

# Modeling Spatial Opportunity Structures and Youths' Transitions from School to Training

Alexandra Wicht<sup>1</sup>, Alexandra Nonnenmacher<sup>2</sup>

<sup>1</sup>GESIS—Leibniz Institute for Social Sciences, Mannheim, Germany

<sup>2</sup>University of Siegen, Siegen, Germany

Email: alexandra.wicht@gesis.org

**How to cite this paper:** Wicht, A. and Nonnenmacher, A. (2017) Modeling Spatial Opportunity Structures and Youths' Transitions from School to Training. *Open Journal of Statistics*, 7, 1013-1038.

<https://doi.org/10.4236/ojs.2017.76071>

**Received:** November 14, 2017

**Accepted:** December 12, 2017

**Published:** December 15, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

This paper examines the significance of spatial externalities for youths' school-to-training transitions in Germany. For this purpose, it is necessary to address the methodological question of how an individual's spatial context has to be operationalized with respect to both its extent and the problem of spatial autocorrelation. Our analyses show that the "zone of influence" comprises of the whole of Germany, not only close-by districts, and that these effects differ between structurally weak and strong regions. Consequently, assuming that only close proximity affects individual outcomes may disregard relevant contextual influences, and for spatial models that require an a priori definition of the weights for spatial units, it may be erroneous to make a decision based on this assumption. Concerning spatial autocorrelation, we found that neglecting local spatial autocorrelation at the context level causes considerable bias to the estimates, especially for districts that are close to the home district.

## Keywords

Spatial Effects, Spatial Autocorrelation, Event History Models, Youths, School-to-Work Transitions, Opportunity Structures

---

## 1. Introduction and State of the Art

There is little doubt that social action and, thus, labor market processes are not just structured in terms of time [1] [2] [3], but are also localized somewhere [4] [5]. But what does this insight mean, exactly, for analyses of contextual effects of opportunity structures on individual labor market outcomes? In recent decades, researchers have become increasingly interested in questions about the regional heterogeneity of labor market conditions, in terms of industry sector and employment structure, for example, as well as in questions of how such regional

opportunity structures of labor markets affect several individual outcomes, including individual labor market success [6] [7] [8] [9]. In the literature on these issues, individuals' opportunity structures are commonly construed as regional containers. Accordingly, labor markets as opportunity structures are regarded as isolated regions within which people compete with each other for jobs or training positions; people from different regions as well as the regions themselves, however, are considered to be independent of one another.

This proposition is unrealistic, to be sure. Opportunity structures are not just point-located; rather, they are distributed in space. In other words, they are spread across the landscape instead of being restricted to a regional container. From this perspective, the impact of local labor market conditions on individual labor market outcomes may be mitigated or reinforced by surrounding areas, which also have an impact on individual outcomes [10] [11]. Hence, neglecting this spatial interrelation may result in underestimating or overestimating regional effects. In this paper, we investigate the relevance of spatial opportunity structures rather than regional opportunity structures for individual labor market processes by examining youths' transitions from school to firm-based vocational training in Germany. To this end, we link retrospectively collected life course data from Germany's National Educational Panel Study with administrative regional time series data at the level of districts (Landkreise).

When analyzing the impact of such spatial externalities [10], the crucial methodological question is how to model these effects, that is, how the wider spatial context has to be operationalized in order to capture its significance for labor market outcomes. There are two significant dimensions: The first concerns the extent of opportunity structures and the issue of whether it is sufficient to only consider immediately adjacent districts [12] or if a much wider "zone of influence" must be taken into account. This issue is related to the Modifiable Areal Unit Problem which states that the results of spatial data analyses are generally sensitive to the spatial scale and spatial zoning system used [13]. The second dimension concerns the problem of spatial autocorrelation [14] [15] of characteristics at the context level, which arises due to social interrelation and exchange between areal units. Spatial autocorrelation of indicators at the context level may lead to underestimating the effects of local and non-local opportunity structures. As yet, both methodological questions and their interrelations have not been considered by previous research.

## **2. On the Relevance of Spatial Externalities**

The concepts "spatial externalities" or "spillover effects" are of particular importance in the theoretical framework of the new economic geography [16]. They refer to interdependencies between observational units in space, such as points, regions, or nations. The idea is that one particular observation somewhere on the landscape is dependent on or influenced by one or several other observations. From a theoretical point of view, spatial interdependencies in economic out-

comes generally arise due to social interactions [17], such as the exchange of resources between relevant actors. However, while having a spatial dimension, these social interactions are not necessarily guided by physical boundaries [11] [18].

Bearing in mind that labor markets comprise social relations, the picture of a regional labor market that makes up an individual's opportunity structure accordingly falls short—social relations do not simply stop at the boundaries of a regional “container”. In contrast, thinking of individuals' opportunity structures in terms of social relations leads to an understanding of individuals' opportunity structures as a wider social system with rather unmarked boundaries. This argument especially holds if the areal units used for data analysis represent (administrative) artificial regions, such as municipalities, districts, or employment agency districts. This is usually the case in the literature: research on the significance of labor market contexts is only rarely based on empirically defined areal units [7], such as travel to work areas (for the delineation of functional labor market regions, see e.g. Kropp and Schwengler) [19].

There are two processes relevant for conceptualizing individuals' opportunity structures as a spatial social system. The first process gives promise to individual spatial mobility and the second process concerns spatial pulling effects. The search behavior of youths is not necessarily a geographically static process [20] [21]. In an aggregate data analysis of youths' work commutes in Germany, Bogai *et al.* [22] point to the importance of mobility in obtaining a training position, depending on the situation in the regional training market and settlement structures as a measure of the degree of (economic) agglomeration of the region. Whereas densely populated regions with a high supply of training positions show a surplus of commuting inflows, the opposite is true for sparsely populated regions with a low supply of training positions.

From an action theoretical perspective, spatial mobility results from youths' subjective assessments of their situation [23] [24] [25]. They will decide to search for and accept a training spot offer away from home if they expect their utility to be maximized this way. The exact utility function, however, is unclear. The utility of spatial mobility has to be balanced against both its monetary and non-monetary costs [9] [26] [27]. For instance, apprentices usually have limited financial resources and are under age in many cases, which impedes setting up a household (e.g., in 2015, the average monthly salary in Germany was 832 euros before deductions). Thus, we may assume that youths prefer to minimize commuting costs as much as possible and tend to look for training position offers in the immediate vicinity of their place of residence. In most cases, they restrict themselves to their home district and will only gradually expand their search radius if they fail to find a suitable offer nearby.

The probability of success, in turn, depends on a) the raw supply of training spots available and b) the supply of training spots in a given area relative to the respective supply in the home district. If, for example, the home district offers very few training positions compared to surrounding districts, the initial search

will most likely be unsuccessful until the search radius is widened, and finding a training position is more likely if other regions offer better opportunities in terms of quantity and/or quality, compared to the local regional situation [28] [29]. Thus, spatial mobility can be assumed to be directed to “escalator regions” [30]. From this perspective, it is not only the supply of training spots in a given area *per se* that determines an individual’s probability of finding one, but also the relative supply of positions, that is, the supply of training spots in a specific location as compared to others.

If only individual utility maximization processes were considered, as we have done thus far, we would only have to take training positions in districts within reasonable proximity to the place of residence into account. Youths would most likely terminate their search once commuting costs become unacceptable. But on the contextual level of local labor markets, we may additionally assume pulling effects, which result from processes of labor exchange in a long line of (adjacent) districts. For example, take a district with a very high supply of training spots. It will attract youths living in adjacent districts, who will leave training positions in their home district unoccupied. These, in turn, will attract youth living in the next districts and so on. Although training places are strongly concentrated in (urban) centers [31], these areas are by no means isolated from each other [32], which means that this cascade will only be interrupted in those districts where no training spots are left unoccupied. Generally, it is conceivable that even districts several hundred kilometers from home have an impact on the search duration or the probability of finding a position. Since these effects are indirect—they are mediated by vacant and filled training spot offers in a long line of adjacent districts—we nevertheless expect nearby districts to have a stronger effect than distant ones.

To sum up, we expect that the probability of finding a training spot will increase given an increasing raw or relative supply of positions. Because of individual spatial mobility and indirect pulling effects on the macro level (district), we not only expect training spots in the home district to have this effect, but training spots in all German districts. However, the effect should decrease with increasing spatial distance.

### **3. Mapping Spatial Opportunity Structures in the Presence of Spatial Interrelation**

On a theoretical level, thinking about spatial interrelations between observations leads to an understanding of an individual’s opportunity structure as a spatial social system that has to be operationalized as such. At the same time, however, spatial interrelations pose statistical problems on a methodological level because spatial dependencies lead to processes of homogenization and, hence, to spatial autocorrelation of the dependent and/or independent variables. Neglecting them, then, may produce biased estimates of regression coefficients and/or standard errors [14] [15].

