Match Me Up Before I Go-Go!—Matching Functions for Spatially Connected VET Labor Markets

Skills2Capabilities Working Paper
November 2025

Dennis Oliver Kubitza
Federal Institute for Vocational Education and Training (BIBB),
Research Centre for Education and the Labour Market (ROA)
Maastricht University

ABSTRACT

Local labor markets in Germany are interconnected through mobile workers who commute or migrate across borders. Therefore, the aggregated number of new matches between employers and employees in one region is influenced by spillovers from neighboring markets, but might also be disproportionately higher due to easier access to suitable jobs in agglomerated areas. Vocational education and training (VET) students in company training, so called apprentices, are younger and therefore less mobile than the general workforce, suggesting that regional spillovers may be less pronounced when looking at VET labor markets. In addition, occupational heterogeneities in the spatial distribution of employers might further amplify or decrease these spillovers. This paper provides the first estimates of spatial matching functions for VET students in general and for multiple profession groups. Using a matching function model with spatially lagged stocks of applicants and vacancies, it analyses efficiency, elasticities, regional influences, and spatial spillovers. To do so, it develops a novel set of spatial weights and ways to measure agglomeration. The findings show that matching efficiency for VET is higher in more agglomerated regions and that measurable spillovers do exist, even for the group of less long-distance mobile adolescents. This underscores the importance of accurate spatial modeling: only indices based on realistic travel times or commuting data show these influences. Finally, this paper finds significantly lower elasticities in the stock of unemployed applicants for regions that are well connected to their surroundings, leading to a more balanced matching process.

ACKNOWLEDGEMENTS

This research was funded by The European Commission under the Horizon Europe Program (Skills2Capabilities; S2C).

Skills2Capabilities, a Horizon Europe study, is about understanding how skills systems need to develop if they are to assist people to make labour market transitions – i.e. between jobs, employers or sectors – and thereby reduce the level of skill mismatch which might otherwise arise.

This Working Paper is part of the Skills2Capabilitiy Work Package 4 entitled 'Challenges and changes in the demand for VET Skills'

For more information please visit skills2capabilities.eu

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Agency. Neither the European Union nor the granting authority can be held responsible for them.

Skills2Capabilities

Co-funded by the European Union

1. Introduction

Assessing how well the labor market works is often done by measuring and interpreting quotas and ratios based on aggregated vacancies, unemployment figures, and new matches. A formal tool for this is the aggregate matching function, which is based on microeconomic search and sorting models (Petrongolo and Pissarides 2001; Chade, Eeckhout, and Smith 2017). Contrary to the perspective of analyzing matching success or quality for an individual, this function models the overall number of matches within a regional labor market as a nonlinear function of two key variables: the stock of vacancies and the stock of unemployment or applicants[1]. Thus, aggregate matching functions enable a model-based analysis of matching efficiency (e.g., the proportions of the vacancy and applicant stocks matched within a period) and of labor market elasticities, which measure the responsiveness of new matches to changes in the vacancy and applicant stocks.

While some labor market processes are invariant to space and place, the job search process certainly is not, especially for the individual. Therefore, it is necessary to account for these factors when estimating performance indicators on spatial units. Additionally, regional labor markets are heterogeneous and therefore not directly comparable. On the demand side, economic trends and locational advantages shape the jobs required within companies. On the supply side, factors such as the local economic situation (Weßling, Hartung, and Hillmert 2015), industry composition (Malin and Jacob 2019; Barbour and Markusen 2007), and urbanity (Schmidt and Uhly 2023) influence the career aspirations of students and, consequently, the long-term composition of the local workforce.

When workers do not find matching jobs in one place, they are forced to become mobile, by changing their careers, migrating or commuting. The share of cross-border commuters and their average commuting distance increased in European industrial nations such as the United Kingdom and Germany (Galvin and Madlener 2014; Lyons and Chatterjee 2008) in the medium and long term. However, COVID-19 quarantines and digitalization introduced remote work for entire industries (Haas, Faber, and Hamersma 2020; Bick, Blandin, and Mertens 2021), reducing job-related commuting.

Since Petrongolo and Pissarides (2001), who provided the first literature survey on the matching function framework, several studies on regional and spatial matching functions have emerged with national focuses (Ilmakunnas and Pesola 2003; Hynninen 2005; Moilanen 2010; Lee 2019; Antczak, Gałecka-Burdziak, and Pater 2018; Mohamedou 2022; Morales et al. 2025). Works by Haller and Heuermann (2016), Fahr and Sunde (2006a, 2006b) and Lottmann (2012) analyze spatial aspects, spillovers and regional influences for the general German labor market in a matching function framework. Fedorets, Lottmann, and Stops (2019) provide the first work that considers both spatial and occupational segmentation for Germany with this tool. By modeling occupational alongside

[1] As the target population in this study will be job-starters, the term applicants is more suitable and will be used to refer to the stock variable in the rest of this work. Also, unemployment will occur later as a regional control variable

Skills2Capabilities
Co-funded by the European Union

spatial spillovers in a fixed-effects model, they show detailed inter-dependencies of professions and regions.

The German Vocational Education and Training (VET) system regulates and defines the processes for obtaining a certification in 328 protected professions and is part of upper secondary education. VET education takes 2-3 years and is typically started by adolescents between the ages of 16 and 20 years who have finished schooling in the general education system. In its predominant form, known as the dual system, it combines theoretical instruction at a vocational school with practical in-company training (apprenticeship) under a training contract. As companies need to advertise open positions for which potential apprentices apply, the VET market is a subset of the labor market, characterized by a younger, less experienced workforce.

Since apprentices exhibit lower long distance mobility than entrants to the general labor market, I expect similar regional influences, but with different magnitudes. This particularly affects the spillovers between labor markets observed in Germany (Lottmann 2012; Fahr and Sunde 2006b; Haller and Heuermann 2016; Fedorets, Lottmann, and Stops 2019), but might also amplify or offset other regional dynamics. This paper specifically investigates how, and to what extent, agglomeration and spatial spillovers impact the regional labor markets in matching employers with apprentices.

In doing so, this paper is the first one to estimate spatial matching functions for VET professions in Germany, considering both the overall VET market and five major occupational groups[2]. Thus, this paper primarily contributes to the literature on regional labor markets and school-to-work transitions.

Beyond estimating these regional influences, the study introduces novel methodological approaches to capture spillovers and agglomeration influences. Specifically, it develops indices and spillover matrices based on population-weighted public transport accessibility between the largest economic centers of each region, both within regions and across their borders. These measures of agglomeration and spillovers are evaluated separately and, in a second step, jointly against existing approaches for modeling such concepts.

The analysis and estimation in this paper are based on register data, which include newly concluded contracts, open positions and unmatched VET applicants for Germany's 148 labor agency districts (Bundesinstitut für Berufsbildung (BIBB) 2023). For spatial and regional integration, I use map data of these labor agency districts to compute neighboring regions and to match included entities, like administrative districts or cities. In addition, this study uses multiple additional data sources: regional data from the INKAR (Indikatoren und Karten zur Raum- und Stadtentwicklung) database provided by the BBSR (Federal Institute for Research on Building, Urban Affairs and Spatial Development) (BBSR 2024); Environmental and land usage statistics from the IÖR Monitor (Monitor of Settlement and Open Space Development) (Krüger, Meinel, and

[2] This study uses the German Classification of Occupations: Klassifikation der Berufe (KLDB), which assigns each occupation a five-digit code based on hierarchical clustering. Occupational groups are defined by the first two digits of these codes.

Schumacher 2013); the exact location and population of cities or municipalities for each labor market region, queried from Wikidata[3]; and finally public transport and car traveling times between them, web-scraped from the Google Distance Matrix API[4].

In combining these different data sources and estimating spatial matching functions, this study serves as an example for the application of spatial econometric modeling.

The results provide the first empirical evidence on the role of spatial spillovers and agglomeration in the German VET market. By focusing on this unique, low-mobility group, this paper reveals significant effects that have been previously undocumented. The comparison between indices for agglomeration and spillover weights show the critical importance of spatial model specification when analyzing less-mobile populations. I show that for apprentices, simplistic measures for measuring spillovers (like centroid distance) and agglomeration (like population density) are inadequate. These measures mask potential influences, leading to the erroneous conclusion that spillovers or agglomeration effects are absent.

While the results are not directly causal, as they estimate formal model parameters of a matching model, the study indicates that agglomeration is associated with higher matching efficiency, while spatial connectivity creates a more balanced matching process by reducing dependency on local applicant pools. I also find that these influences vary considerably across different occupational groups.

[3] **Wikidata SPARQL Endpoint** A query service for retrieving structured data from Wikidata using the SPARQL (SPARQL Protocol and RDF Query Language) query language (Documentation).
[4] 4**Google Distance Matrix API** A service that calculates travel distances and durations between locations for different modes of transportation (Documentation).

Skills2Capabilities
Co-funded by the European Union

2. VET and transitions to dual VET

VET plays a substantial role in the German education system, serving as both a part of secondary education and a transitional phase into the labor market. Education in the VET system takes 2 to 3 years and completing VET leads to a certification in one of 328 recognized occupations (Bundesinstitut für Berufsbildung 2024), providing the formal qualification to practice that profession.

While everyone may start training in the VET system, many job-starters without other professional qualifications seek to start a VET. This is most commonly the case for students after completing secondary schooling either at the first stage (9 - 10 schooling years) or after finishing it with a university entrance certificate (12 - 13 schooling years).

In Germany, approximately 60 percent of an age cohort obtain a professional qualification in the VET system (BMBF 2024). While some chose to participate in school-based VET, which is offered predominantly for health and education occupations, most VET students are trained in the apprenticeship system directly at a company (next to a fixed share of school-based lectures) and thus need a company contract for starting their training. In this dual system the matching between employers and apprentices follows labor market mechanisms. Students apply for open positions, and employers can choose between the pool of applicants. This study focuses on the aggregated outcome of this process: the number of newly signed VET contracts.

Approximately 490,000 new VET contracts were established in 2023, with 73,000 positions remaining unfilled and 26,000 VET students not able to secure a position before the start of the school year[5]. The two drivers of this discrepancy are missing VET applicants and frictions that prevent available applicants from filling available positions. It is important to note, that these numbers only cover positions and students registered with the unemployment agencies and not those who decide to remain unregistered. Therefore, the actual mismatch might even be higher.

[5] These numbers differ slightly from the official statistics, as they differentiate between two types of applicants: Those who also search for alternatives and those who do not. Some regions only list one of the two types, such that an imputation of the missing values raises the numbers.

Personal factors are known to influence the matching process, most notably prior educational attainment. The German education system offers different school-leaving certificates: The lower-level secondary education diploma (9th grade), the intermediate-level secondary education diploma (10th grade), and the upper-level secondary education diploma (12th or 13th grade) differ in the opportunities they provide. The latter grants access to higher education institutions upon completion of school. In theory, anyone can apply for VET positions; however, companies often require or desire specific skills and qualifications for entering professions, creating competition among applicants. Although holders of the higher-level diploma are eligible for higher education, at least 21 percent still enter the VET system (BMBF 2024). Fitzenberger et al. (2025) are so far the only authors to estimate matching functions for the VET market. Their study analyzes the mechanisms driving VET matches, focusing on the evolution of applicants, vacancies, and matching efficiency in Germany between 2013 and 2021. They show that the decline in cohort sizes alone would have exacerbated the decline in matches, but increases in migrant applicants and favorable shifts in school leaver application patterns helped stabilize matches before the COVID pandemic. During COVID, simultaneous declines in applicants, vacancies, and matching efficiency amplified the drop in matches. While Fitzenberger et al. (2025) also examine local labor market tightness and highlight regional differences, they find only minimal effects from regional mismatch and explicitly refrain from jointly analyzing matching across neighboring districts.

As the job search and matching process is not invariant of space and place, the VET market surely will also be influenced by them. However, apprentices exhibit lower long distance mobility than entrants to the general labor market. This raises the question whether regional influences, particularly spillovers, are less pronounced or if they manifest in different ways.

