Research Institute for the Evaluation of Public Polices

http://irvapp.fbk.eu

FBK-IRVAPP Working Paper series

Research Institute for the Evaluation of Public Policies

IRVAPP
RESEARCH INSTITUTE FOR THE EVALUATION OF PUBLIC POLICIES
FONDAZIONE BRUNO KESSLER

The demand for language skills in the European labour market: Evidence from online job ads

Gabriele Marconi

Loris Vergolini

October 2022

FBK-IRVAPP Working Paper No. 2022-08

The demand for language skills in the European labour market: Evidence from online job ads

Gabriele Marconi
Sogeti
gabriele.marconi@sogeti.lu

Loris Vergolini
University of Bologna & FBK-IRVAPP
loris.vergolini@unibo.it

FBK-IRVAPP Working Paper No. 2022-08

October 2022

Research Institute for the Evaluation of Public Policies
Bruno Kessler Foundation
Vicolo dalla Piccola 2, 38122 Trento (Italy)

Phone: (+39) 0461.314209
Fax: (+39) 0461.314240

E-mail: info@irvapp.it
Website: http://irvapp.fbk.eu

The demand for language skills in the European labour market: Evidence from online job ads

Gabriele Marconi & Loris Vergolini

Abstract

We investigate foreign language skill demand and its determinants with a novel dataset, the Web Intelligence Hub's Online Job Advertisement (OJA) database, with information on about 53 million ads posted in 2021 for jobs in Europe. This unique dataset has been built crawling hundreds of job search engines and websites of public employment services, allowing us to identify foreign language requirements in OJAs at the NUTS-3 regional level. Moreover, we analyse how the demand for foreign languages varies at occupational level in the European countries as well as the possible macro factors (GDP, population density; participation rate in education and training; percentage of people employed in the high-tech sector and in the touristic sector) that could influence the request for foreign languages.

JEL-Code: J20, J24, R10

Key words: Language skills, Labour market, Occupational groups, NUTS, Online job ads, Eurostat, English, German, Chinese, French, Spanish, big data, web scraping.

1. Introduction

Foreign language skills yield important benefits for people as well as for economies. Speaking more languages is believed to improve cognitive competencies like alertness and creativity (Bak et al., 2016; Woll and Wei, 2019), as well as improving intercultural understanding and active citizen participation on global issues (Gudykunst, 2003; Della Chiesa et al., 2012). Last but not least, speaking foreign languages is believed to improve employability and individual economic outcomes (Araújo et al., 2015; Isphording, 2015). The importance of some of these benefits is likely to have increased in the last decades, when the interaction among people from different cultures and countries was intensified by globalisation, migration and technological innovation (OECD, 2021). The economic benefits of foreign language skills are particularly clear in linguistically diverse economies. They have been underlined by the Council of the European Union stating that "language competences contribute to the mobility, employability and personal development of European citizens, in particular young people, in line with the objectives of the Europe 2020 strategy for growth and jobs" (Council of the European Union, 2014, p.2).

This paper sheds new light on the economic benefits of foreign language skills by illustrating and analysing the regional distribution of foreign language skill demand in the European labour market. There are two notable studies on foreign language skill demand in national labour markets. Antonietti and Loi (2014) combine three sources of survey and administrative data to analyse foreign language skill demand in the Italian manufacturing sector in 2004. They find a very high foreign language skill demand in high-skilled occupations (over 50% of the vacancies require knowledge of at least one foreign language), but a much lower demand for low-skilled occupations (down to less than 5% for what they label as "elementary occupations"). Fabo et al. (2017) look at foreign language skill demand as recorded in the tags of online job ads in the Visegrad countries (Czechia, Hungary, Poland and Slovakia). They use 1 job portal for each country, and find that a substantial proportion of ads requires knowledge of foreign languages, particularly English. Between 28% (Czechia) and 69% (Poland) of ads refer to the English language in their dataset, and ads requiring English also seem to command a higher salary, on average.