When analyzing spatial effects on individual outcomes, researchers are generally faced with several kinds of spatial dependencies at the individual and/or contextual level. Such dependencies may arise, for example, due to common exposure or social exchange [33]. At the individual level, observations may not be independent of each other when individuals inhabit the same district, or if they live in districts close to each other, where spillover effects are very likely. Such spatial interdependencies concern the dependent variable and can be handled within a multilevel framework by modeling a more complex random effects structure [34] [35]. Various empirical examples in the field of research about neighborhood effects on individual outcomes even deal with such interdependencies not only as a nuisance, but as the main subject of interest [36] [37] [38].

However, when mapping an individual's spatial opportunity structure by considering the availability of non-local training positions as well, the spatial interrelation may also concern the indicators at the context level. Spatial interrelation at the context level generally produces homogenizations of the conditions of local labor markets; therefore, using such spatially correlated context data may result in inappropriate inferences about the effect of spatial opportunity structure on individuals' labor market outcomes. In case of high local spatial autocorrelation, the contextual effect of non-local training positions may even go undetected. Following Tobler's first law of geography, which states that "everything is related to everything else, but near things are more related than distant things", this especially holds true for areas that are in close proximity to one another [39].

Thus, in order to investigate an individual's zone of influence, effects of non-local training positions have to be adjusted for spatial autocorrelation. For that purpose, we have to think about the social process underlying the spatial autocorrelation of the crucial predictor variable, which in our case are comprised of local and non-local training positions. Though training spots are geographically related, the underlying process is not only guided by proximity and distance, but rather by social exchange at the labor market that depends, to a great extent, on infrastructure [31] [40]. The term "infrastructure" refers to the density of training positions that are unequally distributed across the landscape, with a high concentration of training positions in urban centers. Such centers are more attractive to apprentices than other areas that are equally distant or even closer [19]. In an exploratory analysis using data about youth unemployment at the level of districts, Wicht [31] demonstrates that infrastructure, operationalized by commuting flows between home and workplace, best captures the correlation in the data. Observed commuting behavior of apprentices, therefore, can be used to explicitly model the spatial correlation at the context level, and can be expected to be a prerequisite for being able to adequately model spatial effects.

## 4. Data and Methods

### 4.1. Individual Data and Outcome Variable

We use retrospectively collected life course data on the starting cohort 6—adults

from Germany's National Educational Panel Study [41]. The current release of the study<sup>1</sup> consists of five waves and comprises a representative sample of 17,139 individuals born between 1944 and 1986. The data collection of starting cohort 6 partly builds upon the survey "Working and Learning in a Changing World" [42] that has been conducted by the Institute for Employment Research from 2007 to 08. From 2009 on, the NEPS annually collects additional data via CATI by refreshing the ALWA sample and adding older birth cohorts.

In our analyses, we only utilize time-to-event data about youths' transitions from school to training. The outcome variable of interest is individuals' duration of finding a firm-based training position that we measure monthly. We limit ourselves to this particular transition since (in contrast to school-based training, another educational path in the German vocational training system) firm-based training is expected to strongly depend on the labor market situation. The data contains information about the district participants graduated in, but no reliable information on the place of work during apprenticeship. Therefore, we model the probability of a youth finding a training spot somewhere in Germany, conditional on the training market conditions in her home district as well as in all German districts.

Since some of the relevant regional data is only available from 1999 onwards, we are restricted to focus on groups graduating from school between 1999 and 2006. From the entire sample we are able to extract relevant transition data for about 410 individuals and, on the level of our analyses, 2799 person-time units. The large amount of missing cases mainly arises from the restricted observation period and further sample selection procedures: From the valid information about 14,832 school graduates from 1957 onwards, 1683 individuals finish school between 1999 and 2006, of which 803 individuals begin firm-based vocational training. Further missing cases result either from missing information about the place of residence, which is necessary for merging individual data with regional data, or from missing information on relevant control variables.

When investigating life course data, exact operationalization is of high importance, for individual life courses are not necessarily standardized. For example, individuals may attend more than one school during their lives or they may have temporal gaps between school episodes and complete their schooling later; they may begin an apprenticeship somewhere, abandon it, and start another one elsewhere. In our analyses, we treat the end of the first uninterrupted school career as the starting point of the search time needed to find a training position; uninterrupted school episodes are defined as those in which temporal gaps amount to no more than four months (or six months in case of previous primary school attendance). This way, we ensure that individuals are actually searching for a training position rather than continuing their school career. The search

---

<sup>1</sup>[doi:10.5157/NEPS:SC6:5.0.0](https://doi.org/10.5157/NEPS:SC6:5.0.0) From 2008 to 2013, NEPS data were collected as part of the "Framework Programme for the Promotion of Empirical Educational Research" funded by the German Federal Ministry of Education and Research. As of 2014, the Leibniz Institute for Educational Trajectories at the University of Bamberg conducts the NEPS survey in cooperation with a nationwide network.

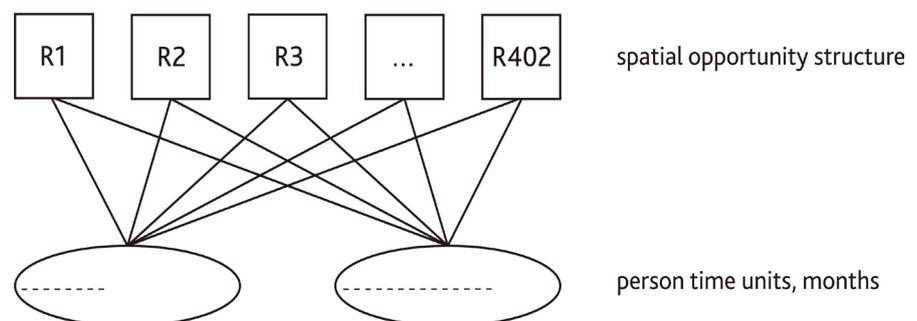
time ends when an individual engages in fully qualifying firm-based vocational training for at least three months. Since we are only interested in events that are directly driven by the spatial opportunity structure of labor markets, we excluded people who begin other forms of training, such as school-based vocational training or university education, as well as right-censored cases.

#### 4.2. Regional Data and the Modeling of Individuals' Spatial Opportunity Structures

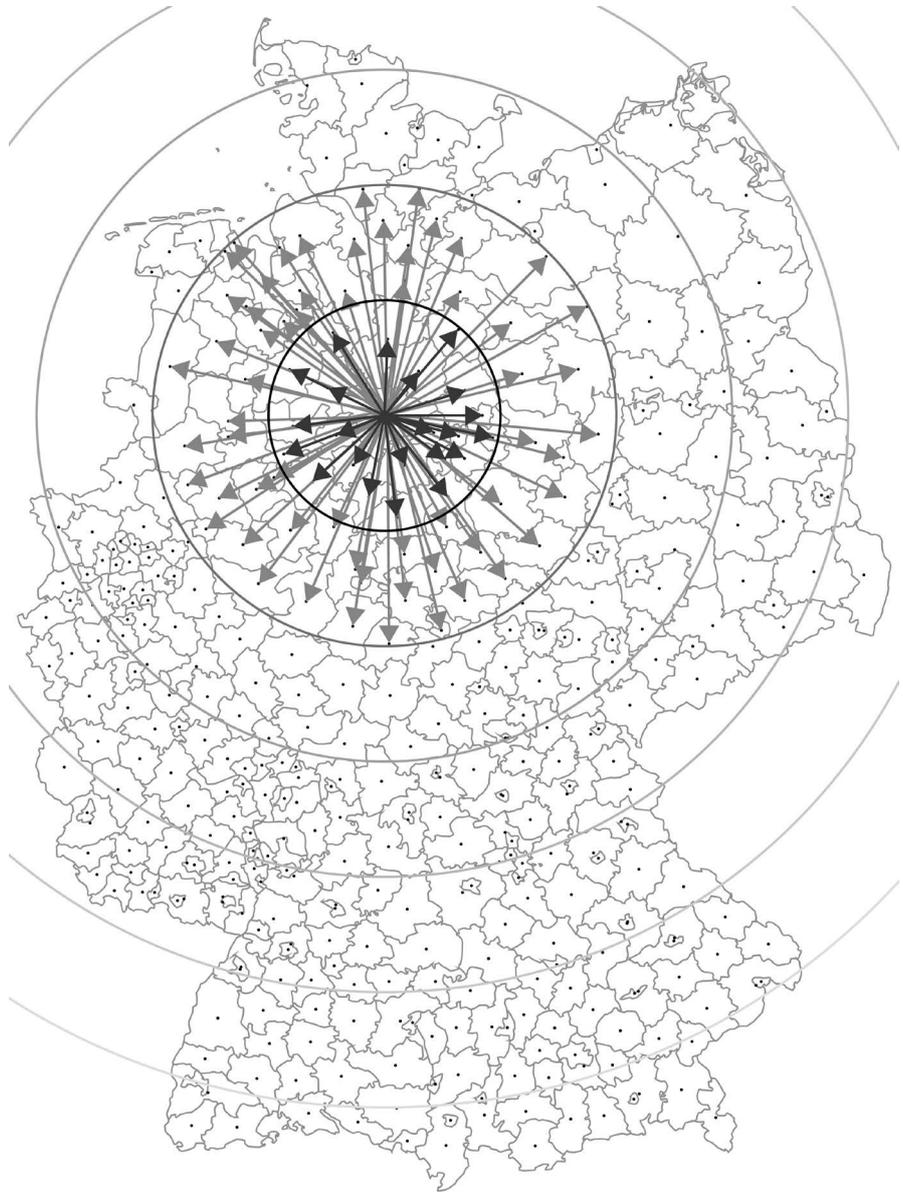
In order to map an individual's spatial opportunity structure, we make use of a dataset consisting of several  $N \times N$  matrices that represent the varied pairwise connectivity between administrative regional units at the district level (Landkreise and kreisfreie Städte, NUTS 3 regions) in Germany. Since Germany is currently divided into 402 of these areal units, all spatial connectivity matrices comprise  $402 \times 402$  districts. In general, merging these matrices with individual data results in a multiple membership dataset, in which each individual or, more precisely, each person-time unit is nested within all of the 402 districts. The structure of the dataset is illustrated in **Figure 1**.

