Cohort size and transitions into the labour market

Preliminary version (June 2016)

Duncan Roth[a]

Abstract

This paper estimates the effect that the size of an individual's labour-market entry cohort has on the subsequent duration of search for employment. Survival analysis is applied to empirically assess this relationship using a sample of apprenticeship graduates who entered the German labour market between 1999 and 2012. The results suggest that within a comparatively short period following graduation members of larger cohorts realise shorter search durations. These findings therefore provide no evidence for the hypothesis that increased crowding negatively affects transition into the labour market. As such the results of this paper are in line with the part of the cohort-size literature that argues that larger youth cohorts are associated with lower unemployment and higher unemployment among young individuals.

JEL classification J21, J64, R23

Key words Survival analysis, entry conditions, cohort size, apprentices, search

[a] 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

The effects of age-specific cohort size has been widely researched for a number of labour-market outcomes: the majority of empirical analyses suggest that individuals in large age groups face depressed earnings (Wright, 1991; Brunello, 2010; Moffat and Roth, 2013; Morin, 2015; Garloff and Roth, 2016), whereas the sign of the effect that cohort size has on (un-)employment continues to be debated (Korenman and Neumark, 2000; Shimer, 2001; Skans, 2005; Lucifora and Biagi, 2008; Garloff et al, 2013; Moffat and Roth, 2014). The central assumption for the relationship between age structures and labour-market outcomes is that individuals sharing certain characteristics (primarily age, though some studies also differentiate by the level of education) can be viewed as perfectly substitutable, whereas there is only imperfect substitutability across groups. As a result, changes in the size of an age group are expected to affect mainly the members of that group.

One major concern about using population-based measures of age-group size in order to estimate the effect on labour-market outcomes is given by the possibility that so-constructed cohort-size variables are measured with error because the size of an age group is not necessarily a good measure of the potential labour supply within that group due to the typically high enrolment rates in these age groups (see discussion in Moffat and Roth, 2014) or because the spatial units for which cohort size is constructed refer to administrative units (countries or sub-national entities such as states or districts) which are not necessarily delineated to economic criteria and therefore do not represent actual labour markets or because individuals within age groups are not necessarily easily substitutable. This study addresses these concerns by constructing a cohort-size variable which arguably provides a better measure for cohort-crowding among substitutable individuals. This is done by looking at the number of individuals who have graduated from an apprenticeship programme and who have entered the German labour market between 1999 and 2012. These individuals have a comparable level of qualification and have actually entered the labour market. Moreover, functional units are used in order to approximate the size of a graduation cohort within an actual labour market.

The relationship of interest to this paper is how variation in the size of graduation cohorts affects the duration that individuals spend searching for employment. Based on the extant literature two competing predictions can be formulated. The cohort-crowding hypothesis argues that due to increased competition individuals in large cohorts face a higher risk of unemployment which could be manifested in longer search periods. To the contrary, Shimer (2001) has argued that belonging to a large cohort may be beneficial for young individuals. Since firms face lower hiring costs in regions in which the potential supply of young workers is high, firms react to expected increases in young cohorts by creating jobs which in turn reduces unemployment rates among the young (as well as among older age groups). In such a case, belonging to a larger cohort may reduce the time it takes to find employment after entering the labour market.

Survival models are estimated in order to empirically assess the relationship between the size of a graduation cohort and the duration of search. The empirical findings provide no support for the hypothesis that there are negative effects from entering the labour market as part of a large cohort. In contrast, there is some evidence that for a comparatively short period following graduation, members of larger graduation cohorts are able to find employment faster, ceteris paribus, than members of smaller cohorts. The remainder of the paper is structured as follows: Section 2 provides and overview of the extant literature; the empirical analysis is the subject of Section 3, while Section 4 contains the results; Section 5 concludes.