Compared to the previous studies by Antonietti and Loi (2014) and Fabo et al. (2017), ours covers a much wider range of countries (all European countries plus the UK) and languages (with a focus on English, German, Chinese, and to a minor extent Spanish and French). This paper presents detailed regional statistics (NUTS level 3), something which was not done by previous authors. It uses a different data source than Antonietti and Loi (2014); compared to Fabo et al. (2017), it has a much more extensive coverage, with between 300 and 400 job portals analysed instead of one per country (we use 53m ads covering all EU countries and the UK, compared to 74000 ads in four countries). This is made possible by the use of a novel data source, the Web Intelligence Hub's Online Job Advertisement (OJA) database jointly developed by Eurostat and Cedefop (Ascheri et al., 2021).

In addition to the distribution of the demand of foreign languages across Europe, we present a set of preliminary analysis regarding the association of foreign language skill requirements with skills and occupational groups; and with regional characteristics (GDP, population density and labour market characteristics). In the first case, the aim is to characterise the ads that require the knowledge of some foreign languages, while in the second one the rationale is to look at the possible determinants in the demand for foreign languages at regional level.

Our main findings show that:

- English finds a place as one of the great transversal skills in the European labour market. It is the sixth most-requested skill, at a similar level as "use a computer" or "communication" (out of 13 000 skills in the ESCO classification); and its presence among the job requirements is best predicted by transversal skills like "tolerate stress" or "adapt to change". 28% of online job ads require some knowledge of English in non-English speaking European countries.
- After English, the most demanded foreign languages in Europe are German, Spanish, Chinese and French. Some 2% of online job ads require German (excluding German speaking countries), and 1% of ads require Chinese. Spanish and French are the best predictors of German language skill requirements, indicating that these languages are often required within the same job.
- English is most in demand in high-skill occupations: for example, over 50% on online ads for manager jobs require English. The converse is true for Chinese, which is more in demand for medium- and low-skill occupations.
- Population density is positively associated with the demand for both English and Chinese, but once this is accounted for, the demand pattern diverge. English demand is positively associated with GDP per capita, while Chinese demand is negatively associated with the share of employment in the high-tech sector. Perhaps surprisingly, the share of employment in the tourism sector is not strongly associated with the share of online job ads requiring English or Chinese.

The data we use is described in the next section. Section 3 highlights the most commonly demanded languages in the European labour market, while Sections 4, 5 and 6 present data on the demand for English, German and Chinese language skills, respectively (some data for French and Spanish, presenting a data pattern similar to German, are presented in Appendix B). Section 7 presents the data for ads where the knowledge of at least one language is explicitly required. Section 8 and Section 9 presents data on the association between English, German and Chinese requirements with other skills and with occupational groups, respectively. Section 10 shows the relationship between the demand for English and a set of macro characteristics at NUTS-3 and NUTS-2 level (with results for German, French and Spanish presented in Appendix D and E), and Section 11 concludes the paper.

2. Data

This is paper uses of a novel data source jointly produced by Eurostat and Cedefop, the Web Intelligence Hub's Online Job Advertisement (OJA) database, of which we use the 2021q4 release. This dataset contains information on 162m ads for jobs in EU countries and the UK between July 2018 and December 2021 (53m ads refer to 2021, the reference year used in this paper). It is based on the systematic crawling of several hundred data sources, for example job search engines (e.g. Indeed, Jooble) and websites of public employment services. The dataset and the procedures used for data collection are described in Cedefop (2019) and Ascheri et al. (2021). Cedefop (2022) made available some summary dashboards describing the skill data.

The OJA dataset contains a number of variables, including for example education required, the name of the entity that posted the ad, the salary requirement when available, etc. In this paper, we use information on skill requirements (at the most granular level), location (NUTS 3 level) and occupation (ISCO 3-digit level). Information on occupation is available for 100% of ads, while location is available at the NUTS 3 level only for 78% of ads. This means that 22% of the ads are discarded from the analysis because detailed geographical information is not available. There are also ads (11%) that do not have explicit skill requirements at the most granular skill level. These are included in the analysis as not requiring any skill.