The German VET market is characterized by significant regional disparities. Some regions serve as training hubs and attract inward commuting from neighboring areas (Jost 2019). Employers in rural areas, which are being abandoned by young, well-educated individuals, struggle to find skilled workers (Rodríguez-Pose, Dijkstra, and Poelman 2024). Figure 1 shows the uneven distribution of the regional deficit. The relative applicant deficit per VET position is visualized in colors, while the positions and size of the five largest cities per district are marked. It shows that particularly in the south and east of Germany, the lack is far more severe than in other parts. In areas with a concentration of major cities, the trend is reversed and open VET positions are missing. Schmidt and Uhly (2023) further support this, claiming that the urbanity of a region influences the professional aspirations of adolescents.

In response to these discrepancies, mobility becomes a crucial adjustment mechanism. Jost (2019) utilizes commuting data and regional statistics to analyze how discrepancies induce commuting patterns. They show that larger companies offer disproportionately more VET positions and that such a surplus of supply induces commuting. They also show that mobility readiness differs by the type of secondary education students obtained. Higher qualified students are more mobile, and the STEM sector shows the highest commuting distances and frequencies.

Underlying these spatial patterns are distinct regional supply and demand forces. On the demand side, Wolter and Ryan (2011) demonstrate that companies expect demographic trends and thus adjust the number of offered positions. On the supply side, local cohort size (Heineck 2011; Hillmert, Hartung, and Weßling 2017) and unemployment rates from neighboring regions, as well as, local unemployment (Weßling, Hartung, and Hillmert 2015; Hillmert, Hartung, and Weßling 2017) significantly influence adolescents' transition probabilities to VET. Researchers often use these contextual indicators such as dominant industry structures (Flohr, Menze, and Protsch 2020) or local unemployment rates (Ebner 2015) to capture how regional conditions influence supply and demand for positions in VET labor markets. Despite these hard factors, social contexts directly influence the decision to start VET, as societal values determine participation patterns, especially for women (Malin and Jacob 2019) and regarding socio-economic status (Hartung, Weßling, and Hillmert 2022).

These regional conditions interact with individual characteristics, especially the type of school leaving certificate. While the type of diploma directly influences career aspirations, its impact is amplified by regional context (Wicht and Ludwig-Mayerhofer 2014). Research shows that young people with lower school-leaving certificates and disadvantaged social backgrounds face greater difficulties in securing a training position, particularly in socioeconomically weaker regions with limited training opportunities (Hartung, Weßling, and Hillmert 2022; Hartung and Weßling 2024). Moreover, these disadvantages manifest differently across rural and urban training markets, especially for holders of lower or intermediate diplomas (Schuster and Margarian 2021).

To analyze these complex interactions at an aggregate level, labor economics focuses on matching efficiency. One key theoretical assumption is that agglomeration influences the aggregated matching efficiency, though the direction of the effect is ambiguous. On one hand, thicker labor markets may be more efficient due to a high concentration of both firms and workers, which can improve match quantity. On the other hand, they might be less efficient because of congestion effects, such as information overload and frictions in the search and screening process.

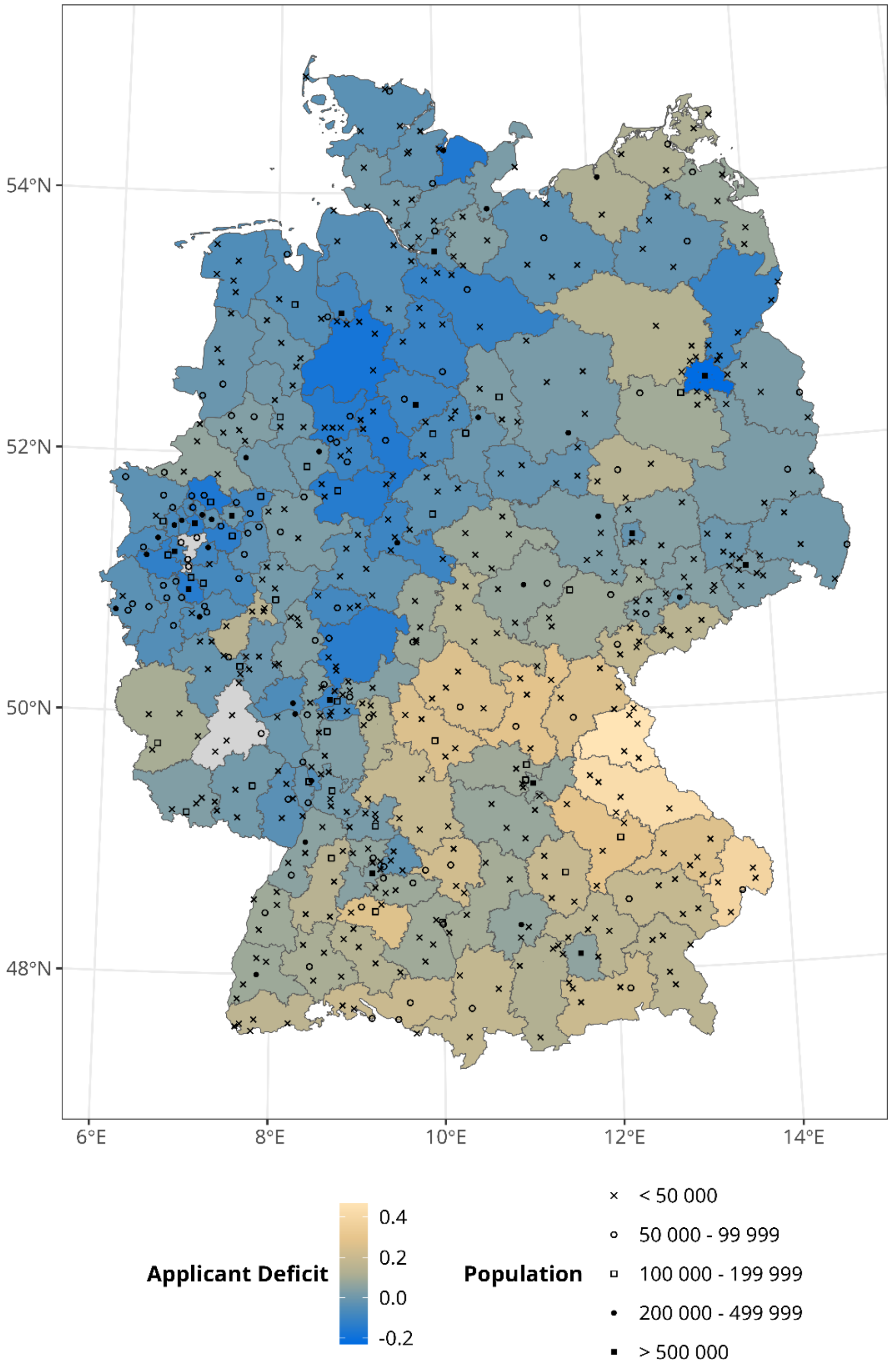
In the context of aggregated matching functions, which measure the overall result rather than the specific underlying mechanisms, it is difficult to assess the precise reasons for these outcomes. The empirical literature reflects this ambiguity, with documented effects of agglomeration pointing in both directions. For example, (Hynninen 2005) found negative estimates for population density, while other studies reported positive estimates for measures like cohort sizes (Fahr and Sunde 2006b) and regional population (Lottmann 2012; Haller and Heuermann 2016; Mohamedou 2022).

Skills2Capabilities

Co-funded by the European Union

Figure 1

Relative applicant deficit and major cities by labor agency region



Notes: Relative applicant deficit per position is shown on the color scale, and the positions of the five largest cities per district are indicated. Data: BIBB-Erhebung 2023, own calculations. Shapefiles are provided by the Bundesagentur für Arbeit (boundaries as of 2023).

Skills2Capabilities

Co-funded by the European Union

3. Data & methods

This paper estimates spatial matching functions across 148 labor agency districts and utilizes additional data to reconstruct the distribution of economic activity at smaller agglomeration levels.

The estimation is divided into four parts and starts from a basic matching function specification, without spatial modeling. The first three parts evaluate the matches and stocks of all VET positions combined: First, the analysis compares and tests different concepts for measuring the influence of agglomeration on the matching function to assess whether agglomeration effects can be seen for VET labor markets. Second, the analysis tests models with only spillovers to investigate if and how VET labor markets are interconnected. Third, the best-fitting versions of both are combined to create a full model that assesses how agglomeration and spillovers interact within the VET labor market. The fourth part re-estimates this full model for separate professional groups to describe professional heterogeneities and determine if these spatial effects differ by occupation compared to the overall VET market.

Throughout all steps, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are calculated and evaluated to assess the quality of the models when comparing different indices of agglomeration and weighing of spillovers. For the final set of nested models, likelihood-ratio (LR) tests are conducted to provide a rigorous statistical comparison for the nested agglomeration and spillovers.

The remainder of this section introduces the matching function framework, the data and data sources. Afterwards, the different measures for agglomeration and spatial weighting matrices are discussed in detail.

3.1 Matching function framework with spatial spillovers

This study is based on the aggregate job matching function model, which relates the inflow to employment (or in this case new VET contracts) to the stock of vacancies and applicants. In this chapter, I extend the aggregate matching function model to include regional influences and spatial spillovers. These extended modeling approaches were already used for the German context (Lottmann 2012; Haller and Heuermann 2016; Fedorets, Lottmann, and Stops 2019) such that the methodological framework is similar enough for comparisons.

In the Cobb-Douglas specification used here (Equation 1), $\alpha \in [0,1]$ corresponds to matching efficiency, while $\beta_U \geq 0$ and $\beta_V \geq 0$ represent the matching elasticities with respect to apprentices searching for a position (U) and vacancies (V). In this model specification, the sum of the elasticities $(\beta_U + \beta_V)$ defines the returns to scale. When assuming fixed efficiency and elasticity parameters, a simultaneous increase in both input parameters U and V by a factor x would lead to an increase of output by a factor of $x^{(\beta_U+\beta_V)}$, thus reflecting the returns to scale. The estimation of $(\beta_U + \beta_V)$ thus directly accounts for potentially different sizes of labor markets. If $(\beta_U + \beta_V = 1)$, this property is known as constant returns to scale. Consequently, the estimated efficiency parameter, α, denotes the effectiveness of the matching process, independent of the scale of the inputs.

The empirical version of this matching function (Equation 2) results from a linearization through log transformation, and can be estimated with the previously introduced data through a linear regression model. As the data used in this study is limited to the year 2023, the time parameter t is neglected. By introducing the vector notion over the set of regions (Equation 3), also the regional index can be dropped (Equation 4).

$$M_{rt} = \alpha (U_{rt})^{\beta_U} (V_{rt})^{\beta_V} \tag{1}$$

$$\ln(M_{rt}) = \ln(\alpha) + \beta_U \ln(U_{rt}) + \beta_V \ln(V_{rt}) \tag{2}$$

$$\ln(\diamond) := \begin{pmatrix} \ln(\diamond_1) \\ \vdots \\ \ln(\diamond_n) \end{pmatrix} \tag{3}$$

$$\ln(M_r) = \ln(\alpha) + \beta_U \ln(U) + \beta_V \ln(V) + \epsilon \tag{4}$$

At this stage, the efficiency parameter α is assumed to be constant. However, there are two main reasons to include spatial variations in the model. First, it is likely that real-world regional differences in matching efficiency exist, which might arise from institutional or geographic variations.

Second, matching functions are often estimated using spatially aggregated data. Consequently, they are simultaneously subject to spatial biases, such as externalities introduced by the Modifiable Areal Unit Problem (MAUP) (Openshaw 1979, 1984; Wong 2004). Accounting for spatial variations can help correct these methodological issues.

While the computation of regional fixed effects requires observations for several time points for the same regions, regional differences can be modeled using regional indicators. The first extension of this framework assumes that the matching efficiency, otherwise considered nationally constant, is influenced by specific regional characteristics. Ilmakunnas and Pesola (2003) find significant influences on the efficiency by the share of young people in the workforce and the GDP per capita. Fahr and Sunde (2006b) compare multiple influences on the efficiency and attest regional differences can be explained by the shares of high educational attainment and the cohort sizes. Such influences can be modeled by incorporating regional indicators into Equation 5 as linear dependencies.