2 Literature review

3 Empirical analysis

3.1 Data

The empirical analysis of this paper utilises two different data sources. To construct the model's main explanatory variable – the number of graduates from an apprenticeship programme – the Integrated Employment Biographies (IEB) are used. This dataset contains information on all individuals who belong to one of the following groups: employees subject to social security contributions, marginal employees, individuals receiving unemployment benefits, individuals registered as seeking employment and participants in the Federal Employment Agency's (FEA) measures of labour-market policy (groups that are not covered are civil servants and the self-employed). For each individual the dataset consists of different records that correspond to episodes in one of the above-mentioned states with specified start and end dates. Moreover, each episode is supplemented with two different sorts of information: first, characteristics of the individual are provided which refer to the beginning of the episode (among others, these characteristics include sex, nationality, year of birth, place of residence and level of education); second, details are provided that describe the state an individual is in (in the case of an employment episode, information would be available on the average daily wage during the episode, the occupation and industry of employment, place of employment as well as on the type of employment).[1]

Participation in apprenticeship programmes constitutes a separate type of employment (employment subject to social security contributions and marginal employment constitute other major categories) and as such it is possible to determine whether an individual is participating in such a training programme at any given point in time. Because a change in the type of employment – e.g. when an individual completes an apprenticeship and takes up another form of employment – entails that a new episode is defined, it is further possible to identify when an apprenticeship has ended. Based on this information, the number of individuals graduating from such a training programme in a given month, year and region can be estimated (Section 3.2. provides further details on the conditions that are imposed for an individual to be expected to have completed training). Due to its size working directly with IEB records can be cumbersome and therefore the regression analysis of this paper uses a 2% sample, the so-called Sample of Integrated Employment Biographies (SIAB).

3.2 Sample and variables

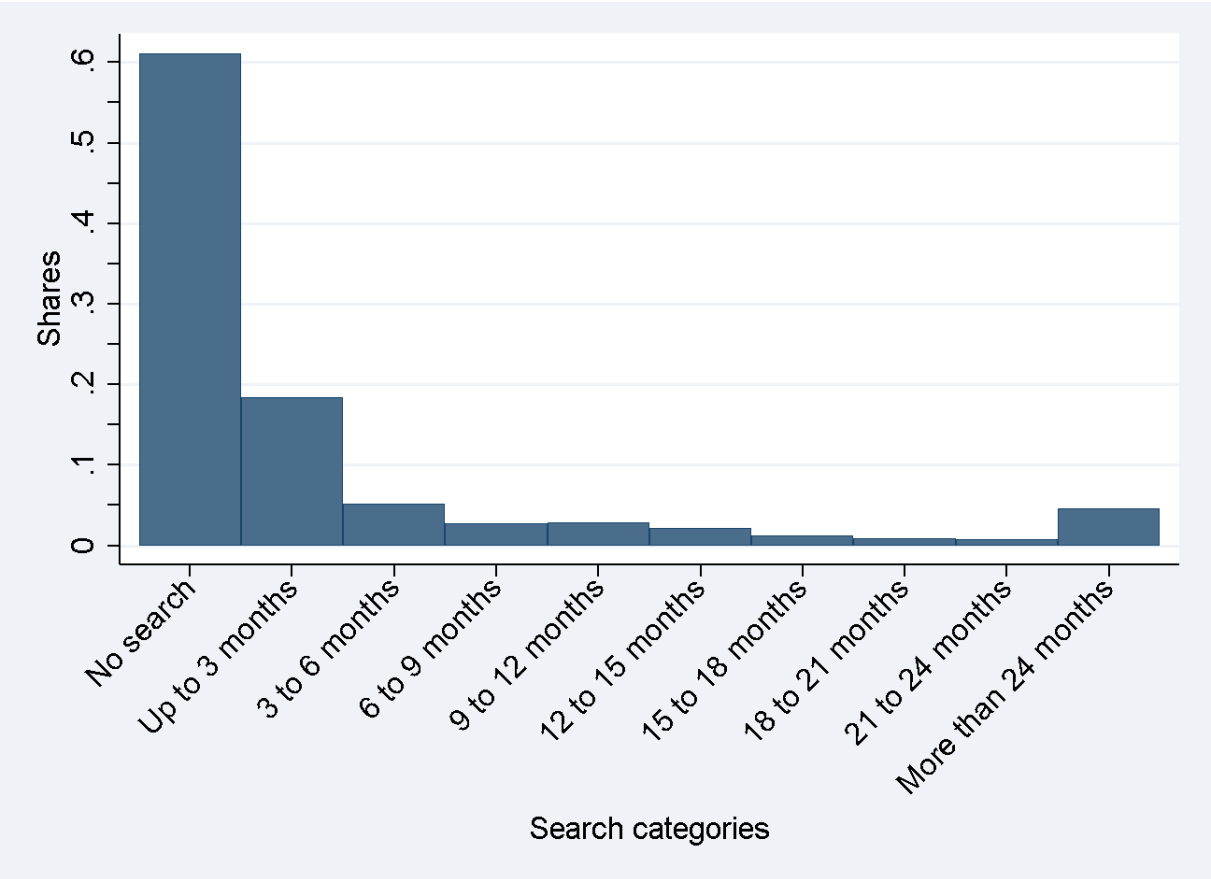
The sample consists of male individuals aged between 19 and 23 who have completed an apprenticeship. Construction of the sample from the SIAB dataset proceeds as follows: first, those individuals without any episode as an apprentice are removed. For the remaining individuals it has then to be decided whether the information on the registered apprenticeship episodes also warrant the assumption that training was also completed. This is done by imposing two criteria: first, the combined duration of apprenticeship episodes has to be at last 730 days. While completion of training can often require more than two years, it is the case that individuals with a higher secondary

