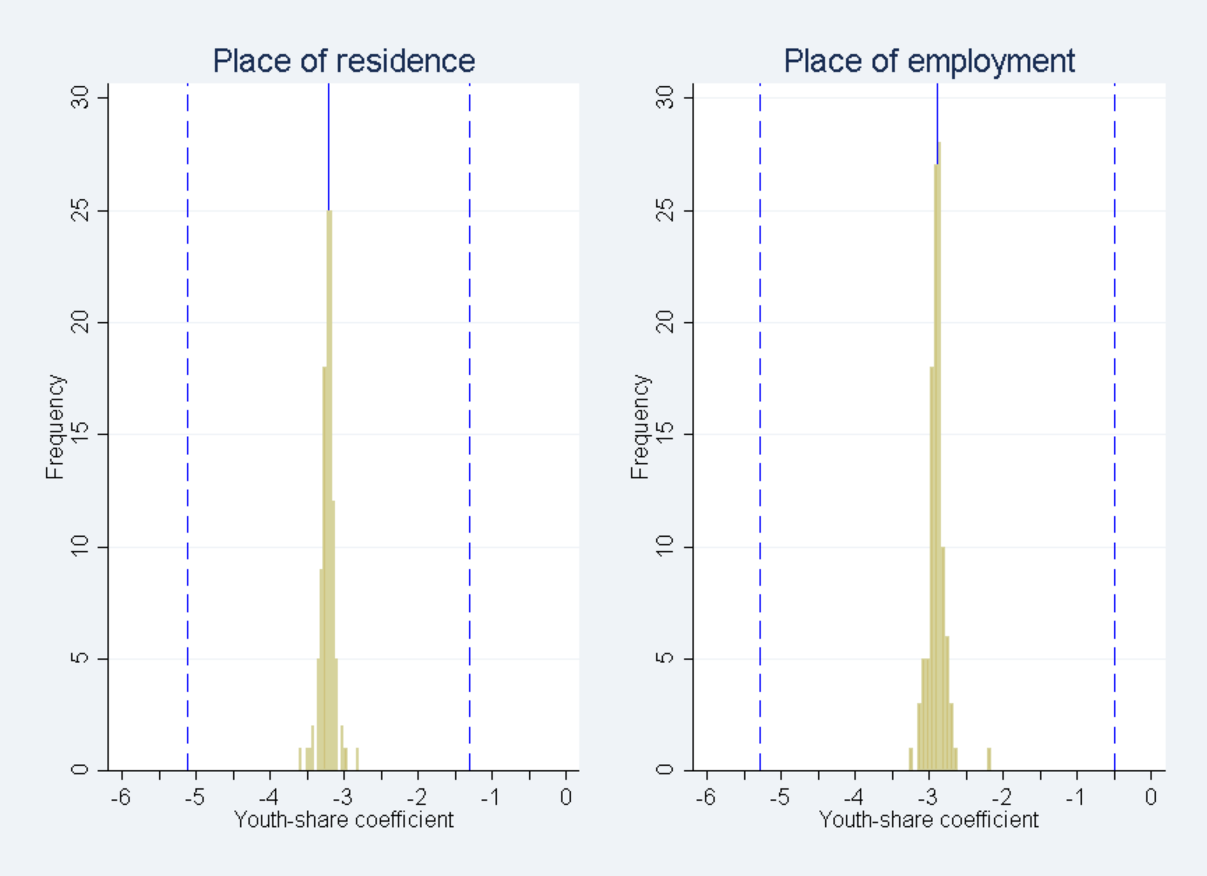
Figure S2: 2SLS baseline coefficients after excluding individual years



Source: Sample of Integrated Employment Biographies (authors' calculations). The youth-share coefficients are derived from the baseline model; the blue solid line represents the youth-share coefficients from the full model, the blue dashed lines the corresponding 95% confidence interval.

Due to the relatively large number of regions (108), illustrating the effects of dropping individual regions would be unwieldy. Instead of showing individual coefficients, their histogram is depicted in Figure S3. As can be seen, the distribution of the coefficients is centred on the coefficient of the full model and displays a spread which is small compared to the 95% confidence interval.