The resulting estimates can then be interpreted in terms of their impact on the national efficiency constant. Negative coefficients indicate a regional reduction in efficiency, whereas positive coefficients suggest an increase. However, due to the log-transformation applied in the estimation context to achieve the empirical version, these effects are not linear.

$$\ln(M_r) = \ln(\alpha) + \beta_U \ln(U) + \beta_V \ln(V_r) \\ + \sum\nolimits_i \gamma_{Ri} \ln(R_i) + \epsilon \tag{5}$$

As the Cobb-Douglas specification has constant returns to scale, it does not consider any scaling effects through larger economies. To measure whether agglomeration has an effect on the matching efficiency of a regional labor market, it needs to be modeled explicitly into the matching function framework. Hynninen (2005) shows a reduction in efficiency for Finnish regions with high population density, unless spatial spillovers from neighboring regions are controlled for. Recent works control the agglomeration in terms of population (Lottmann 2012; Haller and Heuermann 2016; Mohamedou 2022). To differentiate the multiple influences, this paper tests for agglomeration influences, further denoted as A, alongside other regional indicators R_i.

$$\ln(M_r) = \ln(\alpha) + \beta_U \ln(U) + \beta_V \ln(V_r) \\ + \sum\nolimits_i \gamma_{Ri} \ln(R_i) + \gamma_A \ln(A) + \epsilon \tag{6}$$

$$\ln(M) = \ln(\alpha) + \beta_U \ln(U) + \beta_V \ln(V) \\ + \sum\nolimits_i \gamma_{Ri} \ln(R_i) + \gamma_A \ln(A) \\ + \delta_U W \ln(U) + \delta_V W \ln(V) + \epsilon \tag{7}$$

The data do not distinguish between local hires and hires from other regions. Thus, especially the estimates for regions with high inward or outward commuting from neighboring areas can be heavily impacted. Therefore, the matching function framework needs to include potential spillovers in the labor force across labor agency district boundaries.

Coles (1994) argues that well-connected labor markets act as a single bigger one, which alters the returns to scale in the estimation. Burda and Profit (1996) provide an early example of explicit modeling of such spillovers by spatially extended matching functions. They show that neighboring regions in the Czech labor market show significant spillovers in unemployment and vacancies, with a decline in strength in the distance.

Therefore, the second extension of the framework allows for spatial spillovers between such matching functions: apprentices from one region may search for jobs in other regions, and vice versa. The model assumes that M is influenced by independent variables U and V of other regions, proportionally to a spillover weight. This extension corresponds to the Spatially Lagged X (SLX) model in the field of spatial econometrics (Halleck Vega and Elhorst 2015). Spillover weights are modeled to capture the strength of association between two regions relative to all pairs of regions, for example by drawing on physical measures such as proximity. With the vector representation for indices of r, the set of all spillover weights can be described with $r \times r$- weight matrices, as in equation 7. Popular choices include contiguity neighborhoods, commuting distances and commuting flows. Haller and Heuermann (2016) compare such matrices for the German labor

Skills2Capabilities

Co-funded by the European Union

market and state that the best fit is achieved by using the inverse of the geographic distance between regions.

3.2 Data sources

The key variables in this study are the overall number of matches (M), the stock of VET applicants U, and the stock of vacancies V, within a region's labor market.

The first data source, the BIBB survey on newly concluded training contracts (Bundesinstitut für Berufsbildung (BIBB) 2023), provides most of these essential inputs. This study uses the 2023 version of the survey, incorporating data at the level of 148 districts of the German Labor Market Agency. It lists the number of signed contracts within a one-year period M (starting from October 01)[6], as well as the registered unassigned VET applicants and registered vacant positions within each district. The data is agglomerated on professional main groups according to the German Classification of Professions (KLDB 2 digit[7]). These groups aggregate specific jobs (5-digit level) into broader occupational categories, based on their similarity. Because of missing variable entries for certain professions and regions, not all professional main groups can be accounted for in the estimation process, as this would heavily impact the spatial matching framework. This paper focuses on profession main groups with both high geographical coverage and a high number of matches.

Table 3 shows the final selection of professions, their KLDB codes, the number of regions with complete information, and statistics about the distribution of matches, vacancies and applicants. Germany comprises 16 federal states that are officially divided into one or more of Germany's 401 administrative districts (NUTS-3). The geographical unit of a labor agency district is an intermediate step of aggregation: each federal state comprises at least one labor agency district, and each labor agency district comprises one or more administrative districts[8]. As most data is available on the level of 401 administrative districts, this study recalculates statistics to match the 148 districts based on population or area-weighted aggregation.

Regional information and demographics are sourced from the INKAR Regional Database (BBSR 2024). This study uses several regional variables: youth unemployment below an age of 25 (unemp25) to model regional economic prospects, the vote share of conservative party voters (cons) to model societal valuation and norms, and shares of diploma-types to control for the regional education level. It differentiates between the share of annual school leavers with no or only a lower secondary education certificate, the share of individuals with intermediate secondary education certificates (obtained after 10th grade, medSE) and those with higher secondary

[6] Most VET contracts begin on August 1 or September 1, as the vocational school year generally aligns with the standard academic calendar. While it is possible to secure a contract and begin practical training earlier in the year, the school-based component typically starts on these fixed dates. Consequently, the data covers all newly concluded contracts within the one-year period ending on September 30.

[7] Also more granular data on the level of 3-digit codes is available. Due to Data Protection laws, any entries below a threshold of 3 are censored, such that even the most matched KLDB 3 Digit professions, show too many missing

[8] Berlin is the only exception. It is simultaneously a federal state and an administrative district, but because of its size, it is split into 3 labor agency districts.

education certificates (obtained after 12th to 13th grade, highSE). All regional variables are encoded in percentage points (%). Information about the spatial distribution of economic indicators, like urban permeation and density, is extracted from the IOER-Monitor (Krüger, Meinel, and Schumacher 2013).

Spillovers between labor agency districts are computed by summing the stocks of applicants and vacancies in adjacent regions and related occupations, weighted by different abstract or physical proximity measures. Several additional data sources were included in the analysis to calculate such weightings: The first one is the number of commuters and apprentices who commute within and outside a NUTS3 region. These regions have been aggregated to the higher labor agency district levels to account for commuting between them. The data provided by the German Ministry of work for the year 2023 (Statistik der Bundesagentur für Arbeit 2023) represents the most recent available information. Additional weightings like the inverse centroid distances and neighborhoods based on contiguities were computed from the underlying maps.

Furthermore, this study uses information about Germany's municipalities and cities from the Wikidata SPARQL Endpoin[9]t. The list includes the geolocation, last reported population, and administrative district affiliation of all German cities. Results were curated for all cities and the largest municipalities to ensure accurate assignment of city status and district. From this list of 10,820 geolocations, up to 5 of the most populous settlements per labor agency district are sampled, with the target to cover at least 50% of the population per district. 615 settlements made it into the final selection, with an agglomerated population of 44 037 746 inhabitants, who live in the selected settlements.

The last data source is the Google Distance Matrix API to query the transport times between all cities in the selection that are closer to each other than 50 km or within neighboring regions. Since the German railways update their schedule twice a year, two sampling times were chosen to account for potential construction: Monday mornings at the start and end of the schedule. To factor out possible weather events, traffic jams, or accidents, driving and commuting times were queried shortly prior to those days and reflect only the estimated driving times. (Accessed: 09.07.23, 02.12.23)

3.2.1 Measures of agglomeration

The matching function literature uses different measures of agglomeration to control for regional effects, to measure differences between rural and urban environments. These are often indexes that are easy to access: density (Hynninen 2005; Mohamedou 2022) or population numbers (Lottmann 2012).

As previously discussed, agglomeration might have different effects on the VET market than on the general labor market. This study tests whether general labor market agglomeration is associated with a difference in efficiency of VET labor markets. Given that it is unclear how to best measure agglomeration for this specific group, and whether all measures would show similar

[9] query.wikidata.org, Accessed: 22.09.22, ©Wikimedia Foundation; The original query is provided at the Github Repository of this project.

Skills2Capabilities

Co-funded by the European Union

results, this paper systematically compares multiple ones. Ten agglomeration indices, outlined in Table 1, are used for this comparison.

Some studies identify agglomeration in the VET market using administrative definitions, such as the distinction between city and rural districts. More detailed approaches involve measuring continuous variables like urban permeation (Kubitza and Weßling 2025), which capture varying degrees of urban influence. Urban permeation, as defined by Jaeger et al. (2010), measures the dispersion of settlements alongside their size. Lottmann (2012) addresses the Modifiable Areal Unit Problem (MAUP) by summing the populations of neighboring districts, rather than using only a district's population as a measure of agglomeration. This approach helps account for urban districts, which are surrounded by separate administrative units not forming distinct labor agency districts.

In contrast to this study, previous labor market research has often used more disaggregated data at the NUTS-3 region level. Given the potentially lower mobility of younger apprentices and the larger spatial units used here, it is more reasonable to disaggregate population numbers than to inflate them by including neighboring agglomerations.

Therefore, in this study, the population measure is further refined by incorporating the concept of 'average reachable population'. This index is calculated by averaging the population within buffers around all sampled cities within each labor agency district. Those buffers are drawn by car travel times, public transport times, and geographic distances and include everything that is reachable within a certain threshold.

This adjustment addresses labor agency districts that are too large to form a single coherent labor market, as they comprise multiple cities located too far apart for reasonable commuting. Only labor agency districts that have all their cities within each other's buffers will have unadjusted population numbers. For all other cases, the population number is decreased. If none of the cities lie within each other's buffers, the index denotes the average population of all cities. The actual sizes of the buffers are chosen as the best-fitting decile in the distribution of all distances sampled in this study, and correspond to 9.7 km of distance, 25.6 minutes of car driving time, and 49.3 minutes of public transport commuting.

Table 1

Overview of used agglomeration indices.

Description	Min	Max	Mean	SD	Q1	Med	Q3
Indicator for being a Single City labor Market	0	1	0.144	0.352	0	0	0
Share of Population within the biggest city	0.045	1	0.301	0.26	0.13	0.214	0.339
Urban Permeation[a] (Jaeger et al., 2010)	1.61	32	8.72	7.31	4.14	5.28	10.4
Population per area covered by settlements (in 1000 / km^2)	0.693	6.46	2	1.17	1.17	1.56	2.52
Number of Adolescents (in 1k) between 16 and 22 years	5.18	270	23.4	28.3	11.5	15.8	25.4
Population (in 1k) of a certain region	211	3760	518	354	352	461	588
Population (in 1k) that can be reached by an average citizen within 9.7 km distance	17	3660	173	382	38.3	60.2	131
Population (in 1k) that can be reached by an average citizen within 25.6 minutes of car driving	24	3660	201	380	53.5	94.7	185
Population (in 1k) that can be reached by an average citizen within 49.3 minutes of public transport	24.5	3660	223	378	64.1	115	234

Notes: Overview, description, and parametric, as well as non-parametric statistics for the different measures of agglomeration used in this study; data sources: INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, and own

[a] It measures the intensity and spread of buildings in the landscape simultaneously, thus assigning higher values to

Variable	Name
CITY	Single City Region
CITY-SHARE	Biggest City Share
URB	Urban Permeation
DEN	Population Density
COHO	Cohort Size
POP	Population
POP-DIST	Population Distance Weighted
POP-DRIVE	Population Driving Weighted
POP-PUBLIC	Population Public Transport Weighted

3.2.2 Weighting of spillovers

Spatial interactions in the labor market arise when job seekers (in this case apprentices) search for positions in other districts and are willing to commute to find a job. To properly test the second part of the central question whether regional spillovers affect the VET labor market, the matching function framework must model these spillovers. This requires using weight-matrices that can realistically capture the mobility of apprentices. To assess this, this paper systematically compares existing approaches and introduces new measures.

The modeling of neighborhoods using binary relations was first formalized by Moran (1948). Building on this foundation, early works on the spatial labor market employed border-based neighborhood concepts to calculate spillover effects (Burda and Profit 1996). They find significant effects of neighboring stocks of unemployment on local matching. More recent studies in the spatial matching function literature employ weighting matrices based on methods such as inverse centroid distances or commuter flows (Haller and Heuermann 2016; Fedorets, Lottmann, and Stops 2019), offering the potential for more realistic and robust results.

To account for spatial effects among young labor market participants in VET, this study systematically compares different spatial weight matrices. Haller and Heuermann (2016) used a similar approach for the general labor market. Table [tab:spillovers] lists all weighting definitions used in this analysis. Overall, two classes of weighting concepts are considered: conceptual definitions, such as belonging to higher-order regions or shared borders defined by neighborhoods, and spatial measures, such as distances or commuting flows.