The zone of influence can be modeled by bringing distances between the spatial units into the equation. To this end, we use geographical data about the administrative areas from the Federal Agency for Cartography and Geodesy (GeoBasis-DE 2015), which enables us to calculate Euclidean distances between the geometric centers of the districts. For this purpose, we use the projected coordinate system Universal Transverse Mercator (UTM) as geographic coordinate systems generally tend to produce biased distance calculations. The calculated distances between the districts allow us to qualify the spatial relation of each district to all other districts, as illustrated in **Figure 2**. In our analyses, we transform the geographical distances into nine equal intervals, represented by the circles in **Figure 2**. The first category comprises the home district itself (where distance equals zero) and surrounding districts with a distance up to 75 kilometers. Defining this first category any narrower would have led to an insufficient number of cases.

In addition, we use data on commuting flows of apprentices between home and workplace from the Institute for Employment Research [43], which defines



**Figure 1.** Multiple membership data structure for investigating effects of spatial opportunity structures on individual processes.



**Figure 2.** Mapping zones of influence.

the pairwise connection between the districts by the proportion of commuters from district  $i$  to district  $j$  compared with all commuters from district  $i$ . In contrast to the spatial connectivity matrix based on Euclidean distances, this matrix corresponds to a directed spatial network, that is, for each district it represents the importance of all possible destination districts as a training location for non-local youths. We use this measure in order to adjust the effects of the availability of non-local training places for spatial autocorrelation at the context level that mainly arises due to commuting. Likewise, we use discrete measures of the commuting flows from home to the workplace, but define the thresholds according to analytical criteria, as the distribution of the measure is strongly positively skewed.

Beside the category that refers to the connection of one district with itself

(where commuting equals zero), we distinguish between the categories “no commuting flows”, “commuting flows up to 0.05%”, “commuting flows from more than 0.05% to 0.1%”, “commuting flows from more than 0.1% to 0.2%”, “commuting flows from more than 0.2% to 0.5%”, “commuting flows from more than 0.5% to 1%”, and “commuting flows of more than 1%”.

Finally, we use a set of control variables at the individual and at the macro level as adjustment factors in our regression models. At the individual level, we control for individuals’ gender, ethnicity, and educational level as well as a time-varying variable that indicates whether vocational preparation was completed or not. At the macro level, we control for the school graduate cohort as a measure of the overall economic situation in Germany at the time of entering the labor market.

Moreover, we differentiate our analyses between structurally weak and strong areas: Several studies [22] [44] have shown that the mobility behavior of unemployed individuals varies according to regional labor market conditions. In high-unemployment regions, the unemployed are—counter-intuitively—less mobile than in low-unemployment regions, which leads to an “unemployment trap”. If this holds true for youths seeking training positions as well, we have to separately analyze regions with favorable and unfavorable training market situations, respectively. To this end, we divided the sixteen German states (Bundesländer) into structurally strong and weak areas, based on the results presented by Bogai *et al.* [22], and assume that non-local training market effects are weaker in the structurally weak areas.

Structurally weak states are the East German states, Schleswig-Holstein and Lower Saxony in the Northwest, and Rhineland-Palatinate in the far West. All other West German states and the two city-states, Berlin and Hamburg, are considered to be structurally strong.<sup>2</sup> Additionally, in our analyses we control for the state to capture possible variations of mobility levels due to unobserved variables. **Table 1** shows the descriptive results of all individual and contextual variables used in our regression analyses.

### 4.3. Statistical Analyses

The appropriate framework for investigating individual processes is event history (or survival) analysis. We conduct our analyses of youths’ transition from school into firm-based training with the aid of a discrete-time event history model [45] [46], as individuals’ search times are not measured continuously.

The model we apply corresponds to a binary logistic regression model. The dependent variable is the transition probability from state 0 (completion of school) to state 1 (taking a firm-based training position). The transition probability is defined as the log-odds for the conditional probability of taking a firm-based training position at time  $t_i$ . Since the transition probability is dependent

<sup>2</sup>Our data set does not contain any cases in Bremen, the third city-state. To ensure that Hamburg and Berlin do not determine our results for structurally strong regions, we replicated our analyses with Berlin and Hamburg excluded. The results were the same.

**Table 1.** Descriptive statistics of predictor variables.

	Percent/mean (SD)		
	Structurally weak regions	Structurally strong regions	Total sample
Search time (ref. 0 - 4 months)	41.1	34.8	36.8
5 - 16 months	39.6	38.5	38.8
17 - 28 months	11.0	15.0	13.8
>28 months	8.3	11.7	10.6
Female	36.7	41.5	40.0
Native language not German	7.6	26.4	20.4
Educational level (ref. basic secondary education <sup>a</sup> )	35.8	31.5	32.9
Secondary education <sup>b</sup>	44.9	44.6	44.7
Higher secondary education <sup>c</sup>	19.2	23.9	32.9
Vocational preparation completed	14.4	8.9	10.7
Cohort (ref. 1999)	17.1	11.6	13.3
2000	18.1	20.7	19.9
2001	19.0	14.3	15.8
2002	21.7	14.3	16.7
2003	8.4	17.4	14.6
2004	6.4	11.2	9.7
2005	6.0	6.8	6.5
2006	3.4	3.7	3.6
Raw supply of training places	98.4 (5.4)	97.8 (5.4)	98.0 (5.4)
Relative supply of training places	0.98 (0.07)	1.00 (0.08)	0.99 (0.07)
Distance (ref. home district + 75 km)	3.9	6.2	5.5
>75 - 150 km	9.0	13.2	11.9
>150 - 225 km	13.4	17.8	16.4
>225 - 300 km	16.3	18.0	17.4
>300 - 375 km	18.5	15.6	16.5
>375 - 450 km	16.8	12.8	14.1
>450 - 525 km	11.2	10.0	10.3
>525 km	10.8	6.6	7.9
Commuting flows (ref.: >1%, incl. home district)	2.6	2.8	2.7
≤1%, >0.5%	1.1	1.2	1.1
≤0.5%, >0.2%	3.3	3.2	3.2
≤0.2%, >0.1%	5.9	4.9	5.2
≤0.1%, >0.05%	8.1	6.1	6.8
≤0.05%, >0%	5.2	5.5	5.4
0%	73.9	76.4	75.6
N (person-months units)	890	1909	2799

on individual search time, we follow a piecewise-constant modeling strategy, where the transition probability is assumed to be constant only within particular time intervals [47]. In order to depict this, we divide the individual search times into four categories (1 to 4 months, 5 to 16 months, 17 to 28 months and more than 28 months), which we use as additional controls in our model.

In order to take the hierarchical structure of our data into account, we make use of a Huber-White sandwich estimator of variances in order to obtain cluster-robust standard errors. In contrast to hierarchical models [34], which are commonly applied to nested data, this procedure is considerably less computationally demanding. Since we are not interested in the evaluation of variances, we follow this simpler strategy. We specify our model in such a way that person-time-units are nested within the 402 districts and, hence, allow for non-independence of the observations within districts.