[1] Variables differ in the extent to which they are provided. An individual's level of education is an example of a variable for which information can often be missing. Moreover, changes in classifications, e.g. in the coding of occupations, can cause problems in constructing a consistent coding scheme over longer periods of time.

education degree are able to complete an apprenticeship faster than those without a comparable schooling certificate. The rationale for setting a comparatively low threshold is thus to avoid excluding those who have completed secondary school. On the other hand, the risk of including individuals in the sample who have not completed training appears limited since they have already completed at least two years and dropping out of such schemes can be expected to typically happen earlier. Second, it is required that any gaps between two apprenticeship episodes are no longer than 100 days. A possible reason for such breaks is that training also includes a coursework component which does not take place within the training company. No additional restrictions are imposed; in particular changes in the training company, in the occupation or industry during the apprenticeship are disregarded because parts of the training should be sufficiently general as to be transferable to a different company, occupation or industry.

In order to avoid any confounding effects of selected female labour-market participation, the sample is restricted to men. Moreover, the age range of the sample is homogenised to include only those between the age of 19 and 23 because this is the age range at which graduates typically complete their training.[2] However, it is shown in the Appendix that the results of the empirical results are hardly affected by this restriction. Applying this procedure yields a sample of 52,234 individuals who have graduated between January 1999 and October 2012 and for whom transition into employment can be observed.

Figure 1: Distribution of search durations



[2] SIAB only includes an individual's year of birth. Age at the time of graduation is defined as the difference between the year of graduation and the year of birth. Some individuals who are registered as being 25 upon graduation will therefore actually be between 24 and 26. Out of all male observations in SIAB with a completed apprenticeship 92% are aged up to 25.

The model's dependent variable, *search*, is defined as the number of days it takes an individual to find employment after graduating from the apprenticeship programme. Figure 1 shows the distribution of this variable for the sample of individuals described above across different intervals. The distribution is highly skewed as the majority (61%) falls into the category *No search*, which means that the employment episode of these individuals starts the day after graduation from apprenticeship training. Approximately 80% of graduates make a transition into employment within 3 months after graduation, with this figure increasing to over 85% after 6 months.

Despite the large number of individuals who find employment directly upon graduation, this group is excluded from the empirical analysis. The technical reason for this exclusion is that the duration models of Section 3.3 require strictly positive durations. More importantly, zero and strictly positive durations cannot necessarily be regarded as the outcomes of the same process. Instead I argue that firms first decide whether to offer an apprentice a position after graduation from the training programme, with this decision being based on the performance of apprentices during training as well as on the economic condition of the firm. Apprentices are then free to either accept or decline the offer. If no match between training firm and apprentice is reached, individuals enter the labour market and search for employment. The empirical analysis therefore models search duration conditional on an apprentice not being directly employed by his training firm.

The obvious drawback of this approach is that those individuals who are employed directly might not constitute a random sample of graduates. In contrast, it is more likely that firms employ those apprentices which they believe to be especially productive. Put differently, these individuals might possess unobservable which are not directly observable but which are relevant for on-the-job performance. If these characteristics also increase employability at other firms, graduates who are directly employed can be expected to experience shorter search periods in the counter-factual case of not being directly employed by their training firm.