First, this study considers classic neighborhood definitions based on shared borders. Besides these definitions, the analysis also defines neighborhoods based on commuting times and distances. Two labor agency districts are defined as neighboring if at least one city in each district can be reached from the other within a reasonable commuting time or distance. Similarly, the study assesses whether neighboring regions share functional labor markets.

Skills2Capabilities

Co-funded by the European Union

Selecting functional regions as the geographic unit provides a straightforward approach to ensuring spatial control. These regions minimize cross-border connections and maximize internal linkages, thereby enhancing the accuracy of spatial analysis. As they are composed of smaller units, regional statistics can often be calculated more easily. In Germany, the delineation by Eckey, Kosfeld, and Türck (2006) became outdated following a spatial unit reform. The most widely used concept now is the delineation proposed by Kosfeld and Werner (2012), which defines 141 functional regions. The corresponding weighting matrix evaluated in this study is defined by whether two labor agency districts belong to the same functional region. A more recent delineation by Kropp and Schwengler (2016) was not considered, as its aggregation into only 30 to 75 units is unrealistically coarse for this analysis.

In the next step, this work includes data underlying the functional regions: the number of inward and outward VET commuters, using them as regional weights. Theoretically, these measures offer some of the most realistic estimates for assessing spillovers. While not the initial cause of spillovers, they are a direct outcome of people searching for labor in adjacent districts.

Other common ways to define spatial weights include distance-based weightings, such as the inverse distance of centroids, as used by Fedorets, Lottmann, and Stops (2019) and Haller and Heuermann (2016). Unlike commuting data, these measures are easy to access and calculate. Typically, distances above a certain threshold are truncated to avoid bias from small spillovers. Since my data includes city locations within labor agency districts, a natural extension is to average the inverse distances over all sampled cities rather than using the geographic center.

However, distance measures do not fairly capture commuting relationships. 20 km of commuting is far less effort in regions with good public infrastructure. Although incorporating travel times is an obvious solution, such measures have been seldom used in the modeling of spatial matching functions.

For the next weighting matrix, this work considers a concept similar to the reachability-agglomeration indices: the share of the population of another region that can be reached by an average person in that region within a buffer of geographic distance or commuting time by different modes. These measures have the theoretical advantage of being calculated based on realistic commuting times and performing a population-based weighting. Especially for cities close to administrative borders, such measures should be appropriately corrected for spatial interconnection.

Some weight matrices are easier to interpret than others after estimation. For example, binary neighborhood relations can be interpreted directly as the existence of a potential effect. With continuous scaled measurements, increasing the weights by a factor leads to a proportional decrease in the spillover estimate by the inverse of that factor. The choice of normalization affects model interpretation. Row standardization assumes each region has a fixed spillover size, distributed proportionally among connected regions. All non-binary matrices were globally normalized, which scales all connections relative to the strongest connection in the entire dataset.

Table 2

Overview of used spillover matrices.

Variable	Name	Description	Non-	Min	Max	Mean	Q1	Med.	Q3
W-N-1	Neighbors	Binary Weighting for Neighborhoods of degree 1	754	1	1	1	1	1	1
W-N-DIST	Dist Neighbor	Binary Weighting for regions that have at least one town within a fixed buffer of 37.02 km in distance	920	1	1	1	1	1	1
W-N-DRIVE	Car Neighbor	Binary Weighting for regions that have at least one town within a fixed buffer of 51.75 minutes of commuting by	1081	1	1	1	1	1	1
W-N-PUB	Public Neighbor	Binary Weighting for regions that have at least one town within a fixed buffer of 51.28 minutes of commuting by public transport	1340	1	1	1	1	1	1
W-FUNC	Functional Regions	Binary Indicator of whether two regions share the same functional region	216	1	1	1	1	1	1
W-COM-VET-I	VET Commuters In	Inward Commuters participating in VET	1981	0.00369	1	0.0416	0.00369	0.00738	0.0406
W-COM-VET-O	VET Commuters	Outward Commuters participating in VET	1981	0.00369	1	0.0416	0.00369	0.00738	0.0406
W-INV-CENT	Inv. Cent. Distance	Inverse of Centroid Distance	446	0.186	1	0.28	0.204	0.243	0.298
W-INV-DIST	Inv. Distance	Inverse of average distance between cities	1194	0.0299	1	0.153	0.102	0.134	0.176
W-INV-DRIVE	Inv. Car	Inverse of average driving times between cities	1194	0.0598	1	0.202	0.154	0.192	0.231
W-INV-PUB	Inv. Public	Inverse of average commuting times by public transport between cities	1194	0.0735	1	0.217	0.156	0.195	0.253
W-POP-DIST	Dist Population	Share of the population that can be reached by an average citizen within 20.43 km	344	0.0482	1	0.474	0.176	0.368	0.769
W-POP-DRIVE	Car Population	Share of the population that can be reached by an average citizen within 30.37 minutes of driving	342	0.0482	1	0.427	0.16	0.282	0.666
W-POP-PUB	Public Population	Share of the population that can be reached by an average citizen within 51.28 minutes of commuting by	389	0.0747	1	0.541	0.216	0.455	1

Notes: Overview, description, and parametric as well as non-parametric statistics for the distribution of nonzero weights used in this study; non-binary matrices are globally normalized; data sources: INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, and own calculations

3.3 Descriptive Statistics

Figure 2

Scatterplot of applicant and vacancy quotas with fitted inverses

Notes: Scatter of regional vacancy and applicant rates, together with hyperbolic regression curves; professional groups are defined by colors; data: BIBB-Survey 2023, own calculations.

Table A1 presents a detailed overview of the regional average number of matches, vacancies, and VET applicants in 2023, as well as the count of regions with missing observations. For "all professions," there are 146 data entries, with an average of 3,278.11 matches and a stock of 3,774.24 positions, and 3,705.95 applicants, showing a small overhead of vacancies.

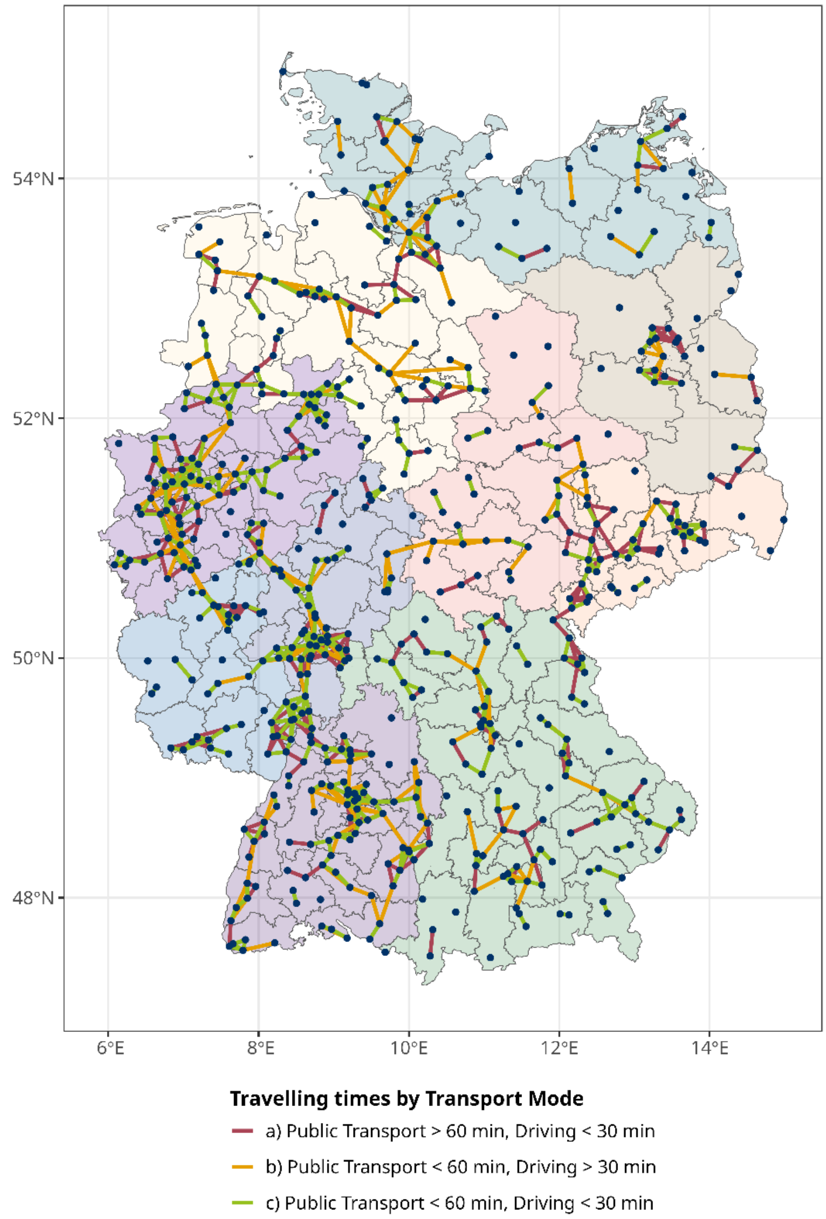
Regional coverage and matches vary across professional groups. Sales professions, for example, with 144 observations, report an average of 340.54 matches. The medical health professions show the lowest number of observations from the selected profession groups, with 133 reported regions. None of the major professional groups achieves complete geographical coverage. A comparison between the professions reveals that not all professions lack applicants. Major unmet demand exists only for certain professions. For instance, in sales professions, the average number of vacancies (442.94) exceeds the average number of applicants (402.72), showing a surplus of job openings. The contrary trend is observed in machine and vehicle technology professions, where the average number of open positions (352.23) is lacking behind the stock of applicants (364.39), although by a smaller margin. In the medical health professions, the number of positions (275.64) almost aligns with the number of applicants (271.64), reflecting a more balanced demand and supply in this field. Figure 2 shows a scatter plot of the regional relationship between the share of unassigned applicants and open vacancies. The x- and y-axes represent the respective rates, and different professions are color-coded. In addition, the figure displays the estimated Beveridge curves.

Occupational efficiency differences are clear with outer curves, showing more unassigned applicants and open vacancies together. The high rates in sales professions show that labor market frictions exist, at least for some professions. Next to introducing the used measures of agglomeration, table 1 also provides the summary statistics for them. Although the scales differ, most of these measures show a strong steepness in the distribution, with most observations close to the minimum or maximum. Most of the labor agency districts cover huge parts of rural areas and only contain a handful of bigger cities that serve as regional centers. This also leads to a strong discrepancy in means between the population and the reachable population measures. While the average population per labor agency district is approximately 518,000, only between 173,000 up to 223,000 inhabitants can reach another in reasonable time or distance. Figure 3 illustrates all labor agency districts in Germany, highlighting their largest cities and the other cities accessible within reasonable travel times by various modes of transportation. The districts are color-coded according to their higher-order spatial units, and connections between cities indicate whether they are reachable within a specified time buffer. The color of the connections represents the types of available transport links. The western part of Germany features several well-connected regions forming a broad network, while some areas in the East and South are well-linked to their neighbors without forming an extensive network. In contrast, certain regions are just isolated. This spatial diversity should be reflected in any weighting matrix when modeling spatial spillovers.

Table 2 also provides summary statistics for the weighting matrices, including the number of nonzero weights and the distribution of these connection weights. All connections are normalized so that the well-connected regions have a value of 1, representing maximum connectivity. For binary-encoded neighborhood matrices, statistics such as the mean and quantiles are less informative. For other matrices, the number of nonzero connections ranges from 344 to 1,981. The distribution statistics for both commuter-based weighting matrices (W-COM-I, W-COM-O) are equivalent, as they represent perturbations of one another. Both commuter-based matrices show low statistics across all quantiles, since only a handful of neighboring districts are simultaneously populated densely enough to generate high commuting flows. A similar distribution is observed for weighting matrices defined by inverse distances (W-INV-CENT to W-INV-PUB), since high inverse values can only occur when centroids or cities are located close to each other. In contrast, the population-weighted measures (W-POP-DIST, W-POP-DRIVE, W-POP-PUB) exhibit higher quantiles and a less tail-heavy distribution. This results from the fact that they only include relations between regions with at least one city within the specified buffers around the other, thereby limiting the number of nonzero connections.