Figure S3: Histogram of 2SLS baseline coefficients after excluding individual regions



Source: Sample of Integrated Employment Biographies (authors' calculations). The youth-share coefficients are derived from the baseline model; the blue solid line represents the youth-share coefficients from the full model, the blue dashed lines the corresponding 95% confidence interval.

The following tables contain the coefficient of the youth-share variable as well as the former's marginal effect for a change of one standard deviation for a number of sensitivity analyses in which either the sample (Tables S1-S10) or the empirical specification are modified. In Tables S11 and S12 the question is addressed why a change from a region-specific to a district-specific youth-share variable leads to a larger decrease in the size of the coefficient when the variable is measured at an individual's place of employment. Finally, Tables S12 and S13 report the youth-share coefficients from a district-specific variable when indicators for an individual's industry and/or occupation are added.

As discussed in the paper, the analysis is restricted to those individuals who are subject to social security contributions. Tables S1 and S2 show the results from further homogenising the sample by either excluding those individuals with more than a full-time job (Table S1) or by dropping individuals whose employment spells contain less than 90 days (Table S2). In the first case the marginal effects are very close to those of the paper's baseline specification, while they become slightly smaller in the second case.

Table S1: Exclusion of observations with more than a full-time job

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.44 (0.68)*	-3.29 (1.01)***	-0.81 (0.77)	-2.90 (1.22)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage statistics				
F-statistic	-	135.19***	-	131.05***
Shea's partial R^2	-	0.32	-	0.32
Observations				
Individuals	102,387	102,387	102,387	102,387
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24
ME(stdev)	-1.81%*	-4.14%***	-1.03%	-3.68%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

Table S2: Exclusion of observations with employment spells of less than 90 days

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.21 (0.62)†	-2.75 (0.97)***	-0.69 (0.71)	-2.53 (1.21)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage statistics				
F-statistic	-	137.34***	-	132.74***
Shea's partial R^2	-	0.32	-	0.32
Observations				
Individuals	103,652	103,652	103,652	103,652
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.23	0.23	0.23	0.23
ME(stdev)	-1.52%*	-3.47%***	-0.88%	-3.22%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

Since the effect of the size of the youth share may vary between differently educated individuals, the sample is homogenised by first excluding those with tertiary education (Table S3) and then those with tertiary or upper secondary education (Table S4). In neither case is the size of the marginal effects substantially changed.

Table S3: Exclusion of observations with tertiary education

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.47 (0.62)*	-3.22 (0.96)***	-0.91 (0.73)	-2.97 (1.22)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage statistics				
F-statistic	-	137.01***	-	132.41***
Shea's partial R^2	-	0.32	-	0.32
Observations				
Individuals	106,422	106,422	106,422	106,422
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24
ME(stdev)	-1.86%*	-4.05%***	-1.16%	-3.77%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

Table S4: Exclusion of observations with tertiary or upper secondary education

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.39 (0.64)*	-3.29 (0.95)***	-0.84 (0.72)	-3.04 (1.21)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage statistics				
F-statistic	-	140.36***	-	135.72***
Shea's partial R^2	-	0.33	-	0.32
Observations				
Individuals	101,820	101,820	101,820	101,820
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24
ME(stdev)	-1.75%*	-4.14%***	-1.06%	-3.85%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

The SIAB dataset contains observations with unrealistically low average daily wages. In order to get a handle on this issue, observations with daily wages below twice the value of the minor-employment threshold were excluded from the sample. Table S5 shows that when these observations are included, the size of the coefficients decrease in size and they become less significant.

Table S5: No truncation of the wage distribution

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-0.99 (0.78)	-2.51 (1.09)*	-0.33 (0.85)	-2.14 (1.26)†
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage statistics				
F-statistic	-	136.17***	-	131.61***
Shea's partial R^2	-	0.32	-	0.32
Observations				
Individuals	110,651	110,651	110,651	110,651
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24
ME(stdev)	-1.25%	-3.16%*	-0.41%	-2.72%†

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

The paper's empirical analysis is restricted to those individuals who are employed on 30 June of a given year and while June values are usually representative of average annual employment levels, the selection of a specific date is essentially arbitrary. Table S6 shows the results when the reference date is set to 31 December. Doing so produces comparable results in terms of sign and significance but the size of the marginal effects increases in magnitude.