## 5. Results

In order to investigate the influence of non-local training positions on youths' probability of finding a spot, we look at both the raw supply of local and non-local training spot offers and the ratio between local and non-local training positions. Both kinds of measures are introduced as interactions with spatial distance. In accordance with our hypotheses, we expect the interaction effects between the raw and the relative supply of training positions, irrespective of their distance to the home district, to increase the probability of finding a spot. However, to reveal the impact of the availability of non-local training positions, the regression coefficients have to be adjusted for spatial autocorrelation. To demonstrate the relevance of spatial autocorrelation, we compare two regression models, as shown in **Table 2** (the complete regression results can be found in **Table A1(a)** and **Table A1(b)** in the appendix). While in models 1a and 2a, we only consider the interactions between the raw and relative supply of training positions and spatial distance, respectively, in models 1b and 2b we also consider commuting flows of apprentices. This is done by introducing further interaction effects between the raw and relative supply of training spots and commuting flows from home to workplace. Accordingly, while models 1a and 2a show biased estimates, models 1b and 2b are corrected for spatial autocorrelation, which arises due to commuting.

As expected, the main effect for the raw supply of training positions shows that the probability of finding a spot increases with an increasing supply of training places in the first zone—the home district plus 75 kilometers. This applies to both structurally weak and strong regions and to both models a and b. However, the expected positive effects of the supply of training positions in districts that are too far away from home to be reached by commuting—in other words, the pulling effects on the district level—are not apparent in model 1. In structurally weak regions, we find unexpected negative interaction effects for both the raw and relative supply of training positions, meaning that a strong

**Table 2.** Spatial effects of the raw and relative supply of training places offered on youths' transitions from school to firm-based vocational training, Germany, 1999-2006, discrete-time event history models, standardized logit coefficients.

	Structurally weak regions		Structurally strong regions	
	<i>Model 1a:</i>	<i>Model 1b:</i>	<i>Model 2a:</i>	<i>Model 2b:</i>
	Not controlling for commuting flows	Controlling for commuting flows	Not controlling for commuting flows	Controlling for commuting flows
	$\beta$	$\beta$	$\beta$	$\beta$
Raw supply of training places	0.363***	0.333***	0.065***	0.098***
Relative supply of training places	0.405***	0.376***	0.021	0.124***
Raw supply $\times$ distance (ref.: home district + 75 km)				
>75 - 150 km	-0.047	-0.128**	-0.000	0.058*
>150 - 225 km	-0.085*	-0.187***	0.034	0.101***
>225 - 300 km	-0.107**	-0.224***	0.001	0.072**
>300 - 375 km	-0.163***	-0.281***	0.049**	0.121***
>375 - 450 km	-0.112**	-0.229***	0.063***	0.136***
>450 - 525 km	-0.111**	-0.238***	0.083***	0.156***
>525 km	-0.113**	-0.249***	0.108***	0.179***
Relative supply $\times$ distance (ref.: home district + 75 km)				
>75 - 150 km	0.092	0.011	-0.026	0.067*
>150 - 225 km	0.017	-0.092	0.021	0.124***
>225 - 300 km	-0.011	-0.141*	0.030	0.136***
>300 - 375 km	-0.094*	-0.227***	0.071***	0.181***
>375 - 450 km	-0.074	-0.206***	0.113***	0.222***
>450 - 525 km	-0.110*	-0.244***	0.115***	0.223***
>525 km	-0.118*	-0.259***	0.172***	0.278***

Notes: \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001 (two-tailed tests), cluster-robust S.E., controls: main effect of distance, main effect of commuting (ref.: home district + strong commuting flows), interaction effects between commuting and the raw supply of training places as well as the relative supply of training places respectively, school-leaving cohort, state (*Bundesland*), individual search duration, sex, ethnic origin, school-leaving qualification, completion of vocational preparation.

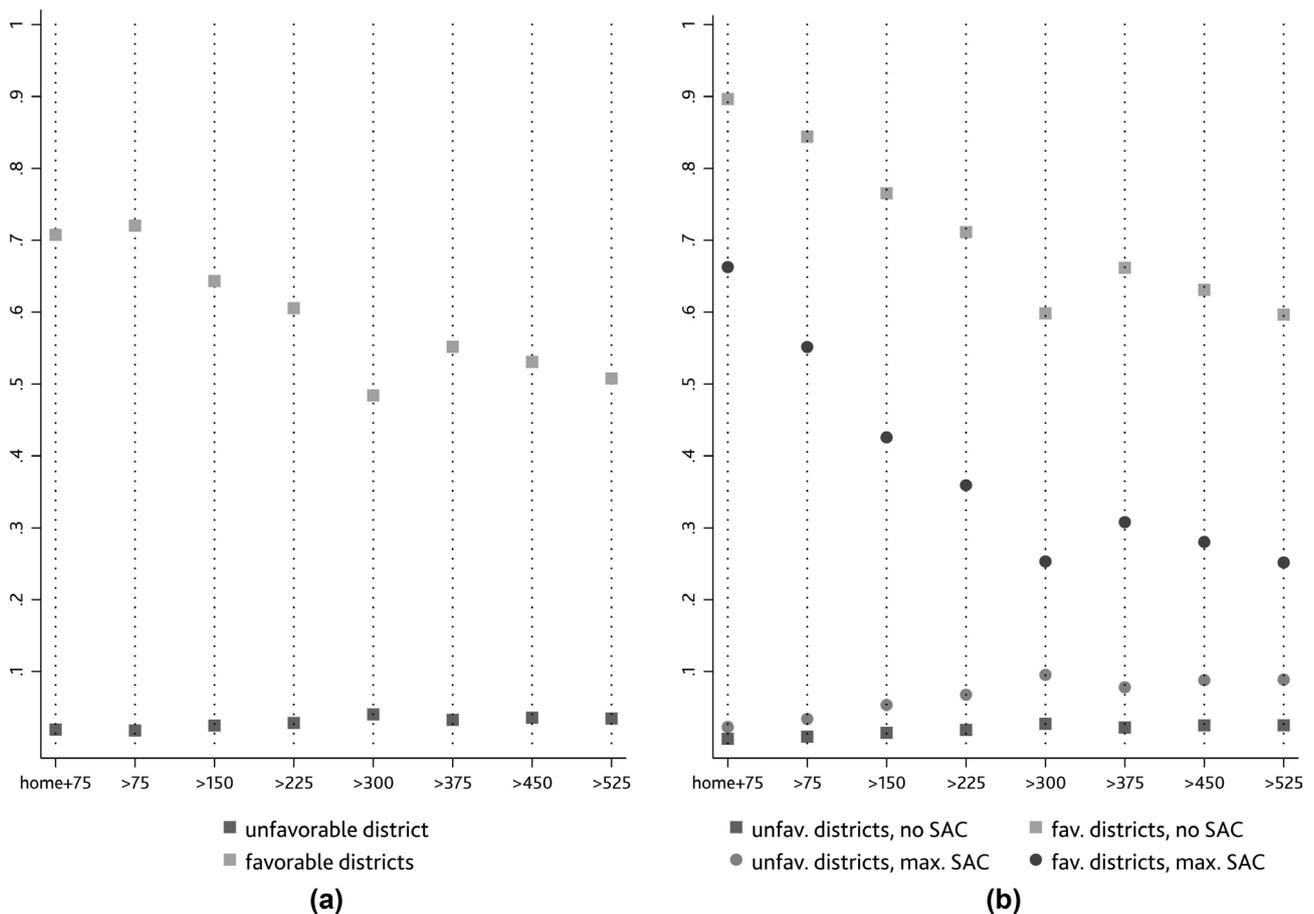
supply in non-local districts diminishes the individual probability of finding a spot. In structurally strong areas, the interaction effects are positive, as expected, but contrary to our expectations they increase with distance, and are only statistically significant for districts that are more than 300 kilometers away from home.

In models 1b and 2b, spatial autocorrelation is taken into account by controlling for commuting flows between home and workplace. The results show that the estimates in models 1a and 2a are considerably biased due to spatial autocorrelation, which arises due to commuting. In models 1b and 2b, the raw and relative supply of training positions in all zones has statistically significant effects, with the exception of the raw supply in districts located in structurally weak regions that are not more than 225 kilometers from home. Also, the coeffi-