Figure 3

Major German cities and traveling times between them.



Notes: Depicted are labor agency districts and their largest cities. Labor agency districts are colored according to their higher-order spatial units. Connections between cities indicate whether both cities are reachable within a specified time buffer; the colors of the connections represent the types of available connections. Data: Google Distance Matrix API, own calculations. Shapefiles are provided by the Bundesagentur für Arbeit (boundaries as of 2023).

4. Results

The analysis of this paper proceeds in four stages: first, all VET positions are evaluated with regional concepts, considering only agglomeration and regional controls. Second, spatial spillovers are modeled without agglomeration. Third, both influences are combined, and the three best-fitting agglomeration indices and weighting matrices are merged into nine final models to estimate matching functions for the overall VET market. These steps are described in subsection 4.1. Finally, the best-fitting combination is re-estimated for different professions. These models are then compared in terms of their regional and spatial influences in subsection 4.2, revealing substantial variation in the importance of spatial and regional determinants.

4.1 Spatial Modeling

The first part of this study disregards spatial spillovers and compares multiple indices of agglomeration for the entire VET market. The models are implemented stepwise by adding regional controls to the base models, as well as different implementations of agglomeration indices. Table A2 shows the regression results of these models. Significant agglomeration influences emerge only for a subset of indices (*CITY-SHARE, COHO, POP-DIST, POP-DRIVE, POP-PUBLIC*). These results align with Lottmann (2012), where agglomeration is only visible in some models. Haller and Heuermann (2016) found a significant positive coefficient for the regional population, suggesting that more populous regions experience higher job matching rates. This link between population and efficiency is not supported by the present VET data. While the cohort size (COHO), which was also identified by Fahr and Sunde (2006b), show weakly significant agglomeration influences, only the agglomeration indices that adjust the population size *(CITY-SHARE, POP-DIST, POP-DRIVE, POP-PUBLIC)* show stronger significance.

Across these indices, I observe consistent associations: regions with higher levels of agglomeration exhibit greater efficiencies, with estimated coefficients of up to 0.018 (ln-scale), implying that a doubling of population size increases efficiency by a factor of approximately 1.012. Overall, the indices *CITY-SHARE*, indicating the percentage of population living in the biggest city of each district, *POP-DIST* and *POP-PUBLIC* indicating the Population that can be reached by an average citizen within 9.7km of distance or 49.3 minutes of public transport usage, provide the best model fit and all show significant estimates for agglomeration.

Introducing agglomeration indices also affects the estimates of regional indicators. In the baseline model, without accounting for agglomeration, only the share of students with intermediate secondary education diplomas shows a significant positive impact on efficiency. Depending on the agglomeration specification, however, this regional variable loses its statistical significance (as with *POP-PUBLIC*), while retaining a positive coefficient. For the best-fitting specifications *(CITY-SHARE, POP-DIST, POP-PUBLIC)* of agglomeration, I additionally observe a weakly significant positive association with youth unemployment.

The second part of the analysis incorporates spillovers, excluding the agglomeration indices discussed above. Tables A3 and A4 present the results for different weighting matrices, which are divided into neighborhood- and commuting-flow-based matrices, as well as distance- and commuting-time-based matrices. Significant spillovers emerge only for four weighting matrices (*W-*

N-PUB, W-COM-VET-O, W-INV-CENT, W-POP-PUB), two of which belong to the most suitable models. *W-POP-DIST*, the third-best-fitting model, yields larger spillover estimates than most others, although these are not statistically significant at the 5% level.

All weighting matrices capture spatial dependence and thus improve model fit relative to both the baseline model and the model with regional controls. However, most improvements are only marginal, raising doubts about the necessity of spillover modeling given the relatively large spatial units available for VET data. Only *W-POP-PUB* shows an improvement in the Bayesian Information Criterion (BIC), which penalizes model complexity.

As this study applies global normalization of the matrices, the estimated sizes are not directly comparable to existing studies that use row-based normalization. Because of this 'global normalization' method, the here presented coefficient values may appear larger than in studies using other methods. The key takeaway is not the exact number, but the statistical significance and direction of the effect.

For the most connected regions, the models used here indicate an elasticity of up to –0.197 in the log-transformed stock of applicants in neighboring regions. Although not directly comparable in effect sizes, the significance and direction of the spillovers are consistent with previous findings, such as those reported by Haller and Heuermann (2016) for the general labor market. In contrast to their results, however, the spatially lagged stocks in vacancies and applicants are of similar magnitude, suggesting a shift in sensitivity between the stocks, rather than changes in returns to scale.

Unlike the agglomeration indices, introducing spillover modeling does not substantially alter the estimates of regional indicators. The share of students with intermediate secondary education diplomas keeps a significant positive influence on efficiency across all tested spillover specifications.

Table 3

Best-fitting matching models with spillovers and agglomeration.

Spillover	W-POP-PUB	W-POP-DIST	W-POP-PUB
Agglomeration	**POP-PUBLIC**	**POP-PUBLIC**	**POP-DIST**
$ln(\alpha)$	-0.509 *	-0.497 *	-0.540 *
	(0.207)	(0.209)	(0.209)
β_U	0.681 ***	0.672 ***	0.684 ***
	(0.039)	(0.038)	(0.039)
β_V	0.316 ***	0.324 ***	0.316 ***
	(0.039)	(0.039)	(0.039)
$\gamma_{unemp25}$	-0.014	-0.015 ^	-0.014
	(0.008)	(0.009)	(0.009)
γ_{cons}	0.051 *	0.052 *	0.044 ^
	(0.025)	(0.025)	(0.025)
γ_{highSE}	-0.002	-0.003	0.006
	(0.024)	(0.024)	(0.024)
γ_{medSE}	0.044	0.045	0.051 ^
	(0.030)	(0.030)	(0.030)
$\gamma_{Agglomeration}$	0.019 ***	0.019 ***	0.014 **
	(0.005)	(0.005)	(0.005)
$\delta_U(lagged)$	-0.070 *	-0.051	-0.079 *
	(0.031)	(0.038)	(0.031)
$\delta_V(lagged)$	0.069 *	0.050	0.079 *
	(0.032)	(0.039)	(0.032)
logLik	299.984	298.831	298.393
AIC	-577.969	-575.662	-574.786
BIC	-545.149	-542.842	-541.966
N. obs.	146	146	146

Notes: Presented are regression estimates for the three best fitting combinations of agglomeration and spillover modeling ; S.E. in parentheses, significance level: ^p< 0.1, *p < 0.05, **p < 0.005, ***p < 0.001; Data sources: BIBB-Survey, INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, and own calculations.

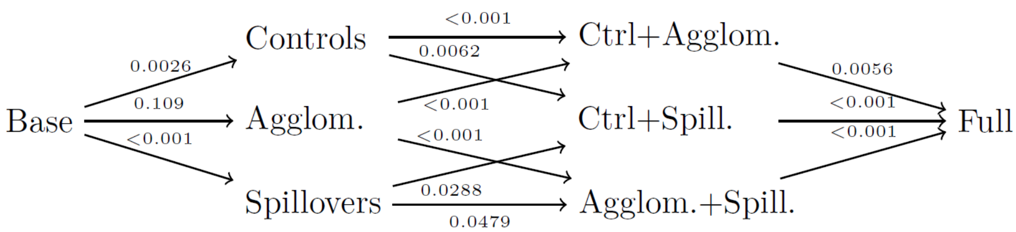
The third part of the analysis jointly incorporates spillover and agglomeration indices into the modeling framework. Among the agglomeration indices *CITY-SHARE*, *POP-DIST*, and *POP-PUBLIC* provide the best fit. Combined with the three best-performing spatial weight matrices (*W-POP-DIST*, *W-COM-VET-O*, and *W-POP-PUB*) this yields nine combinations. Table A5 in the appendix reports the full set of estimates, while Table 3 summarizes the three best-fitting specifications.

The significance, direction, and magnitude of estimates remain consistent across all agglomeration indices, regardless of the imposed spillover matrices. Across all combinations, agglomeration measured by *POP-PUBLIC* provides the best model fit. Similarly, the spatial weight matrix *W-POP-PUB* delivers the best fit regardless of which agglomeration index is employed. While all agglomeration indices yield significant estimates, only spillovers measured by *W-POP-PUB* produce significant results when combined with agglomeration indices. This is plausible, as highly agglomerated regions induce inward commuting. This relation naturally generates spatial spillovers. Thus, both channels of spatial modeling interact meaningfully in this framework. Compared to specifications focusing solely on either agglomeration or spillovers, the joint estimation using *W-POP-PUB* and *POP-PUB* produces smaller coefficient estimates for both components.

Figure 4 summarizes the nested model structures that emerge when agglomeration and spillovers are estimated jointly alongside other regional controls, with this combination of public transport commuting times based agglomeration indices and spillovers. Likelihood ratio tests between the full model and the preceding nested specifications still justify the simultaneous estimation of agglomeration, spillovers, and regional factors. Indeed, any path of model extensions consistently yields at least weakly significant improvements in model fit. Therefore, this full model is selected for the following analysis.

Figure 4

Overview of nested models and test statistics



Notes: Each node represents a model, named after the concepts the model includes. Arrows designate nesting structures and are labeled with the p-values of the LR test. Data: BIBB-Erhebung 2023, own calculations.

4.2 Occupational and Spatial Variations

Section 4.1 analyzed the spatial and regional influences on all VET professions. Likelihood ratio tests indicate that including agglomeration, spillovers, and regional controls improves model fit meaningfully, supporting the use of the full joint model for further analysis. The following part of the analysis re-estimates this spatial matching model for 5 distinct profession groups: sales; machine and vehicle technology; professions in business management and organization; mechatronics, energy, and electrical professions; and medical health professions.

Figure 5 illustrates the respective differences in the estimates for these re-estimated models, along with the corresponding 95% confidence intervals. Table 4 lists the exact estimates for these regressions, alongside the model incorporating all professions.

For the model estimated across all professions in 2023, without spatial modeling (Table A2) the results are broadly consistent with those reported by Fitzenberger et al. (2025) for the Covid period. The estimated elasticities β_U= 0.590 and β_V= 0.419, fall within or slightly below the range observed in that study. In contrast, the estimated base mode efficiency of 0.81 is substantially higher, suggesting a recovery of the VET matching process following the end of the Covid pandemic.

Under the full spatial and regional modeling framework (Table 4), clear efficiency differences emerge, driven primarily by agglomeration, but also by other regional factors. While the base efficiency is only about 60%, agglomeration plays a significant role: a one unit increase in the log-transformed population size raises efficiency by a multiplicative factor of 1.019.

Beyond agglomeration, I find a weak positive association between regional conservatism and labor market efficiency. By contrast, regional controls based on secondary education levels show no significance, diverging from the findings of Fitzenberger et al. (2025). Instead, the spatial model introduces weakly significant coefficients for the spatially lagged variables: -0.070 for δ_U and 0.069 for δ_V. This means that elasticities are not constant across regions; the effective elasticity for any given region depends on its baseline elasticity and its specific level of connectivity.

Isolated regions without connectivity show elasticities of β_U= 0.681 and β_V = 0.316, indicating strong dependence on the stock of applicants in the number of matches. For well-connected regions, this dependence is shifted by the lags, leading to less dependence on the local stock of applicants.

This finding is a key takeaway: spatial connectivity creates a more balanced matching process, as these regions suffer less from missing applicants. This is only the case if I regard the whole VET market, disregarding occupational differences. On a methodological level, this result implies that without spatial modeling (e.g., the "Controls" model in Table A3), the models slightly under-estimate the true dependence on applicants for isolated regions, and over-estimate it for well-connected regions.

In terms of regional influences, each profession group is shaped by different factors (Figure 5). Significant agglomeration influences are found for *medical and health professions*, as well as *sales professions*, while spillovers are only evident for *machine and vehicle technicians*. The absence of spillovers in other professions, contrary to the overall VET market, may reflect selection bias, since this study focuses only on occupational groups with the highest number of matches (which account for roughly 42% of all VET market matches).

In professional labor markets with sparser distribution of employers, commuting is likely to play a greater role, which could amplify spillovers. Figure 5 shows these differences visually together with confidence intervals of the estimates. While the most significant occupational differences still occur in the elasticities, suggesting limited mobility between sectors for VET students, also efficiency estimates show occupational differences.