Most categorical variables are classified based on an ontology matching model. This means that each category of each variable (for example, the category "accountant" in the occupation variable) is associated to one or more tags (e.g. "accountant", "accounting analyst" and "internal auditor"). When a tag referring to an occupation is found in the ad text, the ad is classified in that category. In the case of occupation (one of the variables used in this paper), ads that are not classified through the ontology model are classified through a machine learning model, with an estimated accuracy of 80% (Ascheri et al., 2021). For location and skills, other variables used in this model, only the ontology matching is used. This means that, for example, job ads that are classified as demanding the skill "English" are those in which "English" is found within the ad.

The kind of automatic classification that is used in big data classification is prone to measurement error (AAPOR, 2015). Accordingly, classifications based on ontology matching have some limitations. One important limitation is that tags could be identified out of context. For example, a job ad stating that the position requires "coordination with our English headquarters" would also be classified as requiring English. Another important source of measurement in error is for jobs in which language knowledge is not explicitly required, but it is assumed. We try to deal with this limitation by including all jobs written in a certain language (e.g., English) among the jobs requiring that language, but this is unlikely to be sufficient. For example, around 11% of jobs do not have any skill requirement recorded in the dataset, even though it is natural to expect that some skills will be required in those jobs.

3. The most demanded languages among online job ads in Europe

Table 1 reports the number of ads explicitly requiring knowledge of each language that is found in European online job ads data. English is by far the most requested language with 11 million ads requiring it in 2021 (22% of all ads). To give an idea of the importance of English in the labour market, it is the sixth most-required skill in online job ads among the about 13 000 skills classified in the ESCO classification. It comes after “adapt to change”, “work in teams”, “use a computer”, “teamwork principles” and “use Microsoft office”; and before “assist customers”, “create solutions to problem” and “communication”. This places English among the most widely requested transversal skills in the labour market. The second most required language is German, with around 900 000 ads, followed by Spanish (800 000), Chinese (700 000 ads) and French (600 000). Basque fares surprisingly well (400 000 ads require it), perhaps because it is not known by many people within the Basque countries themselves, implying that jobs requiring its knowledge always need to make it explicit.

The following sections looks more in detail at the geographic distribution of the proportion of ads requiring English, German and Chinese; and of those requiring at least one of the languages outlined in Table 1. The same charts for Spanish and French are reported in Appendix B. All the charts have been generated with the help of the package “Eurostat” for R, version 3.1 (Lahti et al., 2017), after preparing the data in Base R (R Core Team, 2021). The data has been queried from its repository in the cloud using the packages “noctua” (Jones, 2021) and “DBI” (R-SIG-DB R Special Interest Group on Databases, Wickham, & Müller, 2021), following the procedure outlined in Ascheri et al. (2022).

Table 1 - Number of ads explicitly requiring knowledge of a language, 2021

Language	Ads	Language	Ads
English	11383304	Russian	11310
German	892281	Czech	8382
Spanish	812593	Danish	5265
Chinese	659671	Hungarian	3879
French	604827	Greek	2785
Basque	417394	Icelandic	2221
Dutch	374298	Slovak	1293
Arabic	244481	Croatian	1113
Finnish	235692	Turkish	1024
Italian	193635	Romanian	968
Polish	78858	Slovenian	921
Welsh	75368	Bulgarian	843
Norwegian	70239	Bihari	520
Swedish	35318	Portuguese	416
Latvian	11603	Maltese	357