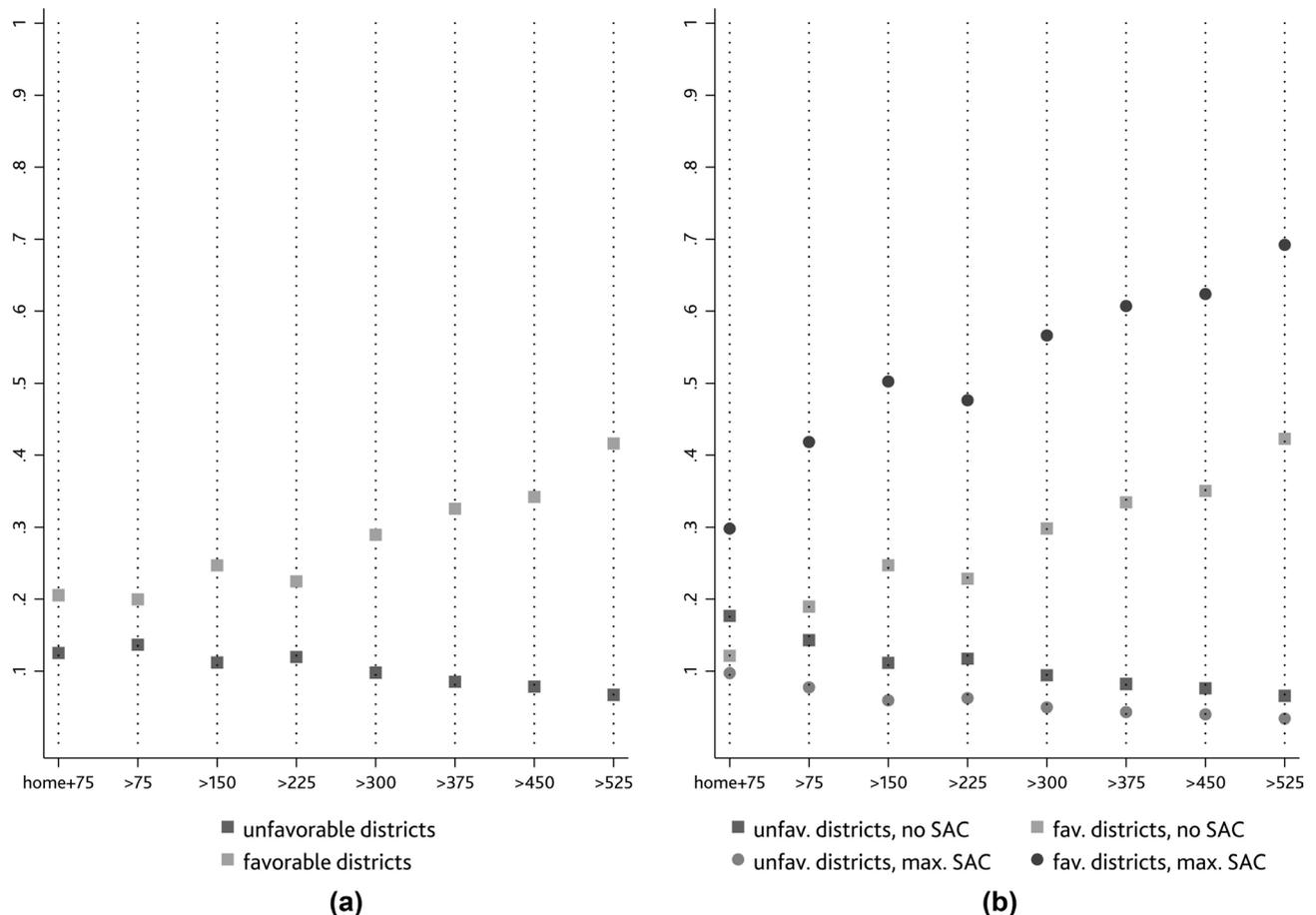
cients rise considerably, compared to models 1a and 2a. Two results are unexpected, though: For structurally weak regions, the interaction effects are still negative in model 2b, and for both regions the effects roughly increase with distance, a result that does not lend itself to a substantial interpretation. We will discuss these problems in the last section in detail.

All in all, our results show that both the raw and relative supply of training positions throughout Germany have an effect on the probability of finding a spot. To be able to verify these effects, it is necessary to control for spatial autocorrelation; neglecting to do so produces a significant underestimation of the effects of the supply of training spots in all non-local districts, and to false conclusions, specifically for areas relatively close to home.

The relevance of our methodical results can best be shown by means of probabilities instead of logit coefficients. **Figure 3** and **Figure 4** show these probabilities as a function of the raw and relative supply of training positions in the home district plus 75 kilometers, in areas that are 75 to 150 kilometers away from the home district, and so on. The other variables are held constant at the following values: male, individual search duration:  $\leq 4$  months; school graduation cohort: 2006; ethnic origin: German; school-completion qualification: higher



**Figure 3.** Predicted probabilities based on model 1a (not controlling for spatial autocorrelation, SAC) and model 1b (SAC fixed to 0), for structurally weak regions. (a) not controlling for SAC; (b) controlling for SAC.



**Figure 4.** Predicted probabilities based on model 2a (not controlling for spatial autocorrelation, SAC) and model 2b (SAC fixed to 0), for structurally strong regions. (a) not controlling for SAC; (b) controlling for SAC.

secondary education, no vocational preparation; states: Hamburg (structurally strong areas) and Schleswig-Holstein (structurally weak areas). We distinguish two scenarios: the probability of finding a training spot when a relevant district offers a high/low supply of training positions. A high supply is represented by a raw supply of 120, *i.e.*, a ratio of 1.2 between successfully concluded training contracts (plus vacant positions) and unsuccessful applicants (plus vacant positions), and a relative supply of 1.2, *i.e.*, the ratio between the supply of training places in the home district  $i$  and a given district  $j$ . For districts with a low supply, the values are 80 and 0.8, respectively.

**Figure 3** shows the predicted probabilities for structurally weak regions as calculated by models 1a and 1b, and **Figure 4** shows the predicted probabilities for structurally strong regions based on models 2a and 2b. In line with the presentation of our models, in **Figure 3(a)** and **Figure 4(a)** we do not control for spatial autocorrelation, and **Figure 3(b)** and **Figure 4(b)** take local spatial autocorrelation into account, *i.e.*, spatial autocorrelation between the home district and all other districts. We distinguish between two conditions: First, a spatial autocorrelation between each individual's home district and all other districts that equals zero, that is, in contrast to models 1a and 2a. Although there may be

commuting flows between other districts, they do not occur between the home district and the respective districts depicted (square markers). This condition is the most realistic in the sense that between three quarters of all districts there are no commuting flows (see **Table 1**). Second, the probabilities of maximum spatial autocorrelation are identified by round markers.

A comparison between figures a and b shows that neglecting spatial autocorrelation leads to biased estimates. In structurally weak regions, the predicted probabilities of finding a training position are underestimated in case of zero autocorrelation between a favorable home district and the respective non-local district. If autocorrelation is at its maximum, they are overestimated. For example, in **Figure 3(a)**, the probability for the home district plus a 75-kilometer radius is 70.7% for favorable districts. In contrast, in **Figure 3(b)**, this probability is 89.0% in case of zero autocorrelation (an underestimation of 18.3 percentage points), and 66.2% in case of maximum autocorrelation (an overestimation of 4.5 percentage points). The estimates for unfavorable districts are equally biased, but due to the overall small probability of finding a training spot in these districts, the differences between **Figure 3(a)** and **Figure 3(b)** are less obvious.

For structurally strong regions, the directions of over- and underestimation are exactly the reverse. For favorable districts, we find an overestimation in case of zero autocorrelation and an underestimation when autocorrelation is at its maximum (see **Figure 4(a)** and **Figure 4(b)**). Here, the most severe misestimation occurs for districts that are more than 525 kilometers from home: only  $-0.01$  percentage points in case of zero autocorrelation, but  $-28.2$  percentage points in case of maximal autocorrelation. Again, for unfavorable districts, the overall probability of finding a training position is so small that the differences between **Figure 4(a)** and **Figure 4(b)** are negligible.

Because we measured spatial autocorrelation by means of commuting flows in our analyses, our results can be interpreted with respect to the integration of districts into the (trainee) labor market as well. The findings presented here show that not controlling for spatial autocorrelation leads to an underestimation of spatial effects for a) isolated districts in structurally weak regions and b) well-integrated districts in structurally strong regions. In other words, if extreme conditions are obtained in a district, *i.e.*, a combination of regional structural weakness and isolation or regional structural strength and pronounced connection to other districts, contextual effects are underestimated if spatial autocorrelation is not controlled for. For “mixed” districts—those located in structurally weak but well-connected areas, and isolated districts located in structurally strong areas—contextual effects are overestimated. All this refers to favorable districts that offer a high raw and relative supply of training positions. For unfavorable districts, we find under- and overestimation to be mirror-inverted—*i.e.*, an underestimation in “mixed” districts and an overestimation in “extreme” districts—but the bias is small enough to be negligible.

Our results show that the differing patterns of probabilities in structurally

weak and strong regions are remarkable. In structurally weak regions—the five eastern states, Lower Saxony, and Schleswig-Holstein—finding a training position depends far more on local conditions than it does in structurally strong regions. We find the biggest difference between favorable and unfavorable districts for the home district (plus a 75-kilometer radius); with distance this gap closes (see **Figure 3(b)**). In structurally strong regions—the remaining western states and the city-states—the home district conditions are not decisive. The probabilities for favorable and unfavorable districts, respectively, with or without connections to other regions do not vary significantly (see **Figure 4(b)**). What matters more is whether regions further away offer training positions; if they do, the probability of finding a training spot increases markedly for youths living in favorable districts, resulting in an increasing gap between favorable and unfavorable districts. All in all, youths benefit more from an interconnection between districts in structurally strong regions.