In sum, the analysis shows that occupational and spatial differences emerge simultaneously in the VET matching process, underscoring the need to disentangle spatial dependencies at the level of individual occupational groups. While elasticities remain the strongest source of occupational heterogeneity, indicating limited mobility of VET applicants across professions, measurable agglomeration influences and spillovers arise in the whole VET market and show substantial differences across occupational groups. This suggests that although apprentices are less mobile, spatial frictions persist like in the normal labor market, but their identification requires more detailed spatial modeling. From the matching model's perspective, it remains unclear why spillovers and agglomeration influences only affect certain professions. These disparities are likely driven by the distinct regional concentration patterns and demand structures inherent to each professional group and thus require a profession-specific analysis for their explanation.

Skills2Capabilities

Co-funded by the European Union

Table 4

Regression results by profession groups.

	All	Machine & Vehicle Tech.	Mechatronics, Energy & Electrical	Medical Health	Business & Managment	Sales
$ln(\alpha)$	-0.509 *	-0.324	-0.482 *	-0.916 **	-0.551 *	-0.944 **
	(0.207)	(0.218)	(0.196)	(0.338)	(0.225)	(0.313)
β_U	0.681 ***	0.508 ***	0.694 ***	0.566 ***	0.500 ***	0.691 ***
	(0.039)	(0.041)	(0.038)	(0.052)	(0.046)	(0.034)
β_V	0.316 ***	0.503 ***	0.320 ***	0.446 ***	0.519 ***	0.319 ***
	(0.039)	(0.042)	(0.038)	(0.050)	(0.045)	(0.033)
$\gamma_{unemp25}$	-0.014	-0.006	-0.021 *	0.017	-0.008	0.005
	(0.008)	(0.009)	(0.008)	(0.016)	(0.010)	(0.013)
γ_{cons}	0.051 *	0.002	0.065 *	0.127 **	0.028	0.034
	(0.025)	(0.027)	(0.025)	(0.042)	(0.029)	(0.036)
γ_{highSE}	-0.002	0.005	0.017	-0.002	0.007	0.034
	(0.024)	(0.027)	(0.024)	(0.039)	(0.028)	(0.039)
γ_{medSE}	0.044	0.042	0.007	0.033	0.050	0.092 ^
	(0.030)	(0.033)	(0.029)	(0.048)	(0.034)	(0.047)
$\gamma_{Agglomeration}$	0.019 ***	-0.003	0.007	0.028 **	0.006	0.021 **
	(0.005)	(0.005)	(0.005)	(0.009)	(0.007)	(0.008)
$\delta_U(lagged)$	-0.070 *	-0.084 **	-0.036	0.095	-0.015	0.005
	(0.031)	(0.032)	(0.033)	(0.059)	(0.041)	(0.029)
$\delta_V(lagged)$	0.069 *	0.086 **	0.035	-0.095	0.015	-0.006
	(0.032)	(0.033)	(0.034)	(0.059)	(0.042)	(0.029)
logLik	299.984	276.552	276.644	193.812	247.251	232.410
AIC	-577.969	-531.104	-531.288	-365.623	-472.503	-442.819
BIC	-545.149	-498.825	-499.412	-335.053	-441.303	-410.228
N. obs.	146	139	134	119	126	143

Notes: Presented are regression estimates for the three best fitting combinations of agglomeration and spillover modeling ; S.E. in parentheses, significance level: ^p< 0.1, *p < 0.05, **p < 0.005, ***p < 0.001; Data sources: BIBB-Survey, INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, and own calculations.

5. Summary and conclusion

Local labor markets are subject to regional factors and spatially interconnected with their surroundings. These spillovers arise through worker mobility during job searches across different local or regional units. In Germany, with its 148 labor agency districts, researchers observed agglomeration, as well as spillovers for spatial matching functions for the general labor market. Entering the German VET system is an educational path for many of Germany's students (BMBF 2024) leading after 2-3 years of training to a certified profession. Depending on the profession, this education is school-based or part of the dual system, with company-based training and a share of VET school. Within the dual system, the matching between employers and VET students follows labor market mechanisms. Students apply for VET positions offered by employers.

This paper investigates the influence of regional agglomeration and spatial spillovers on the matching efficiency of the German VET market in a spatial matching function framework. It is the first work to provide empirical estimates of these spatial dynamics specifically for apprentices and distinct professional groups in Germany. To achieve this, it integrated a variety of data sources, including administrative records and web-scraped travel times, developed novel indices for agglomeration and spillover weights, and systematically compared them using spatial regression models. Thus, this paper serves not only as an example for spatial modeling but also provides a set of tools and measures to quantify spatial relations in the analysis of labor markets.

The primary finding is that agglomeration significantly shapes the efficiency in this spatial matching function framework for VET labor markets. Regional spillovers exist like in the general labor market, but these effects are masked unless modeled with appropriate realistic measures. Specifically, the results show that regions that are well-connected to their surroundings exhibit a more balanced VET matching process, significantly reducing the dependency on local stocks of applicants.

Spillovers and agglomeration were statistically significant parts of the models when using the novel measures based on realistic public transport commuting times. Among the agglomeration indices, the newly constructed measures of 'average reachable population within a district' provide the best fit with the data. These indices quantify the population that can be reached by an average citizen within roughly 50 minutes of public transport and were constructed by analyzing the position of major cities within each district and querying public transport travel times between them.

For spillovers, the spatial weight matrix that captures the share of the population that can be reached within a neighboring district by an average citizen within roughly 50 minutes of commuting by public transport consistently produces significant results, suggesting that commuting flows and regional connectivity generate measurable externalities not only for the general labor market (Haller and Heuermann 2016; Lottmann 2012; Fedorets, Lottmann, and Stops 2019), but also for the VET market.

After evaluating the influence of spatial modeling, this study analyzed the efficiency and elasticities of different professional groups to identify occupational differences in regional characteristics. The fact that spillovers and agglomeration affect only certain professional groups indicates that spatial dependencies are specific to each occupational structure. Consequently, these findings underscore the need to consider occupational heterogeneity alongside spatial modeling to accurately describe the VET market, demonstrating that professional groups are not directly comparable in terms of their regional dynamics.

To situate this study in the context of existing and future research, it is important to note that most prior literature focuses on the general German labor market. I can confirm the validity of the matching function estimates for the German VET market reported by Fitzenberger et al. (2025), who analyzed COVID-period data without explicitly accounting for spatial dependencies. Without incorporating the spatial concepts introduced here, our 2023 estimates are of a similar magnitude to those reported by Fitzenberger et al. (2025). Once spatial and regional modeling is applied, I observe even stronger elasticities in the stock of applicants, as well as slightly higher overall efficiency in the VET market, with efficiency rising in regions with greater agglomeration.

Since agglomeration effects and spillovers can be observed individually in VET markets, and occupational groups do not exhibit both simultaneously, future research should assess whether these dynamics have changed in recent years, particularly in light of COVID-19 and political reforms, and investigate why certain occupational groups achieve higher matching efficiency in urban areas while others do not.

While this study offers new insights into the spatial dynamics of VET markets, it is not without limitations. The exclusion of certain professional groups because of data constraints highlights the need for more comprehensive data collection. The study's reliance on a single year of data (2023) is directly related to the presence and quality of data.

Occupational data were missing for different regions in different years, which creates estimation problems for the spatial regressions, but also if a region-based time component or region-based fixed effects were to be used. Currently, post-COVID data are only available for the additional years 2022 (which precedes the most recent reform of the spatial units in 2023) and 2024. In the future, a replication of this study that also considers year-based effects would be feasible as soon as multiple reported data points per region and occupation exist. Future studies could also benefit from integrating more spatially detailed data on matches, vacancies and applicants, as the current spatial units are too large to adequately capture agglomeration and spillovers, as observed in the general labor market.

Beyond the data's level of aggregation, its quality is also a drawback. While the signed contracts must be officially reported and are thus correct as the dependent variable, the data only covers registered vacant positions and applicants (the independent variables), and thus both numbers may be incomplete. Estimates from a representative survey among employers show a rate of 35.4 percent unassigned positions (Fitzenberger, Leber, and Schwengler 2024), compared to the 15.1 percent in the data used here. As the number of unregistered applicants for VET is unknown and even more difficult to assess, the absolute level of matching efficiency is likely lower for all professions.

Consequently, the absolute magnitudes of the estimated elasticities and spatial spillover coefficients must be interpreted with caution. However, since the dependent variable (signed contracts) is accurately measured, the relative findings of this study should remain robust. This includes the comparisons of elasticities between different occupational groups and, most importantly, the identification of the travel-time-based agglomeration index as the best-fitting model specification.

Overall, this research contributes to a deeper understanding of the complexities inherent in regional labor markets, particularly within the context of VET. It provides a set of tools and measures to quantify spatial relations in the analysis of labor markets and labor market decisions for apprentices. These new tools offer a more robust toolkit for future socio-spatial research, particularly as they are capable of handling data aggregated at spatial units that might otherwise be too large or have too many missing values for a more detailed analysis.

[10]Furthermore, the empirical findings provide insights for policy, as they suggest that regions that are better connected internally and to their surroundings show a more balanced matching process between apprentices and companies and are therefore less affected by applicant shortages. Thus, these results underscore the need for regionally tailored policies that operate across administrative boundaries and consider potential commuting patterns.

6. References

Antczak, Elżbieta, Ewa Gałecka-Burdziak, and Robert Pater. 2018. "Unemployment and Vacancy Flows in Spatial Labour Market Matching at the Regional Level. The Case of a Transition Country." *Journal of Applied Economics* 21 (1): 25–43. https://doi.org/10.1080/15140326.2018.1526874.

Barbour, Elisa, and Ann Markusen. 2007. "Regional Occupational and Industrial Structure: Does One Imply the Other?" *International Regional Science Review* 30 (1): 72–90. https://doi.org/10.1177/0160017606296727.

BBSR. 2024. "Laufende Raumbeobachtung Des BBSR - INKAR." In *Bundesinstitut Für Bau- Stadt-Und Raumforschung (BBSR) Im Bundesamt Für Bauwesen Und Raumordnung (BBR)*. https://www.inkar.de.

Bick, Alexander, Adam Blandin, and Karel Mertens. 2021. "Work from Home Before and After the COVID-19 Outbreak." *Federal Reserve Bank of Dallas, Working Papers* 2020 (2017). https://doi.org/10.24149/wp2017r2.

BMBF. 2024. "Berufsbildungsbericht 2024." Bundesministerium Für Bildung Und Forschung (BMBF).

Bundesinstitut für Berufsbildung. 2024. *Verzeichnis Der Anerkannten Ausbildungsberufe 2024*. Edited by Bundesinstitut für Berufsbildung. Bonn: Bundesinstitut für Berufsbildung (BIBB).

Bundesinstitut für Berufsbildung (BIBB). 2023. "BIBB-Erhebung "Neu Abgeschlossene Ausbildungsverträge Zum 30.09.23"." https://www.bibb.de/de/179138.php.

Burda, Michael C., and Stefan Profit. 1996. "Matching Across Space: Evidence on Mobility in the Czech Republic." *Labour Economics* 3 (3): 255–78. https://doi.org/10.1016/S0927-5371(96)00013-9.

Chade, Hector, Jan Eeckhout, and Lones Smith. 2017. "Sorting Through Search and Matching Models in Economics." *Journal of Economic Literature* 55 (2): 493–544. https://doi.org/10.1257/jel.20150777.

Coles, Melvyn G. 1994. Cross-Section Estimation of the Matching Function: Evidence from England and Wales. CEPR Discussion Papers.

Ebner, Christian. 2015. “Labour Market Developments and Their Significance for VET in Germany: An Overview.” *Research in Comparative and International Education* 10 (4): 576–92. https://doi.org/gtnfq2.

Eckey, Hans-Friedrich, Reinhold Kosfeld, and Matthias Türck. 2006. “Abgrenzung Deutscher Arbeitsmarktregionen.” *Raumforschung Und Raumordnung* 64 (4): 299–309. https://doi.org/10.1007/BF03183178.

Fahr, René, and Uwe Sunde. 2006a. “Regional Dependencies in Job Creation: An Efficiency Analysis for Western Germany.” *Applied Economics* 38 (10): 1193–1206. https://doi.org/ds6v2s.