Table 1 assess this hypothesis by comparing average values of a number of characteristics between those apprentices who are employed directly and those who experience a strictly positive search duration. The first three variables refer to characteristics of the apprenticeship episode and while the difference in average duration of training and the share of Germans is statistically significant at the 0.01 level, the absolute difference in the variables is very small compared to the mean values of both groups. There is no statistically significant difference in average age at graduation. In contrast, there are sizeable and significant differences between characteristics of the employment spell that follows graduation: average daily earnings are about 20 Euros smaller for individuals who are not directly employed and the share of individuals working in part time is higher by about 15 percentage points. These latter findings suggest that both groups differ with respect to characteristics that are relevant for labour-market performance.

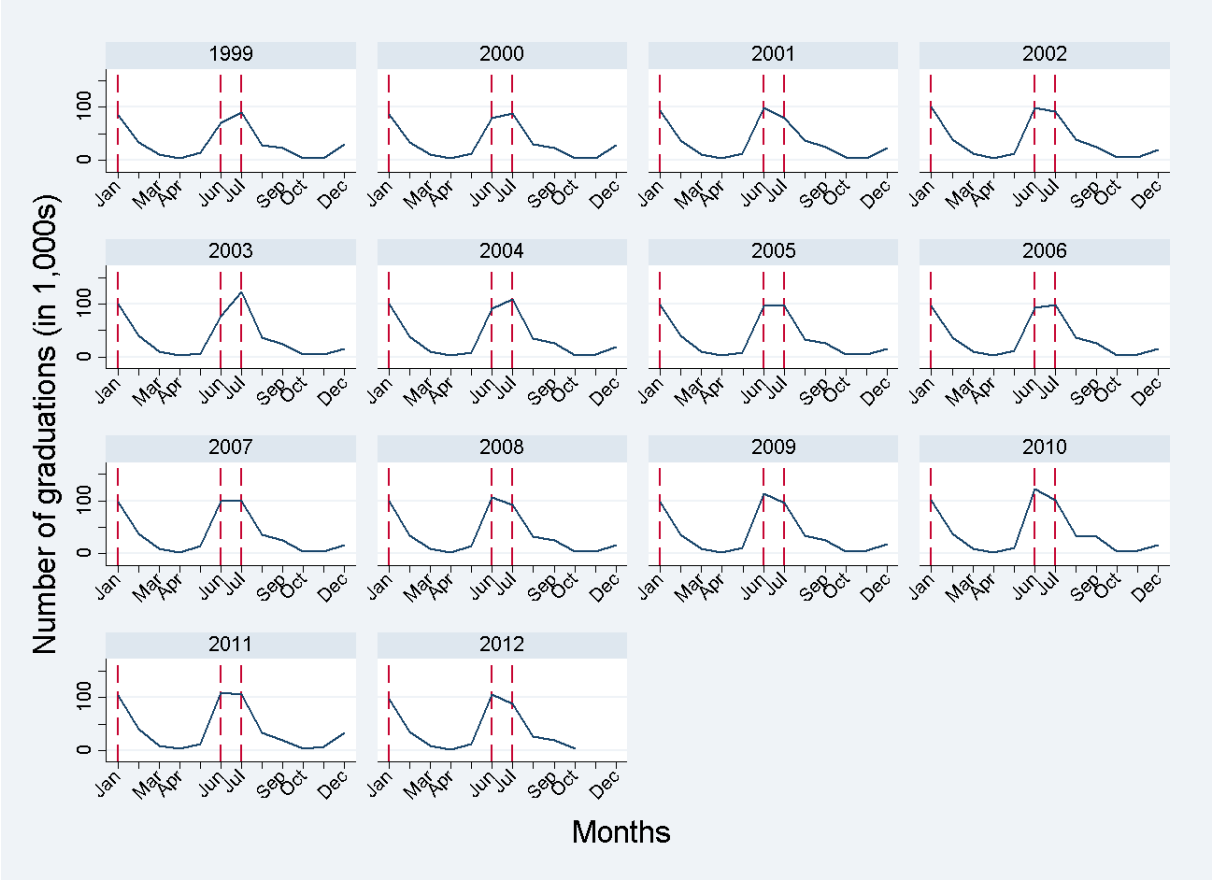
Table 1: Mean differences between individuals with no and strictly positive search duration