With regard to our hypotheses, we were able to verify the assumed positive influence of the raw and relative supply of training positions in the home district and in all other German districts, but only for structurally strong regions. Contrary to our second assumption, however, those benefits gained from non-local districts do not decrease with increasing distance. Our findings for structurally weak regions contradicted the expectations in several aspects. We found effects of non-local districts on the probability of finding a training position; however, not all non-local districts have this effect. The relative supply of training places only has an effect for districts that are more than 225 kilometers away from home (relatedly, in structurally weak regions, the relative supply has an overall weaker effect than the raw supply, while in structurally strong regions the opposite is true). Additionally, the probability of finding a training spot decreases with an increasing supply in non-local districts. We will discuss all open questions in detail in the following section.

## 6. Discussion and Conclusions

In this paper we have examined the significance of the spatial opportunity structure for individual labor market outcomes by utilizing the example of youths' transitions from school to firm-based vocational training. When investigating opportunity structures, two methodological issues have to be addressed: the extent of such structures as well as the presence of spatial autocorrelation. Regarding the extent of opportunity structures, we must ask whether it is sufficient to only consider the local context. In our case, we have two reasons to assume that non-local conditions additionally affect the outcome. The first is that generally youths searching for a training position are spatially mobile: they are likely to accept offers that require commuting or even moving house. Second, local conditions themselves are a function of individual actions in adjacent areas, which in turn are affected by other areas, and so on. Such pulling effects on the contextual level point to the necessity of acknowledging that labor market op-

portunity structures are not only point-located. An individual opportunity structure does not end at the boundaries of an isolated regional container but, rather, extends across the landscape. From this perspective, opportunity structures correspond to a wider social system with unmarked boundaries. While our first argument, spatial mobility, is topic-specific and may not be applicable to other research, we believe that our second point is universally valid.

Our findings show that in the case of youths' transitions from school to firm-based vocational training, non-local labor market conditions matter. The probability of finding a training spot is affected not only by the supply of positions offered in the home district and its vicinity, but additionally by those presented in districts up to more than 500 kilometers from home. These effects, though, differ markedly depending on the structural strength of the wider region in which the home place is situated. In structurally strong regions—in our case, regions with a positive net trainee balance—the probability of finding an apprenticeship increases with an increasing supply of training spots in non-local districts, as expected. In structurally weak regions, we assumed that these effects would be weaker, but instead an increasing supply of positions in non-local districts is associated with a decreasing probability. The reason for this unexpected result has to be determined by future research. If we consider Windzio's findings [9] [44], we may assume that youths in structurally weak areas are perhaps less mobile and do not widen their search radius beyond the home district and adjacent areas. If, at the same time, training positions are heavily concentrated in a small number of cities, the different distributions of training spots and places of residence may lead to negative correlations such as those we have found. For the moment, we were able to establish that in structurally weak regions, finding a training position depends on local conditions to a far greater extent than in structurally strong regions.

Turning to our second methodological issue, our findings demonstrate that ignoring spatial autocorrelation leads to severely biased estimates. In the sample we analyzed, spatial autocorrelation does not occur at the individual level. We have no clustering of individual observations within districts or in adjacent or nearby districts, as is often the case with countrywide surveys. But at the context level of administrative districts, spatial interrelations lead to spatial autocorrelation, *i.e.*, to a homogenization of the value of the contextual predictor variables of strongly tied areal units. In order to map the actual extent of an individual's spatial opportunity structure, our regression results had to be adjusted for local spatial autocorrelation between the home district and all other districts, respectively. For this purpose, we used data on commuting flows between home and workplace and introduced into our models interaction terms between these commuting flows and our main predictor variables, *i.e.*, the raw and relative supply of training spots.

Our comparison showed that neglecting local spatial autocorrelation leads to considerably biased estimates especially for districts with a distance of up to 300 kilometers to the home district. For youth living in unfavorable districts with a

small raw and relative supply of training position offers, the probability of finding a spot is already so small that the bias is negligible. For favorable districts, however, not controlling for spatial autocorrelation is associated with over- or underestimation, depending on the wider region's structural strength and the level of connectedness between the home district and other districts. If either regional structural weakness and isolation or regional structural strength and pronounced connection to other districts are combined, contextual effects are underestimated; for "mixed" districts, *i.e.*, well-connected districts located in structurally weak areas, and isolated districts located in structurally strong areas, contextual effects are overestimated.

However, even though we controlled for spatial autocorrelation, our results give rise to the assumption that our estimations are still biased, because contrary to our second hypothesis, contextual effects for structurally strong areas do not decline with spatial distance. Three possible explanations are conceivable: First, in structurally strong regions indirect pulling effects on the macro level do not decrease with distance; second, they are stronger than the individual level effects of spatial mobility (which are still assumed to decrease with distance); and third, our measure for local spatial autocorrelation—apprentices' commuting flows between the home-district and all other districts respectively—does not entirely capture all spatial interdependencies in the data. There may be indirect pulling effects that result from unobserved processes of labor exchange between all districts in Germany. Such further reaching complex interrelations of districts have to be subject for future research about spatial effects on individual outcomes. Furthermore, the effects of remote districts may especially concern particular group(s) of people, like youths searching for training spots that are only available in a handful of locations in Germany. This may happen if industries or sectors are heavily concentrated, as is the case of biotechnology in southern Germany, North Rhine-Westphalia, and Berlin, or car manufacturing in southern and northern Germany. Lastly, for some occupational fields, the German training system offers only school-based training, so the next question may be whether these transitions are also structured by local and non-local conditions as well.

Our results show that for future analyses of spatial effects certain assumptions may be premature. First, assuming that only close proximity affects individual outcomes may disregard relevant contextual influences of the wider surroundings of the home, workplace, and so on, leading to biased estimates, because the characteristics of the wider surroundings may constitute essential control variables. It is also possible that contextual effects do not decrease with distance. For spatial models that require an a priori definition of the weights for spatial units, for example, the Spatial Durbin Model, it may therefore be erroneous to make a decision based on this assumption.

## References

- [1] Blossfeld, H.-P. (1986) Career Opportunities in the Federal Republic of Germany: A Dynamic Approach to the Study of Life-Course, Cohort, and Period Effects. *Euro-*

- pean Sociological Review*, **2**, 208-225.  
<https://doi.org/10.1093/oxfordjournals.esr.a036419>
- [2] Card, D. and Lemieux, T. (2001) Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics*, **116**, 705-746. <https://doi.org/10.1162/00335530151144140>
- [3] Kleinert, C. and Jacob, M. (2013) Demographic Changes, Labor Markets and Their Consequences on Post-School-Transitions in West Germany 1975-2005. *Research in Social Stratification and Mobility*, **32**, 65-83.  
<https://doi.org/10.1016/j.rssm.2013.01.004>
- [4] Abott, A. (1997) Of Time and Space: The Contemporary Relevance of the Chicago School. *Social Forces*, **75**, 1149-1182. <https://doi.org/10.1093/sf/75.4.1149>
- [5] Goodchild, M.F., Anselin, L., Appelbaum, R.P. and Harthorn, B.H. (2000) Toward Spatially Integrated Social Science. *International Regional Science Review*, **23**, 139-159. <https://doi.org/10.1177/016001760002300201>
- [6] Arntz, M. and Wilke, R.A. (2009) Unemployment Duration in Germany: Individual and Regional Determinants of Local Job Finding, Migration and Subsidized Employment. *Regional Studies*, **43**, 43-61. <https://doi.org/10.1080/00343400701654145>
- [7] Furlong, A., Biggart, A. and Cartmel, F. (1996) Neighbourhoods, Opportunity Structures and Occupational Aspirations. *Sociology*, **30**, 551-565.  
<https://doi.org/10.1177/0038038596030003008>
- [8] Gross, C. and Kriwy, P. (2013) The Influence of Regional Social Inequality and Labour Market Characteristics on Health. *Comparative Population Studies*, **38**, 769-794.
- [9] Windzio, M. (2008) The “Exit Option” of Labour Migration from East to West Germany: Individual and Contextual Determinants of Geographic Mobility of Unemployed Workers. In: Kolb, H. and Egbert, H., Eds., *Migrants and Markets: Perspectives from Economics and Other Social Sciences*, Amsterdam University Press, Amsterdam, 56-80.
- [10] Anselin, L. (2003) Spatial Externalities, Spatial Multipliers, and Spatial Econometrics. *International Regional Science Review*, **26**, 153-166.  
<https://doi.org/10.1177/0160017602250972>
- [11] Logan, J.R. (2012) Making a Place for Space. Spatial Thinking in Social Science. *Annual Review of Sociology*, **38**, 507-524.  
<https://doi.org/10.1146/annurev-soc-071811-145531>
- [12] Weßling, K., Hartung, A. and Hillmert, S. (2015) Spatial Structure Counts: The Relevance of Regional Labour-Market Conditions for Educational Transitions to Vocational Training. *Empirical Research in Vocational Education and Training*, **7**, 12.
- [13] Wong, D. (2009) The Modifiable Areal Unit Problem. In: Fotheringham, A.S. and Rogerson, P.A., Eds., *The SAGE Handbook of Spatial Analysis*, SAGE Publications, London, 105-124. <https://doi.org/10.1016/B978-008044910-4.00475-2>
- [14] Anselin, L. (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis*, **27**, 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- [15] Ward, M.D. and Gleditsch, K.S. (2008) Spatial Regression Models. Series: Quantitative Applications in the Social Science, SAGE Publications, Inc., Thousand Oaks.  
<https://doi.org/10.4135/9781412985888>
- [16] Fujita, M., Krugman, P. and Venables, A.J. (1999) The Spatial Economy-Cities, Regions, and International Trade. MIT Press, Cambridge.
- [17] Akerlof, G.A. (1997) Social Distance and Social Decisions. *Econometrica*, **65**,

1005-1027.