Table S6: Alternative reference date (31 December)

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.36 (0.65)*	-3.80 (0.96)***	-0.75 (0.74)	-3.26 (1.16)***
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage statistics				
F-statistic	-	135.93***	-	130.83***
Shea's partial R^2	-	0.32	-	0.32
Observations				
Individuals	113,748	113,748	113,748	113,748
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.30	0.30	0.30	0.30
ME(stdev)	-1.71%*	-4.79%***	-0.95%	-4.14%***

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

The youth-share variable is meant to measure the potential supply of young individuals to the labour market. In the paper this variable is constructed from the size of the population in the corresponding age group. However, it is likely that parts of this group are not available to the labour market and as such a population-based variable might provide an inaccurate measure of age-specific labour supply. When a youth-share variable is used instead that is defined as the number of employees aged between 15 and 24 relative to the number of employees between 15 and 64, similarly sized coefficients are obtained, but since the standard deviation of the employment-based youth-share variable is larger than in the case of the population-based variable the marginal effects increase in size (Table S7).

Table S7: Employment-based youth share variable

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-0.45 (0.37)	-3.06 (1.08)***	-0.31 (0.41)	-2.79 (1.28)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.48 (0.01)***	-	0.47 (0.01)***
First-stage statistics				
F-statistic	-	56.91***	-	56.01***
Shea's partial R^2	-	0.18	-	0.18
Observations				
Individuals	107,351	107,351	107,351	107,351
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24
ME(stdev)	-0.69%	-4.72%***	-0.48%	-4.30%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

The paper estimates the effect of the youth-share on individual wages. Alternatively, it is possible to first average individual-level variables within a region-year cell and to then regress the average daily wage within such a cell on the youth share and weighting the regression by the number of observations in a region-year cell (see Angrist and Pischke, 2009). As can be seen from Table S8, the aggregate-level analysis yields comparable results, though the marginal effects are slightly larger.

Table S8: Aggregated model (variables averaged at the level of the region-year cell)

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.65 (0.68)*	-3.71 (1.14)***	-0.89 (0.77)	-3.37 (1.43)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.45 (0.04)***	-	0.44 (0.04)***
First-stage statistics				
F-statistic	-	134.84***	-	124.29***
Shea's partial R^2	-	0.30	-	0.30
Observations				
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.82	0.81	0.82	0.81
ME(stdev)	-2.08%*	-4.68%***	-1.14%	-4.28%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

In the empirical specification the effects of age and experience on an individual's wage are approximated through the inclusion of these variables' first two polynomials. However, very similar results are obtained if mutually exclusive dummy variables are used instead (Table S9).

Table S9: Age and experience dummies

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.47 (0.63)*	-3.27 (0.98)***	-0.86 (0.73)	-2.96 (1.23)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes
Experience	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.00)***	-	0.46 (0.00)***
First-stage statistics				
F-statistic	-	136.64***	-	131.81***
Shea's partial R^2	-	0.32	-	0.32
Observations				
Individuals	107,351	107,351	107,351	107,351
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.29	0.29	0.29	0.29
ME(stdev)	-1.85%*	-4.12%***	-1.10%	-3.75%*

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

Table S10 contains the results of estimating a double-log specification. The coefficients of the youth-share variable continue to be negative and significant.

Table S10: Double-log specification

Dependent variable: log real daily earnings	Place of residence		Place of employment	
	OLS	2SLS	OLS	2SLS
Youth share	-0.24 (0.11)*	-0.52 (0.15)***	-0.13 (0.13)	-0.46 (0.19)*
Dummies				
Year	Yes	Yes	Yes	Yes
Labour-market regions	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.42 (0.00)***	-	0.42 (0.00)***
First-stage statistics				
F-statistic	-	129.62***	-	124.52***
Shea's partial R^2	-	0.35	-	0.35
Observations				
Individuals	107,351	107,351	107,351	107,351
Labour-market regions-year cells	1,296	1,296	1,296	1,296
Labour-market regions (clusters)	108	108	108	108
R^2	0.24	0.24	0.24	0.24

Cluster-robust standard errors in parentheses (clustered at the level of the labour-market region). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression.