Variable	Observations	Group 1 (*search=0*)	Group 2 (*search>0*)	Mean difference
Apprenticeship episode				
Duration of training	52,234	1,095.00	1,086.84	-8.16***
Age at graduation	52,234	21.16	21.17	0.01
German	52,226	0.97	0.95	-0.01***
Employment episode				
Average daily earnings	52,234	61.78	42.32	-19.46***
Part-time share	52,196	0.00	0.16	0.15***

Values derived from a regression on a group indicator as well as dummies for period and region of graduation. Robust standard errors are used. ***/**/* signifies significance at the .01/0.05/0.1 level. Differences in the number of observations are due to missing values of the corresponding variables.

The main explanatory variable, *cohort*, measures the regional supply of apprenticeship graduates and is based on the number of individuals that complete training and thus become available to the labour-market within a given six-months period. Figure 2 shows the monthly number of graduates for the years 1999-2012. The annual distribution displays two peaks – one in January and another in June and July – which suggests that the bulk of apprentices complete training at two distinct points in time each year. To better reflect this pattern, the size of graduation cohorts is not computed for the whole year, but separately for two six-months periods that are centred on the peaks: November-April and May-October.

Figure 2: Monthly number of graduates (1999-2012)



It is assumed that the duration of search for employment is influenced by the conditions of the labour market that an individual enters after graduation. The variable *cohort* measures the characteristic that is most relevant to this analysis: the degree of labour-market crowding among recently graduated apprentices. In order to avoid measurement error, graduation cohorts are constructed at the level of the 141 labour-market regions that are defined by Kosfeld and Werner (2012). These entities combine one or more districts (equivalent to the third layer of the Nomenclature of Territorial Units for Statistics (NUTS)) based on how closely these are integrated as measured by commuting flows. These labour-market regions therefore approximate self-contained units in which the employed population is exclusively recruited from the resident population. Since administrative units do not necessarily delineated according to economic criteria, they cannot be relied upon to provide an accurate measure of the graduation cohorts within an actual labour

market. This argument is supported by findings of Garloff and Roth (2016) that the effect of cohort size on wages appears to be biased downwards when measured at the district level as compared to the level of labour-market regions.[3] Finally, to ensure comparability of the site of the graduation cohort across different labour-market regions, this quantity is standardised by the size of total employment in the region.[4]

Additional control variables are given by dummy variables for an individual's age at graduation, for whether an individual is of German nationality, for the occupation and industry of the apprenticeship[5] and the unemployment rate at the level of the labour market. Summary statistics of these variables are given in the Appendix.

3.3 Model

To empirically assess the effect that the size of the graduation period has on an individual's search duration, the following Cox model is specified where subscripts *i*, *r* and *p* refer to the individual, the region and the period of graduation:

$$h_i(t) = h_0(t)e^{(\beta cohort_{rp} + \boldsymbol{\gamma' x_{ipt}})} \quad [1]$$

Instead of formulating a relationship between the search duration and covariates this model is specified in terms of the hazard rate $h_i(t)$, which can be interpreted as the instantaneous probability that an individual realises a transition from search into employment. The term $h_0(t)$ represents the baseline hazard, i.e. the hypothetical hazard rate of an individual for whom all covariates take on the value zero. The Cox model belongs to the class of proportional hazard models meaning that changes in covariates shift the hazard rate up or down relative to the baseline hazard. The variable $cohort_{rp}$ captures the size of the graduation cohort in region *r* and period *p* relative to the number of employed individuals in that region. Sign and significance of the coefficient β therefore provide the basis for assessing the effect that cohort-crowding has on the duration of job search. The vector x_{irp} contains the above-mentioned set of control variables as well as dummy variables for period and region of graduation. The coefficients of the model are derived by Maximum Partial Likelihood Estimation (MPLE).[6] To account for the different levels of aggregation of the dependent variable, which is individual-specific, and the main explanatory variables, which only varies across regions and periods, standard errors are clustered at the level of the labour-market region.

Four different specifications of this model are estimated which differ with respect to the specified period of time during which transitions into employment are observed. An inherent asymmetry of the data is given by the fact that individuals that complete their apprenticeship training earlier can be observed for a longer period of time (up to 31 December 2014) and as such can also accumulate

[3] While the share of individuals in the sample who live and work in the same region is high (…%), there are also cases where the region of residence and the region of employment do not coincide. In the empirical analysis we construct the *cohort* variable at the level of the region of employment as this would appear the more relevant place for labour-market outcomes. However, using the place of residence instead yields comparable results.