———. 2006b. “Spatial Mobility and Competition for Jobs: Some Theory and Evidence for Western Germany.” *Regional Science and Urban Economics* 36 (6): 803–25. https://doi.org/10.1016/j.regsciurbeco.2006.05.001.

Fedorets, Alexandra, Franziska Lottmann, and Michael Stops. 2019. “Job Matching in Connected Regional and Occupational Labour Markets.” *Regional Studies* 53 (8): 1085–98. https://doi.org/gtnfqn.

Fitzenberger, Bernd, Anna Heusler, Anna Houštecká, and Leonie Wicht. 2025. “The Composition of Applicants, Mismatch, and Matching Efficiency in the German VET Market.” *Labour Economics*, August, 102755. https://doi.org/10.1016/j.labeco.2025.102755.

Fitzenberger, Bernd, Ute Leber, and Barbara Schwengler. 2024. “IAB-Betriebspanel: Rekordhoch Beim Anteil Der Unbesetzten Ausbildungsstellen.” IAB-Kurzbericht 16/2024. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB). https://doi.org/10.48720/IAB.KB.2416.

Flohr, Matthias, Laura Menze, and Paula Protsch. 2020. “Berufliche Aspirationen Im Kontext Regionaler Berufsstrukturen.” *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie* 72 (S1): 79–104. https://doi.org/10.1007/s11577-020-00665-4.

Galvin, Raymond, and Reinhard Madlener. 2014. “Determinants of Commuter Trends and Implications for Indirect Rebound Effects: A Case Study of Germany’s Largest Federal State of NRW, 1994-2013.” https://doi.org/10.2139/ssrn.2534281.

Haas, Mathijs de, Roel Faber, and Marije Hamersma. 2020. “How COVID-19 and the Dutch ‘Intelligent Lockdown’ Change Activities, Work and Travel Behaviour: Evidence from Longitudinal Data in the Netherlands.” *Transportation Research Interdisciplinary Perspectives* 6 (July): 100150. https://doi.org/10.1016/j.trip.2020.100150.

Halleck Vega, Solmaria, and J. Paul Elhorst. 2015. “THE SLX MODEL.” *Journal of Regional Science* 55 (3): 339–63. https://doi.org/10.1111/jors.12188.

Haller, Peter, and Daniel F. Heuermann. 2016. “Job Search and Hiring in Local Labor Markets: Spillovers in Regional Matching Functions.” *Regional Science and Urban Economics* 60 (September): 125–38. https://doi.org/10.1016/j.regsciurbeco.2016.06.007.

Hartung, Andreas, and Katarina Weßling. 2024. “Discouraged and Hedged – Why Students Enter VET After Obtaining University Eligibility.” *Journal of Vocational Education & Training* 0 (0): 1–24. https://doi.org/10.1080/13636820.2024.2307536.

Hartung, Andreas, Katarina Weßling, and Steffen Hillmert. 2022. “Interplay Between Family Background and Labour-Market Conditions in Shaping Students’ Occupational Status Expectations.” *Journal of Education and Work* 35 (4): 405–21.

Heineck, Guido. 2011. “Regionale Typisierung: Was Ausbildungsmärkte Vergleichbar Macht.” *IAB-Kurzbericht*, no. 2011-13.

Hillmert, Steffen, Andreas Hartung, and Katarina Weßling. 2017. “A Decomposition of Local Labour-Market Conditions and Their Relevance for Inequalities in Transitions to Vocational Training.” *European Sociological Review* 33 (4): 534–50. https://doi.org/10.1093/esr/jcx057.

Hynninen, Sanna-Mari. 2005. "Matching Across Space: Evidence from Finland." *Labour* 19 (4): 749–65. https://doi.org/c68bx4.

Ilmakunnas, Pekka, and Hanna Pesola. 2003. "Regional Labour Market Matching Functions and Efficiency Analysis." *LABOUR* 17 (3): 413–37. https://doi.org/d7pcjd.

Jaeger, Jochen, René Bertiller, Christian Schwick, Duncan Cavens, and Felix Kienast. 2010. "Urban Permeation of Landscapes and Sprawl Per Capita: New Measures of Urban Sprawl." *Ecological Indicators* 10 (March): 427–41. https://doi.org/10.1016/j.ecolind.2009.07.010.

Jost, Oskar. 2019. "Regionale Mobilität von Lehrlingen: Auszubildende in MINT-Berufen Pendeln Besonders Häufig." *IAB-Kurzbericht*, no. 2019-2.

Kosfeld, Reinhold, and Alexander Werner. 2012. "Deutsche Arbeitsmarktregionen – Neuabgrenzung Nach Den Kreisgebietsreformen 2007–2011." *Raumforschung Und Raumordnung* 70 (February): 49–64. https://doi.org/10.1007/s13147-011-0137-8.

Kropp, Per, and Barbara Schwengler. 2016. "Three-Step Method for Delineating Functional Labour Market Regions." *Regional Studies* 50 (3): 429–45. https://doi.org/gc4n52.

Krüger, Tobias, Gotthard Meinel, and Ulrich Schumacher. 2013. "Land-Use Monitoring by Topographic Data Analysis." *Cartography and Geographic Information Science* 40 (3): 220–28. https://doi.org/10.1080/15230406.2013.809232.

Kubitza, Dennis Oliver, and Katarina Weßling. 2025. "Whole Lotta Training - Studying School-to-Training Transitions by Training Artificial Neural Networks." Kiel, Hamburg: ZBW - Leibniz Information Centre for Economics. https://hdl.handle.net/10419/310974.

Lee, Woong. 2019. "The Role of Labor Market Flexibility in the Job Matching Process in India: An Analysis of the Matching Function Using State-Level Panel Data." *Emerging Markets Finance and Trade* 55 (4): 934–49. https://doi.org/10.1080/1540496X.2017.1294982.

Lottmann, Franziska. 2012. "Spatial Dependencies in German Matching Functions." *Regional Science and Urban Economics* 42 (1): 27–41. https://doi.org/fjjhc3.

Lyons, Glenn, and Kiron Chatterjee. 2008. "A Human Perspective on the Daily Commute: Costs, Benefits and Trade-offs." *Transport Reviews*, March. https://doi.org/10.1080/01441640701559484.

Malin, Lydia, and Marita Jacob. 2019. "Gendered Occupational Aspirations of Boys and Girls in Germany: The Impact of Local VET and Labour Markets." *Journal of Vocational Education & Training* 71 (3): 429–48. https://doi.org/10.1080/13636820.2018.1517128.

Mohamedou, Nasser Dine. 2022. "Labor Mobility Within Japanese Regional Labor Markets and Spillover Effects." *The Japanese Economic Review* 73 (4): 625–45. https://doi.org/10.1007/s42973-020-00059-3.

Moilanen, Mikko. 2010. "Matching and Settlement Patterns: The Case of Norway." In *Papers in Regional Science*, 89:607–24. https://doi.org/10.1111/j.1435-5957.2009.00264.x.

Morales, Leonardo-Fabio, Mauricio Quiñones, Eleonora Dávalos, and Luis-Felipe Gaviria. 2025. "Spatial Spillover Effects in the Labour Market in a Middle-Income Country." *Spatial Economic Analysis* 20 (1): 53–71. https://doi.org/10.1080/17421772.2024.2378792.

Moran, P. a. P. 1948. "The Interpretation of Statistical Maps." *Journal of the Royal Statistical Society: Series B (Methodological)* 10 (2): 243–51. https://doi.org/10.1111/j.2517-6161.1948.tb00012.x.

Openshaw, Stan. 1979. "A Million or so Correlated Coefficients: Three Experiment on the Modifiable Areal Unit Problem." *Statistical Applications in the Spatial Sciences*. https://cir.nii.ac.jp/crid/1571135650655620224.

———. 1984. “The Modifiable Areal Unit Problem.” *Concepts and Techniques in Modern Geography*. https://cir.nii.ac.jp/crid/1570291225725496704.

Petrongolo, Barbara, and Christopher A. Pissarides. 2001. “Looking into the Black Box: A Survey of the Matching Function.” *Journal of Economic Literature* 39 (2): 390–431. https://doi.org/10.1257/jel.39.2.390.

Rodríguez-Pose, Andrés, Lewis Dijkstra, and Hugo Poelman. 2024. “The Geography of EU Discontent and the Regional Development Trap.” *Economic Geography* 100 (3): 213–45. https://doi.org/10.1080/00130095.2024.2337657.

Schmidt, Robyn, and Dr Alexandra Uhly. 2023. “Florierende Städte Und Abgehängtes Land? Empirische Auswertungen Zum Regionaltypischen Geschehen in Der Dualen Berufsausbildung Auf Basis Der Berufsbildungsstatistik.” *BIBB Discussion Paper*.

Schuster, Katja, and Anne Margarian. 2021. “Vocational Training Choice from a Regional Perspective.” *Empirical Research in Vocational Education and Training* 13: 1–39. https://doi.org/10.1186/s40461-020-00105-9.

Statistik der Bundesagentur für Arbeit. 2023. “Pendler — Jahresdaten Zu Ein- Und Auspendlern Für Kreise Und Gemeinden in Deutschland (Stand: Juni 2023).” https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Pendler/Pendler-Nav.html;jsessionid=A85D4083727A882CA4F1E1CE7D9BC98D. https://statistik.arbeidsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Pendler/Pendler-Nav.html;jsessionid=A85D4083727A882CA4F1E1CE7D9BC98D.

Weßling, Katarina, Andreas Hartung, and Steffen Hillmert. 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 (1): 12. https://doi.org/gk7qbz.

Wicht, Alexandra, and Wolfgang Ludwig-Mayerhofer. 2014. “The Impact of Neighborhoods and Schools on Young People’s Occupational Aspirations.” *Journal of Vocational Behavior* 85 (3): 298–308. https://doi.org/10.1016/j.jvb.2014.08.006.

Wolter, Stefan C., and Paul Ryan. 2011. “Apprenticeship.” In *Handbook of the Economics of Education*, 3:521–76. Elsevier. https://linkinghub.elsevier.com/retrieve/pii/B9780444534293000119.

Wong, David W. S. 2004. “The Modifiable Areal Unit Problem (MAUP).” In *WorldMinds: Geographical Perspectives on 100 Problems*, edited by Donald G. Janelle, Barney Warf, and Kathy Hansen, 571–75. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-2352-1_93.

Appendix A Tables:

Table A1

Professions, geographic coverage of data and average statistics.

Profession	KLDB	N.obs	Matches	Vacancies	Applicants
All professions	——	148	3278.11	3774.24	3705.95
Sales professions	62—	144	340.54	442.94	402.72
Machine and vehicle technology professions	25—	145	325.45	352.23	364.39
Professions in business management and organization	71—	135	288.62	316.98	330.19
mechatronics, energy, and electrical professions	26—	140	282.36	310.48	307.45
Medical health professions	81—	133	240.27	275.64	271.64

Notes: Presented are number of matches, vacancies and applicants per profession group averaged over all reported unites; N.obs denoting the number of non-missing observations; data sources: BIBB-Survey 2023, own calculations;

Table A2

Matching model with agglomeration indices.