Comparing the results of Tables 1 and 2 in the paper shows that when a district-specific youth-share variable is used rather than one based on labour-market regions, the decrease in the size of the coefficient is considerably stronger for the place of employment than the place of residence. In the following, all districts are ordered according to the difference between the number of observations in the sample that live and that work in a district. The model of Equation 1 is then estimated separately for the districts from the top half (i.e. for which the difference is largest) and for those from the bottom half (i.e. for which the difference is smaller) of this ordering. The results are shown in Tables S11 and S12, respectively. As was already discussed in the paper, negative and significant effects are only found for the set of districts from the top half of the ordering. The row *Fraction of full sample* shows that for the place-of-residence specification the majority of observations (55%) are from such districts. In the place-of-employment specification the corresponding figure stands at only 44%.

Assuming that the districts from the top half represent those which individuals are more likely to live in than to work in, the larger decrease in the size of the youth-share coefficient (i.e. a larger degree of attenuation) in the place-of-employment specification may be explained by the fact that the degree of measurement error is more pronounced in regions that people are more likely to live in and that this type of district is over-represented in the place-of-employment specification. To support this argument, the bottom rows of Tables S11 and S12 show the mean difference between the youth-share variable at the level of the labour-market region and of the district in the corresponding sample. A comparison of these differences between Table S11 and Table S12 shows that regardless of whether the place-of-residence

(0.51 as opposed to 0.42) or the place-of-employment specification (0.55 as opposed to 0.40) is used, the extent of measurement error is larger for those districts that individuals are more likely to work in than to live in.

Table S11: Districts in which individuals are more likely to live

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.78 (0.62)***	-3.63 (0.98)***	-1.07 (0.67)	-2.21 (1.12)*
Dummies				
Year	Yes	Yes	Yes	Yes
Districts	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.46 (0.03)***	-	0.45 (0.03)***
First-stage statistics				
F-statistic	-	253.16***	-	230.56***
Shea's partial R^2	-	0.45	-	0.43
Observations				
Individuals	58,705	58,705	47,146	47,146
Fraction of full sample	54.69%	54.69%	43.92%	43.92%
District-year cells	1,884	1,884	1,884	1,884
Districts (clusters)	157	157	157	157
R^2	0.23	0.23	0.23	0.23
ME(stdev)	-2.12%***	-4.33%***	-1.27%	-2.63%*
Mean difference	0.42		0.40	

Cluster-robust standard errors in parentheses (clustered at the district level). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation. Mean difference gives the average absolute difference between the district-level youth share and the value at the level of the corresponding labour-market region (multiplied by 100).

Table S12: Districts in which individuals are more likely to work

Dependent variable: log real daily earnings	**Place of residence**		**Place of employment**	
	OLS	2SLS	OLS	2SLS
Youth share	-1.05 (0.61)†	-1.75 (1.39)	0.11 (0.52)	-0.87 (1.43)
Dummies				
Year	Yes	Yes	Yes	Yes
Districts	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage statistics				
Instrument	-	0.43 (0.05)***	-	0.42 (0.05)***
First-stage statistics				
F-statistic	-	85.01***	-	60.10***
Shea's partial R^2	-	0.17	-	0.15
Observations				
Individuals	48,646	48,646	60,205	60,205
Fraction of full sample	45.31%	45.31%	56.08%	56.08%
District-year cells	1,872	1,872	1,872	1,872
Districts (clusters)	156	156	156	156
R^2	0.27	0.27	0.29	0.29
ME(stdev)	-1.63%†	-2.70%	0.18%	-1.40%
Mean difference	0.51		0.55	

Cluster-robust standard errors in parentheses (clustered at the district level). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation. Mean difference gives the average absolute difference between the district-level youth share and the value at the level of the corresponding labour-market region (multiplied by 100).