[4] For the first period (November-April) employment numbers refer to 31 March of the year, while it is 31 October for the second period (May-October).

[5] Occupation indicators refer to the 2-digit level of the coding scheme *Klassifikation der Berufe 2010*,industry indicators are based on the 2-digit codes of the *Klassifikation der Wirtschaftszweige 1993*.

[6] The term partial refers to the fact that in contrast to fully parametric models, information on the search durations themselves is not used in the estimation. Instead, the relationship between the hazard rate and the covariates is derived solely from the ordering of the search durations.

longer search durations. To ensure comparability between graduates from different periods, four common periods of observation following graduation are defined: 3 months, 6 months, 12 months and 24 months. Individuals that find employment after the end of the common observation period are treated as not having realised a transition (i.e. they are right-censored).[7] When a common period of observation of 3 months is chosen the share of right-censored observations in the sample is 47%. The corresponding values are 61%, 75% and 89% for the 6-months, 12-months and 24-months periods.

To consistently estimate the effect that the size of the graduation cohort has on the duration of search, it has to be assumed that individuals did not systematically select a region in which to undertake the apprenticeship training on the basis of their expectations regarding the probability of finding employment upon graduation. While the absence of regional selection appears unlikely in the context of other studies on cohort-size effects (see Moffat and Roth, 2013), I argue that this possibility is less of a concern in this case. First, the region in which the apprenticeship is completed will in most cases be the same region in which training episode started (except for individuals who move from one region to another during training). Since individuals are typically young when they start training, the region in which an apprenticeship is being undertaken will usually be determined by the region they live in at that time. Second, it appears unlikely that reliable expectations can be formed about the economic conditions prevailing in a region at the time of graduation. Moreover, if self-selection occurs into regions that constantly provide better employment opportunities for apprentices (and hence shorter search durations), this effect would be captured by the region dummies.

4 Results

Table 2 contains the coefficients from estimating the model of Equation 1 for each of the four common observation periods (3 months, 6 months, 12 months and 24 months).

Table 2: Regression results

	3 months	**6 months**	**12 months**	**24 months**
Cohort	26.68**	15.40	11.97	6.07
	(11.23)	(9.42)	(9.39)	(8.61)
Unemployment rate	-1.98*	-0.98	-1.54*	-1.41*
	(1.09)	(0.98)	(0.87)	(0.81)
German	0.09	0.02	-0.05	-0.05
	(0.06)	(0.05)	(0.05)	(0.04)
Age				
20	-0.04	-0.05	-0.06	-0.06
	(0.03)	(0.03)	(0.03)	(0.03)**
21	-0.07**	-0.06*	-0.07**	-0.05
	(0.03)	(0.03)	(0.03)	(0.03)*
22	-0.13***	-0.12***	-0.12***	-0.09***
	(0.04)	(0.04)	(0.03)	(0.03)
23	-0.07	-0.04	-0.04	-0.04
	(0.04)	(0.04)	(0.04)	(0.04)
Dummies				
Occupation	Yes	Yes	Yes	Yes

[7] Right-censored observations are not dropped from the regression. While no transition time is recorded for these observations, they are included in the 'risk set', i.e. the set of observations that are at risk of realising a transition into employment at each of the recorded transition times.

Industry	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Log pseudo-likelihood	-81,391.31	-103,443.45	-126,154.12	-145,261.56
Observations	18,133	18,133	18,133	18,133

All variables refer to the time of graduation from apprenticeship training. Coefficients are expressed as proportional hazard estimates. ***/**/* signifies significance at the 0.01/0.05/0.1 level of significance. Standard errors are clustered at the level of the labour-market region. The Breslow method is used to handle tied observations.

The coefficients of the variable measuring the size of the graduation cohort are positive which indicates that an increase in the degree of labour-market crowding increases the rate of transiting into employment. Equivalently, these results suggest that search durations are shorter in larger cohorts. This effects is, however, only significant in the case of the 3-months observation period. For longer periods the coefficient becomes smaller in size and insignificant.