- [18] Goodchild, M.F. and Janelle, D.G. (2004) *Spatially Integrated Social Science*. Oxford University Press, New York.
- [19] Kropp, P. and Schwengler, B. (2016) Three-Step Method for Delineating Functional Labour Market Regions. *Regional Studies*, **50**, 429-445. <https://doi.org/10.1080/00343404.2014.923093>
- [20] Venhorst, V., Van Duk, J. and Van Wissen, L. (2011) An Analysis of Trends in Spatial Mobility of Dutch Graduates. *Spatial Economic Analysis*, **6**, 57-82. <https://doi.org/10.1080/17421772.2010.540033>
- [21] Lemistre, P. and Moreau, N. (2009) Spatial Mobility and Returns to Education: Some Evidence from a Sample of French Youth. *Journal of Regional Science*, **49**, 149-176. <https://doi.org/10.1111/j.1467-9787.2008.00574.x>
- [22] Bogai, D., Seibert, H. and Wiethölter, D. (2008) Duale Ausbildung in Deutschland. Die Suche nach Lehrstellen macht junge Menschen mobil. IAB Kurzbericht, 9.
- [23] Coleman, J.S. (1986) Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology*, **91**, 1309-1335. <https://doi.org/10.1086/228423>
- [24] Cadwallader, M. (1989) A Conceptual Framework for Analysing Migration Behaviour in the Developed World. *Progress in Human Geography*, **13**, 494-511. <https://doi.org/10.1177/030913258901300402>
- [25] Esser, H. (1998) Why Are Bridge Hypotheses Necessary? In: Blossfeld, H.-P. and Prein, G., Eds., *Rational Choice Theory and Large-Scale Data Analysis*, Westview Press, Boulder, 94-111.
- [26] Lemistre, P. and Magrini, M.-B. (2010) Job Qualification, Distance between Towns and Geographical Relocation for French Youth. *Urban Studies*, **48**, 2141-2161. <https://doi.org/10.1177/0042098010382675>
- [27] Sjaastad, L.A. (1962) The Costs and Returns of Human Migration. *Journal of Political Economy*, **70**, 80-93. <https://doi.org/10.1086/258726>
- [28] Decressin, J.W. (1994) Internal Migration in West Germany and Implications for East-West Salary Convergence. *Weltwirtschaftliches Archiv*, **130**, 231-257. <https://doi.org/10.1007/BF02707708>
- [29] Détang-Dessendre, C. (1999) Reciprocal Link between Exit from Unemployment and Geographical Mobility. *Environment and Planning A*, **31**, 1417-1431. <https://doi.org/10.1068/a311417>
- [30] Fielding, A.J. (1992) Migration and Social Mobility: South East England as an Escalator Region. *Regional Studies*, **26**, 1-15. <https://doi.org/10.1080/00343409212331346741>
- [31] Wicht, A. (2016) Regional Contexts in Quantitative Educational Sociology. In: Million, A., Heinrich, A.J. and Coelen, T., Eds., *Education, Space and Urban Planning: Education as a Component of the City*, Springer International Publishing, Luxembourg, 299-307.
- [32] Granato, N., Haas, A., Hamann, S. and Niebuhr, A. (2015) The Impact of Skill-Specific Migration on Regional Unemployment Disparities in Germany. *Journal of Regional Science*, **55**, 513-539. <https://doi.org/10.1111/jors.12178>
- [33] Haining, R. (2009) The Special Nature of Spatial Data. In: Fotheringham, A.S. and Rogerson, P.A., Eds., *The SAGE Handbook of Spatial Analysis*, SAGE Publications, London, 5-23. <https://doi.org/10.4135/9780857020130.n2>
- [34] Goldstein, H. (2011) *Multilevel Statistical Models*. Wiley & Sons, Chichester.
- [35] Rasbash, J. and Browne, W. (2008) Non-Hierarchical Multilevel Models. In: de

- Leeuw, J. and Meijer, E., Eds., *Handbook of Multilevel Analysis*, Springer, New York. [https://doi.org/10.1007/978-0-387-73186-5\\_8](https://doi.org/10.1007/978-0-387-73186-5_8)
- [36] Chaix, B., Merlo, J. and Chauvin, P. (2005) Comparison of a Spatial Approach with the Multilevel Approach for Investigating Place Effects on Health: The Example of Healthcare Utilisation in France. *Journal of Epidemiology & Community Health*, **59**, 517-526. <https://doi.org/10.1136/jech.2004.025478>
- [37] Sampson, R.J. and Raudenbush, S.W. (1999) Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods. *American Journal of Sociology*, **105**, 603-651. <https://doi.org/10.1086/210356>
- [38] Xu, H. (2014) Comparing Spatial and Multilevel Regression Models for Binary Outcomes in Neighborhood Studies. *Sociological Methodology*, **44**, 229-272. <https://doi.org/10.1177/0081175013490188>
- [39] Tobler, W.R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234-240. <https://doi.org/10.2307/143141>
- [40] Shen, Q. (2004) Updating Spatial Perspectives and Analytical Frameworks in Urban Research. In: Goodchild, M.F. and Janelle, D.G., Eds., *Spatially Integrated Social Science*, Oxford University Press, New York, 263-279.
- [41] Blossfeld, H.-P., Roßbach, H.-G. and Maurice, J.V. (2011) Education as a Lifelong Process. The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft*, No. 14. <https://doi.org/10.1007/s11618-011-0179-2>
- [42] Kleinert, C., Matthes, B., Antoni, M., Drasch, K., Ruland, M. and Trahms, A. (2011) ALWA New Life Course Data for Germany. *Schmollers Jahrbuch*, **131**, 625-634. <https://doi.org/10.3790/schm.131.4.625>
- [43] IAB (2012) Beschäftigtenhistorik (BEH), V08.07.00, Nürnberg.
- [44] Windzio, M. (2004) Kann der regionale Kontext zur Arbeitslosenfrage werden? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, **56**, 257-278. <https://doi.org/10.1007/s11577-004-0034-z>
- [45] Allison, P.D. (1982) Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology*, **13**, 61-98. <https://doi.org/10.2307/270718>
- [46] Yamaguchi, K. (1991) Event History Analysis. Applied Social Research Methods Series, Vol. 28, SAGE Publications, Newbury Park.
- [47] Blossfeld, H.-P. and Rohwer, G. (2002) Techniques of Event History Modeling: New Approaches to Casual Analysis. Lawrence Erlbaum Associates, Mahwah.

## Appendix

**Table A1.** Individual and spatial effects on youths' transitions from school to firm-based vocational training, Germany, 1999-2006, discrete-time event history models, standardized logit coefficients. (a) (structurally weak regions); (b) (structurally strong regions).

(a)				
	<i>Model 1a:</i>		<i>Model 1b:</i>	
	Not controlling for commuting flows		Controlling for commuting flows	
	$\beta$	s.e.	$\beta$	s.e.
Search time (ref. 0 - 4 months)				
5 - 16 months	-1.332***	0.001	-1.327***	0.001
17 - 28 months	-2.741***	0.001	-2.718***	0.002
>28 months	-2.917***	0.002	-2.881***	0.004
Female	0.133***	0.001	0.126***	0.001
Native language not German	0.243***	0.001	0.256***	0.002
Educational level (ref. higher secondary education <sup>a</sup> )				
Secondary education <sup>b</sup>	-0.142***	0.001	-0.151***	0.002
Basic secondary education <sup>c</sup>	-1.214***	0.001	-1.240***	0.003
Vocational preparation	2.254***	0.001	2.257**	0.001
Cohort (ref. 1999)				
2000	0.149***	0.001	0.141***	0.002
2001	-0.175***	0.001	-0.175***	0.002
2002	-0.083***	0.002	-0.089***	0.003
2003	0.383***	0.003	0.391***	0.003
2004	-0.071***	0.004	-0.049***	0.004
2005	-0.261***	0.003	-0.244***	0.004
2006	-0.343***	0.003	-0.335***	0.005
State (ref. Schleswig-Holstein)				
Lower Saxony	0.754***	0.003	0.751***	0.003
Rhineland-Palatinate	1.418***	0.005	1.436***	0.005
Brandenburg	1.500***	0.006	1.543***	0.008
Mecklenburg-Western Pomerania	2.431***	0.006	2.492***	0.008
Saxony	0.962***	0.006	0.973***	0.007
Saxony-Anhalt	1.086***	0.006	1.108***	0.006
Thuringia	1.027***	0.006	1.059***	0.006
Raw supply of training places	0.363***	0.032	0.333***	0.038
Relative supply of training places	0.405***	0.045	0.376***	0.059
Distance (ref. home district + 75 km)				
>75 - 150 km	0.006	0.030	0.021	0.036