	Base	Controls	CITY	CITY SHARE	URB	DEN	COHO	POP	POP DIST	POP DRIVE	POP PUBLIC
$ln(\alpha)$	-0.210 *** (0.054)	-0.662 ** (0.216)	-0.669 ** (0.217)	-0.630 ** (0.211)	-0.671 ** (0.215)	-0.641 ** (0.216)	-0.556 * (0.219)	-0.723 ** (0.237)	-0.539 * (0.215)	-0.580 ** (0.215)	-0.504 * (0.213)
β_U	0.590 *** (0.029)	0.655 *** (0.040)	0.657 *** (0.040)	0.672 *** (0.039)	0.648 *** (0.040)	0.653 *** (0.040)	0.633 *** (0.041)	0.662 *** (0.041)	0.651 *** (0.039)	0.644 *** (0.039)	0.648 *** (0.038)
β_V	0.419 *** (0.030)	0.362 *** (0.039)	0.360 *** (0.040)	0.340 *** (0.039)	0.364 *** (0.039)	0.360 *** (0.039)	0.375 *** (0.039)	0.362 *** (0.040)	0.351 *** (0.039)	0.360 *** (0.039)	0.350 *** (0.038)
$\gamma_{unemp25}$		-0.010 (0.008)	-0.011 (0.009)	-0.019 * (0.009)	-0.013 (0.009)	-0.010 (0.008)	-0.011 (0.008)	-0.009 (0.008)	-0.018 * (0.009)	-0.014 ^ (0.008)	-0.018 * (0.008)
γ_{cons}		0.019 (0.024)	0.022 (0.026)	0.046 ^ (0.026)	0.028 (0.025)	0.029 (0.025)	0.031 (0.025)	0.020 (0.025)	0.044 ^ (0.025)	0.040 (0.025)	0.050 * (0.025)
γ_{highSE}		0.017 (0.025)	0.017 (0.025)	0.009 (0.024)	0.015 (0.024)	0.012 (0.025)	0.003 (0.025)	0.022 (0.026)	-0.000 (0.025)	0.002 (0.025)	-0.008 (0.025)
γ_{medSE}		0.076 * (0.031)	0.075 * (0.031)	0.069 * (0.030)	0.078 * (0.030)	0.072 * (0.031)	0.057 ^ (0.031)	0.083 * (0.033)	0.057 ^ (0.031)	0.062 * (0.031)	0.048 (0.031)
$\gamma_{Agglomeration^-}$			0.004 (0.010)	0.014 ** (0.005)	0.010 ^ (0.006)	0.010 (0.007)	0.013 * (0.006)	-0.008 (0.013)	0.013 ** (0.005)	0.012 * (0.005)	0.018 *** (0.005)
logLik	280.639	288.797	288.894	292.896	290.238	289.773	291.054	289.003	292.977	291.886	294.804
AIC	-553.279	-561.593	-559.789	-567.791	-562.477	-561.545	-564.107	-560.006	-567.953	-565.772	-571.607
BIC	-541.344	-537.724	-532.937	-540.939	-535.624	-534.693	-537.255	-533.153	-541.101	-538.920	-544.755
N. obs.	146	146	146	146	146	146	146	146	146	146	146

Notes: Presented are regression estimates for base, regional and agglomeration modeling; S.E. in parentheses, significance level: ^p< 0.1, *p < 0.05, **p < 0.005, ***p< 0.001; Data sources: BIBB-Survey, INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, and own calculations.

Table A2

Matching models with neighborhood and commuting spillovers.

	Controls	W-N 1	W-N DIST	W-N DRIVE	W-N PUB	W-N FUNC	W-COM VET-I	W-COM VET-O
$ln(\alpha)$	-0.662 **	-0.669 **	-0.699 **	-0.717 **	-0.646 **	-0.655 **	-0.612 **	-0.733 ***
	(0.216)	(0.222)	(0.217)	(0.219)	(0.214)	(0.217)	(0.220)	(0.215)
β_U	0.655 ***	0.664 ***	0.707 ***	0.696 ***	0.718 ***	0.673 ***	0.668 ***	0.693 ***
	(0.040)	(0.047)	(0.050)	(0.049)	(0.050)	(0.044)	(0.041)	(0.041)
β_V	0.362 ***	0.352 ***	0.310 ***	0.322 ***	0.297 ***	0.343 ***	0.344 ***	0.322 ***
	(0.039)	(0.047)	(0.051)	(0.049)	(0.050)	(0.044)	(0.041)	(0.041)
$\gamma_{unemp25}$	-0.010	-0.010	-0.008	-0.008	-0.007	-0.010	-0.008	-0.007
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
γ_{cons}	0.019	0.019	0.022	0.022	0.015	0.015	0.023	0.018
	(0.024)	(0.025)	(0.024)	(0.025)	(0.024)	(0.025)	(0.025)	(0.024)
γ_{highSE}	0.017	0.019	0.025	0.025	0.021	0.019	0.016	0.035
	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)
γ_{medSE}	0.076 *	0.076 *	0.077 *	0.079 *	0.074 *	0.077 *	0.069 *	0.082 **
	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.031)	(0.031)	(0.030)
$\delta_U (lagged)$		-0.001	-0.012	-0.008	-0.036 *	-0.018	-0.098	-0.197 *
		(0.004)	(0.011)	(0.009)	(0.016)	(0.024)	(0.068)	(0.094)
$\delta_V (lagged)$		0.001	0.012	0.008	0.036 *	0.018	0.099	0.195 *
		(0.004)	(0.011)	(0.009)	(0.016)	(0.023)	(0.068)	(0.094)
logLik	288.797	289.452	290.783	290.268	291.333	289.419	290.427	292.555
AIC	-561.593	-558.905	-561.566	-560.536	-562.665	-558.838	-560.854	-565.109
BIC	-537.724	-529.069	-531.730	-530.700	-532.829	-529.002	-531.018	-535.273
N. obs.	146	146	146	146	146	146	146	146

Notes: Presented are number of matches, vacancies and applicants per profession group averaged over all reported unites; N.obs denoting the number of non-missing observations; data sources: BIBB-Survey 2023, INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, own calculations;

Table A4

Matching models with distance, driving and public transport spillovers.

	W-INV CENT	W-INV DIST	W-INV DRIVE	W-INV PUB	W-POP DIST	W-POP DRIVE	W-POP PUB
$ln(\alpha)$	-0.669 **	-0.659 **	-0.674 **	-0.667 **	-0.650 **	-0.638 **	-0.653 **
	(0.213)	(0.216)	(0.217)	(0.217)	(0.214)	(0.215)	(0.211)
β_U	0.685 ***	0.680 ***	0.680 ***	0.686 ***	0.673 ***	0.680 ***	0.685 ***
	(0.041)	(0.043)	(0.044)	(0.043)	(0.040)	(0.041)	(0.040)
β_V	0.333 ***	0.337 ***	0.337 ***	0.331 ***	0.343 ***	0.337 ***	0.331 ***
	(0.041)	(0.043)	(0.044)	(0.043)	(0.040)	(0.041)	(0.040)
$\gamma_{unemp25}$	-0.003	-0.007	-0.008	-0.007	-0.005	-0.006	-0.004
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
γ_{cons}	0.024	0.020	0.021	0.020	0.017	0.016	0.017
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
γ_{highSE}	0.017	0.020	0.021	0.021	0.021	0.017	0.022
	(0.024)	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)
γ_{medSE}	0.067 *	0.071 *	0.073 *	0.071 *	0.071 *	0.072 *	0.069 *
	(0.030)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)
$\delta_U (lagged)$	-0.121 *	-0.053	-0.032	-0.047	-0.077 ^	-0.073	-0.095 **
	(0.061)	(0.049)	(0.038)	(0.034)	(0.039)	(0.048)	(0.032)
$\delta_V (lagged)$	0.122 *	0.053	0.032	0.047	0.077 ^	0.072	0.095 **
	(0.061)	(0.049)	(0.038)	(0.034)	(0.039)	(0.048)	(0.032)
logLik	291.858	289.976	289.752	290.399	292.025	291.231	293.883
AIC	-563.716	-559.951	-559.504	-560.798	-564.049	-562.462	-567.767
BIC	-533.880	-530.115	-529.668	-530.962	-534.213	-532.626	-537.931
N. obs.	146	146	146	146	146	146	146

Notes: Presented are number of matches, vacancies and applicants per profession group averaged over all reported unites; N.obs denoting the number of non-missing observations; data sources: BIBB-Survey 2023, INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, own calculations;

Table A5

Matching models with best-fitting agglomeration indices and best-fitting spillovers.

Spillover	W-POP-DIST			W-COM-VET-I			W-POP-PUB		
Agglomeration	CITY SHARE	POP DIST	POP PUBLIC	CITY SHARE	POP DIST	POP PUBLIC	CITY SHARE	POP DIST	POP PUBLIC
$ln(\alpha)$	-0.622 **	-0.531 *	-0.497 *	-0.686 **	-0.624 **	-0.586 **	-0.629 **	-0.540 *	-0.509 *
	(0.209)	(0.211)	(0.209)	(0.214)	(0.216)	(0.215)	(0.206)	(0.209)	(0.207)
β_U	0.689 ***	0.674 ***	0.672 ***	0.696 ***	0.683 ***	0.679 ***	0.704 ***	0.684 ***	0.681 ***
	(0.040)	(0.039)	(0.038)	(0.041)	(0.041)	(0.041)	(0.040)	(0.039)	(0.039)
β_V	0.323 ***	0.325 ***	0.324 ***	0.317 ***	0.321 ***	0.321 ***	0.308 ***	0.316 ***	0.316 ***
	(0.040)	(0.039)	(0.039)	(0.041)	(0.041)	(0.040)	(0.040)	(0.039)	(0.039)
$\gamma_{unemp25}$	-0.014	-0.016 ^	-0.015 ^	-0.015	-0.017 ^	-0.017 ^	-0.014	-0.014	-0.014
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)
γ_{cons}	0.043 ^	0.046 ^	0.052 *	0.039	0.040	0.047 ^	0.044 ^	0.044 ^	0.051 *
	(0.025)	(0.025)	(0.025)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
γ_{highSE}	0.013	0.005	-0.003	0.024	0.017	0.009	0.015	0.006	-0.002
	(0.024)	(0.024)	(0.024)	(0.025)	(0.026)	(0.026)	(0.024)	(0.024)	(0.024)
γ_{medSE}	0.065 *	0.053 ^	0.045	0.075 *	0.066 *	0.058 ^	0.064 *	0.051 ^	0.044
	(0.030)	(0.030)	(0.030)	(0.030)	(0.031)	(0.031)	(0.029)	(0.030)	(0.030)
$\gamma_{Agglomeration}$	0.014 **	0.014 **	0.019 ***	0.011 *	0.011 *	0.016 **	0.014 **	0.014 **	0.019 ***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$\delta_U\,(lagged)$	-0.070 ^	-0.055	-0.051	-0.149	-0.121	-0.112	-0.087 **	-0.079 *	-0.070 *
	(0.039)	(0.039)	(0.038)	(0.096)	(0.097)	(0.095)	(0.031)	(0.031)	(0.031)
$\delta_V\,(lagged)$	0.070 ^	0.055	0.050	0.149	0.120	0.110	0.087 **	0.079 *	0.069 *
	(0.039)	(0.039)	(0.039)	(0.096)	(0.098)	(0.096)	(0.031)	(0.032)	(0.032)
logLik	295.894	296.911	298.831	294.687	295.524	297.251	298.072	298.393	299.984
AIC	-569.788	-571.823	-575.662	-567.373	-569.047	-572.502	-574.145	-574.786	-577.969
BIC	-536.968	-539.003	-542.842	-534.554	-536.227	-539.682	-541.325	-541.966	-545.149
N. obs.	146	146	146	146	146	146	146	146	146

Notes: Presented are regression estimates for base, regional and agglomeration modeling; S.E. in parentheses, significance level: ^p< 0.1, *p < 0.05, **p < 0.005, ***p< 0.001; Data sources: BIBB-Survey, INKAR, IÖR-Monitor, Wikidata, Google Distance Matrix API, and own calculations.

Data Availability

The survey data supporting the findings of this study are available online at https://www.bibb.de/de/179138.php. Additional regional data sources and their origins are documented in the manuscript.

Code Availability

The repository at https://github.com/DennisKubitza/MatchMeUp contains the code used to prepare the datasets and perform the analyses presented in this paper.

This working paper was authored for Skills2Capabilities by Dennis Oliver Kubitza (Federal Institute for Vocational Education and Training (BIBB); Research Centre for Education and the Labour Market (ROA), Maastricht University). This paper is a deliverable from the work package 4 entitled "Challenges and changes in the demand for VET Skills", led by BIBB.

This working paper represents the views of the authors based on the available research. It is not intended to represent the views of all Skills2Capabilities affiliates.

www.skills2capabilities.eu

Skills2Capabilities Partner Institutions:

3s

bibb Federal Institute for Vocational Education and Training

dSEA Dipartimento di Scienze Economiche e Aziendali 'Marco Fanno'

Fafo

Institute of Philosophy and Sociology

KRIVET

Maastricht University ROA

Tallinn University

tu dortmund university

University of Jyväskylä Finnish Institute for Educational Research

Warwick Institute for Employment Research IER

Skills2Capabilities

Co-funded by the European Union