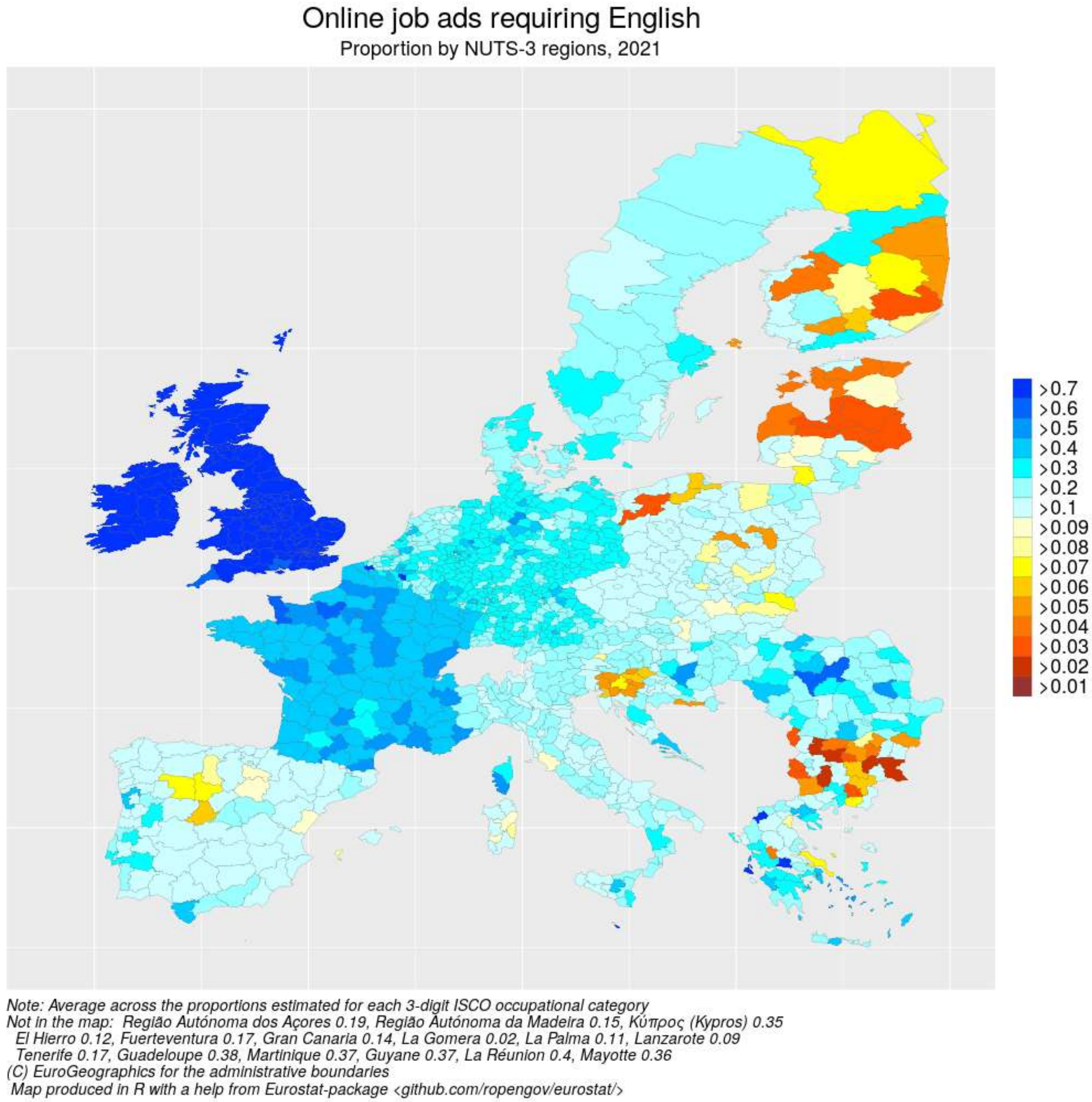
Note: Un-weighted ad count across European Union countries and the UK.