Finally, Tables S13 and S14 provide the analogues to Tables 3 and 4 but use a youth-share variable that is constructed from districts rather than labour-market regions. First, the results continue to be considerably larger in magnitude for the place of residence than the place of employment when industry and occupation dummies are added; for the place of employment, none of the coefficients are statistically significant. Second, the youth-share coefficients of the district-specific model remain smaller than the ones from the region-specific model. In the case of the place of residence the district-specific coefficients are smaller by between 27% (industry dummies) and 10% (occupation dummies). In contrast, the differences in size are much more pronounced at the place of employment where the inclusion of dummies for an individual's industrial or occupational affiliation further reduces the magnitude of the youth-share coefficients relative to those from the labour-market specification. This finding illustrates that the distinction between place of employment and place of residence is of particular importance for the estimated size and significance of the effects at the district level.

Table S13: Industry and occupation indicators (place of residence)

Dependent variable: log real daily earnings	Baseline	+industry	+occupation	+industry +occupation
Youth share (2SLS)	-2.79 (0.81)***	-2.05 (0.69)***	-1.68 (0.75)**	-1.40 (0.66)**
Youth share (OLS)	-1.31 (0.45)***	-1.10 (0.35)***	-0.93 (0.44)**	-0.86 (0.34)**
Dummies				
Year	Yes	Yes	Yes	Yes
Labour market region	Yes	Yes	Yes	Yes
Industry	No	Yes	No	Yes
Occupation	No	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage regression				
Instrument	0.44 (0.00)***	0.44 (0.00)***	0.44 (0.00)***	0.44 (0.00)***
First-stage test statistics				
F-statistic	300.92***	301.78***	301.96***	302.71***
Shea's partial R^2	0.27	0.27	0.27	0.27
Observations				
Individuals	107,351	107,351	107,351	107,351
District-year cells	3,756	3,756	3,756	3,756
Districts (clusters)	313	313	313	313
R^2 (2SLS)	0.25	0.46	0.41	0.51
R^2 (OLS)	0.25	0.46	0.41	0.51
ME(stdev, 2SLS)	-3.84%***	-2.82%***	-2.31%**	-1.92%**
ME(stdev, OLS)	-1.80%***	-1.51%***	-1.28%**	-1.19%**

Cluster-robust standard errors in parentheses (clustered at the district level). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. *Instrument* shows the coefficient of the instrument in the first-stage regression. *ME(stdev)* gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

Table S14: Industry and occupation indicators (place of residence)

Dependent variable: log real daily earnings	Baseline	+industry	+occupation	+industry +occupation
Youth share (2SLS)	-1.50 (0.92)	-1.19 (0.76)	-0.87 (0.79)	-0.83 (0.69)
Youth share (OLS)	-0.12 (0.42)	-0.23 (0.34)	0.04 (0.40)	-0.05 (0.32)
Dummies				
Year	Yes	Yes	Yes	Yes
Labour market region	Yes	Yes	Yes	Yes
Industry	No	Yes	No	Yes
Occupation	No	No	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
First-stage regression				
Instrument	0.43 (0.00)***	0.43 (0.00)***	0.43 (0.00)***	0.43 (0.00)***
First-stage test statistics				
F-statistic	181.95***	182.17***	182.12***	182.29***
Shea's partial R^2	0.22	0.22	0.22	0.22
Observations				
Individuals	107,351	107,351	107,351	107,351
District-year cells	3,756	3,756	3,756	3,756
Districts (clusters)	313	313	313	313
R^2 (2SLS)	0.26	0.47	0.42	0.52
R^2 (OLS)	0.26	0.47	0.42	0.52
ME(stdev, 2SLS)	-2.18%	-1.74%	-1.27%	-1.21%
ME(stdev, OLS)	-0.17%	-0.33%	0.05%	-0.08%

Cluster-robust standard errors in parentheses (clustered at the district level). ***/**/*/† indicate significance at the 0.005/0.01/0.05/0.10 level, respectively. Instrument shows the coefficient of the instrument in the first-stage regression. ME(stdev) gives the percentage change in daily earnings given an increase in the youth share by one standard deviation.

References

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

Eckey, H.-F., Kosfeld, R. and Türck, M. (2006) Abgrenzung deutscher Arbeitsmarktregionen, *Raumforschung und Raumordnung*, **64**, 299-309.

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