These results provide no evidence for the hypothesis that an increased degree of labour-market crowding leads to longer periods of search. In contrast, appear to suggest that apprentices benefit from belonging to large graduation cohorts in terms of being able to transit into employment at a faster rate, though this beneficial effect seems to be restricted to a relatively short period following graduation. As such these findings are in line with the mechanism described in Shimer (2001) which lead to the unemployment (employment) rate being smaller (larger) among members of large cohorts. While the empirical evidence suggests that there are advantages for apprentices of entering the labour market as part of a large cohort, other aspects so far remain unaddressed: does the size of the graduation cohort affect the quality of the subsequent job or the duration of employment?

5 Conclusion

References

Biagi, F. and Lucifora, C. (2008). Demographic and education effects on unemployment in Europe. *Labour Economics, 15*(5)*,* 1076-1101. doi:10.1016/j.labeco.2007.09.006

Brunello, G. (2010). The effects of cohort size on European earnings. *Journal of Population Economics*, *23*(1), 273–290. doi:10.1007/s00148-009-0250-y

Garloff, A., Pohl, C., and Schanne, N. (2013). Do small labor market entry cohorts reduce unemployment? *Demographic Research*, *29*, 379–406. doi:10.4054/DemRes.2013.29.15

Garloff, A., and Roth, D. (2016). Regional population structure and young workers' wages (IAB Discussion Paper No. 06/2016). Retrieved from http://doku.iab.de/discussionpapers/2016/dp0616.pdf

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.), *NBER Comparative labor markets series. Youth Employment and Joblessness in Advanced Countries* (pp. 57–105). Chicago: University of Chicago Press.

Kosfeld, W., and Werner, A. (2012). Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007-2011. *Raumforschung und Raumordnung 70*, 49-64.

Moffat, J. and Roth, D. (2013). *The CohortSize-Wage Relationship in Europe* (MAGKS Papers on Economics, Philipps-Universität Marburg No. 201346). Retrieved from http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/46_2013_moffat.pdf

Moffat, J. and Roth, D. (2014). *Cohort size and youth unemployment in Europe: a regional analysis* (MAGKS Papers on Economics, Philipps-Universität Marburg No. 201440). Retrieved from http://www.uni-marburg.de/fb02/makro/forschung/magkspapers/40-2014_roth.pdf

Morin, L.-P. (2015). Cohort size and youth earnings: Evidence from a quasi-experiment. *Labour Economics, 32*, 99-111. doi:10.1016/j.labeco.2015.01.001

Shimer, R. (2001). The Impact of Young Workers on the Aggregate Labor Market. *The Quarterly Journal of Economics*, *116*(3), 969–1007. doi:10.1162/00335530152466287

Skans, O. N. (2005). Age effects in Swedish local labor markets. *Economic Letters, 86*(3), 419-426. Doi:10.106/j.econlet.2004.09.004

Wright, R. E. (1991). Cohort size and earnings in Great Britain. *Journal of Population Economics*, *4*(4), 295-305. doi:10.1007/BF02426373

Appendix

Table A1: Descriptive statistics

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
Search duration	18,133	313.301	546.772	1	5179
Cohort	18,133	0.009	0.003	0.004	0.020
Unemployment rate	18,133	0.120	0.053	0.024	0.317
German	18,133	0.954	0.209	0	1
Age					
19	18,133	0.144	0.351	0	1
20	18,133	0.292	0.455	0	1
21	18,133	0.278	0.448	0	1
22	18,133	0.175	0.380	0	1
23	18,133	0.111	0.314	0	1

Table A2: Regression results (no age restriction)

	3 months	6 months	12 months	24 months
Cohort	23.61**	12.73	10.27	4.29
	(9.40)	(8.14)	(8.18)	(7.73)
Control variables	Yes	Yes	Yes	Yes
Dummies				
Occupation	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
Log pseudo-likelihood	-105,305.34	-134,329.00	-163,828.24	-187,466.92
Observations	22,828	22,828	22,828	22,828

All variables refer to the time of graduation from apprenticeship training. Coefficients are expressed as proportional hazard estimates. ***/**/* signifies significance at the 0.01/0.05/0.1 level of significance. Standard errors are clustered at the level of the labour-market region. The Breslow method is used to handle tied observations.