**Continued**

>150 - 225 km	0.008	0.030	0.033	0.037
>225 - 300 km	0.008	0.027	0.038	0.035
>300 - 375 km	-0.031	0.025	-0.001	0.034
>375 - 450 km	-0.019	0.027	0.004	0.035
>450 - 525 km	-0.013	0.027	0.009	0.036
>525 km	-0.075**	0.028	-0.055	0.036
Raw supply × distance (ref.: home district + 75 km)				
>75 - 150 km	-0.047	0.038	-0.128**	0.047
>150 - 225 km	-0.085*	0.039	-0.187***	0.048
>225 - 300 km	-0.107**	0.037	-0.224***	0.049
>300 - 375 km	-0.163***	0.039	-0.281***	0.050
>375 - 450 km	-0.112**	0.039	-0.229***	0.050
>450 - 525 km	-0.111**	0.037	-0.238***	0.048
>525 km	-0.113**	0.038	-0.249***	0.049
Relative supply × distance (ref.: home district + 75 km)				
>75 - 150 km	0.092	0.049	0.011	0.057
>150 - 225 km	0.017	0.048	-0.092	0.058
>225 - 300 km	-0.011	0.048	-0.141*	0.058
>300 - 375 km	-0.094*	0.048	-0.227***	0.058
>375 - 450 km	-0.074	0.048	-0.206***	0.057
>450 - 525 km	-0.110*	0.049	-0.244***	0.057
>525 km	-0.118*	0.047	-0.259***	0.057
Commuting flows (ref.: >1%, incl. home district)				
≤1%, >0.5%	-	-	-0.006	0.062
≤0.5%, >0.2%	-	-	0.077	0.045
≤0.2%, >0.1%	-	-	-0.008	0.043
≤0.1%, 0.05%	-	-	-0.165***	0.044
≤0.05%, >0%	-	-	0.102*	0.044
0%	-	-	-0.022	0.040
Raw supply × commuting (ref.: >1%, incl. home district)				
≤1%, >0.5%	-	-	0.120	0.080
≤0.5%, >0.2%	-	-	0.128*	0.058
≤0.2%, >0.1%	-	-	0.071	0.058
≤0.1%, 0.05%	-	-	0.002	0.055
≤0.05%, >0%	-	-	-0.080	0.056
0%	-	-	0.216***	0.051

## Continued

Relative supply × commuting (ref.: >1%, incl. home district)				
≤1%, >0.5%	–	–	0.110	0.095
≤0.5%, >0.2%	–	–	0.117	0.064
≤0.2%, >0.1%	–	–	0.114	0.067
≤0.1%, 0.05%	–	–	–0.026	0.063
≤0.05%, >0%	–	–	0.032	0.064
0%	–	–	0.232***	0.060
Constant term	–1.687***	0.023	1.696***	0.029

Notes: \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001 (two-tailed tests), cluster-robust S.E. The values of the raw and relative supply of training places are standardized; the S.E. of individual level effects are not interpretable due to the data structure. <sup>a</sup> *Abitur/Fachabitur*, <sup>b</sup> *Mittlere Reife/Realschulabschluss*, <sup>c</sup> *Hauptschulabschluss*.

(b)

	<i>Model 2a:</i> Not controlling for commuting flows		<i>Model 2b:</i> Controlling for commuting flows	
	$\beta$	s.e.	$\beta$	s.e.
Search time (ref. 0 - 4 months)				
5 - 16 months	–1.316***	0.000	–1.313***	0.001
17 - 28 months	–1.599***	0.001	–1.596***	0.001
>28 months	–2.234***	0.002	–2.229***	0.002
Female	0.141***	0.000	0.139***	0.001
Native language not German	–0.643***	0.001	–0.638***	0.001
Educational level (ref. higher secondary education <sup>a</sup> )				
Secondary education <sup>b</sup>	0.404***	0.001	0.403***	0.001
Basic secondary education <sup>c</sup>	–0.486***	0.001	–0.493***	0.001
Vocational preparation	1.025***	0.001	1.029***	0.002
Cohort (ref. 1999)				
2000	–0.705***	0.001	–0.706***	0.002
2001	–0.548***	0.002	–0.545***	0.002
2002	–0.305***	0.001	–0.301***	0.002
2003	–0.639***	0.002	–0.634***	0.003
2004	–0.545***	0.003	–0.542***	0.003
2005	–0.701***	0.003	–0.694***	0.003
2006	–0.150***	0.003	–0.144***	0.003
State (ref. Hamburg)				
North Rhine-Westfalia	1.066***	0.003	1.070***	0.004
Hesse	0.785***	0.004	0.780***	0.005
Baden-Wuerttemberg	0.997***	0.004	1.000***	0.005
Bavaria	0.812***	0.004	0.820***	0.005

**Continued**

Saarland	2.073***	0.004	2.070***	0.005
Berlin	0.727***	0.006	0.779***	0.013
Raw supply of training places	0.065***	0.013	0.098***	0.019
Relative supply of training places	0.021	0.017	0.124***	0.030
Distance (ref. home district + 75 km)				
>75 - 150 km	0.035	0.020	0.113***	0.021
>150 - 225 km	0.042*	0.017	0.119***	0.022
>225 - 300 km	0.027	0.018	0.104***	0.022
>300 - 375 km	0.065***	0.018	0.142***	0.022
>375 - 450 km	0.065***	0.017	0.142***	0.022
>450 - 525 km	0.051**	0.016	0.129**	0.021
>525 km	0.109***	0.017	0.189***	0.022
Raw supply × distance (ref.: home district + 75 km)				
>75 - 150 km	-0.000	0.019	0.058*	0.024
>150 - 225 km	0.034	0.018	0.101***	0.025
>225 - 300 km	0.001	0.019	0.072**	0.025
>300 - 375 km	0.049**	0.017	0.121***	0.024
>375 - 450 km	0.063***	0.017	0.136***	0.025
>450 - 525 km	0.083***	0.016	0.156***	0.023
>525 km	0.108***	0.017	0.179***	0.024
Relative supply × distance (ref.: home district + 75 km)				
>75 - 150 km	-0.026	0.024	0.067*	0.028
>150 - 225 km	0.021	0.021	0.124***	0.027
>225 - 300 km	0.030	0.022	0.136***	0.027
>300 - 375 km	0.071***	0.020	0.181***	0.026
>375 - 450 km	0.113***	0.020	0.222***	0.026
>450 - 525 km	0.115***	0.021	0.223***	0.026
>525 km	0.172***	0.018	0.278***	0.025
Commuting flows (ref.: >1%, incl. home district)				
≤1%, >0.5%	-	-	-0.113**	0.038
≤0.5%, >0.2%	-	-	-0.095**	0.033
≤0.2%, >0.1%	-	-	-0.142***	0.030
≤0.1%, 0.05%	-	-	-0.176***	0.031
≤0.05%, >0%	-	-	-0.351***	0.031
0%	-	-	-0.166***	0.029
Raw supply × commuting (ref.: >1%, incl. home district)				
≤1%, >0.5%	-	-	0.003	0.039

**Continued**

≤0.5%, >0.2%	–	–	0.033	0.036
≤0.2%, >0.1%	–	–	–0.027	0.033
≤0.1%, 0.05%	–	–	–0.103**	0.035
≤0.05%, >0%	–	–	–0.272***	0.036
0%	–	–	–0.097***	0.030
Relative supply × commuting (ref.: >1%, incl. home district)				
≤1%, >0.5%	–	–	–0.042	0.054
≤0.5%, >0.2%	–	–	–0.052	0.042
≤0.2%, >0.1%	–	–	–0.097**	0.037
≤0.1%, 0.05%	–	–	–0.178***	0.037
≤0.05%, >0%	–	–	–0.329***	0.039
0%	–	–	–0.209***	0.036
Constant term	–1.674***	0.015	–1.588***	0.023

*Notes:* \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001 (two-tailed tests), cluster-robust S.E. The values of the raw and relative supply of training places are standardized; the S.E. of individual level effects are not interpretable due to the data structure. <sup>a</sup>*Abitur/Fachabitur*, <sup>b</sup>*Mittlere Reife/Realschulabschluss*, <sup>c</sup>*Hauptschulabschluss*