4. The demand for English language skills

Figure 1 shows the proportion of online job ads requiring English, by far the most demanded language on the European labour market. This proportion combines job ads that explicitly require English and job ads that are written in English. Therefore, the data in Figure 1 are not directly comparable with those seen in the previous section, but they give a more complete picture of the demand for English in the labour market.

Following Ascheri et al. (2021), Figure 1 reports the average proportion across ISCO occupational categories (3-digit), to limit the problem of selectiveness of online job ad data (i.e., the fact that some occupations are likely over-represented among online ads). The results are not very dissimilar when calculating an overall proportion without averaging across occupations, as done in Annex A. The tables underlying all the charts in this paper are available in the GitHub repository created by Marconi and Vergolini (2022).

Figure 1



On average across European regions, 37% of online job ads (and 28% of all those outside Ireland, Malta and the UK) required some knowledge of English in 2021. For the analysis in this section, we classify an ad as “requiring some knowledge of English” if it either explicitly mentions English among the job requirements, or is written in English. Excluding Ireland, Malta and the UK, the proportion of ads requiring English varies between 1.5% in La Gomera (Spain) and 95% in Ithaki-Kefallinia (Greece). Perhaps surprisingly, the country with the highest proportion of online job ads requiring English is France (47%), followed by Luxembourg (42%) (see also Table 2). Besides the difference across countries, no obvious regional pattern emerges at a first glance. Jobs in coastal and big urban areas do not have visibly higher requirements in terms of English language skills.

Table 2 - Percentage of ads requiring English, German or explicitly requiring at least one language, by country

Country	English	German	Spanish	Chinese	French	At least one language explicitly required
AT – Austria	22.09	90.63	0.74	0.18	0.45	17.44
BE – Belgium	33.69	6.12	0.63	0.08	38.63	25.5
BG – Bulgaria	8.86	3.82	0.35	0.00	0.28	1.21
CY – Cyprus	35.15	4.15	0.27	0.00	0.35	0.00
CZ – Czechia	15.47	1.92	0.20	0.00	0.12	11.16
DE – Germany	33.25	96.99	1.14	1.50	0.35	31.65
DK – Denmark	30.47	0.21	0.03	0.00	0.07	20.12
EE – Estonia	8.13	2.30	0.08	0.00	0.04	1.60
EL – Greece	31.48	3.62	1.10	0.00	1.97	2.94
ES – Spain	14.16	0.95	87.6	0.00	0.56	6.42
FI – Finland	11.36	0.43	0.02	0.00	0.02	0.36
FR – France	47.55	1.49	1.06	1.20	94.36	45.32
HR – Croatia	22.36	1.18	9.82	0.00	1.00	11.13
HU – Hungary	26.20	3.62	0.77	0.00	0.63	15.17
IE – Ireland	96.54	0.99	0.57	0.55	1.12	1.95
IT – Italy	19.61	1.54	0.94	0.40	1.49	15.60
LT – Lithuania	11.53	0.36	0.26	0.00	0.00	8.02
LU – Luxembourg	41.78	14.51	1.24	0.00	49.47	6.05
LV – Latvia	4.72	0.05	0.00	0.00	0.00	2.49
MT – Malta	90.14	7.68	0.16	0.00	0.01	0.00
NL – Netherlands	27.52	2.56	0.34	0.17	0.62	17.76
PL – Poland	15.95	2.07	0.98	0.07	0.33	8.54
PT – Portugal	22.40	3.60	17.32	0.07	1.14	17.57
RO – Romania	30.47	2.07	1.65	0.01	0.98	14.32
SE – Sweden	25.84	0.24	0.18	0.03	0.02	20.47
SI – Slovenia	6.83	1.44	0.01	0.00	0.02	4.02
SK – Slovakia	11.73	1.92	0.12	0.00	0.78	3.29
UK – United Kingdom	98.37	1.19	0.39	2.36	1.27	7.92
EU cross-country average	22.34	2.75	1.54	0.16	3.86	11.49

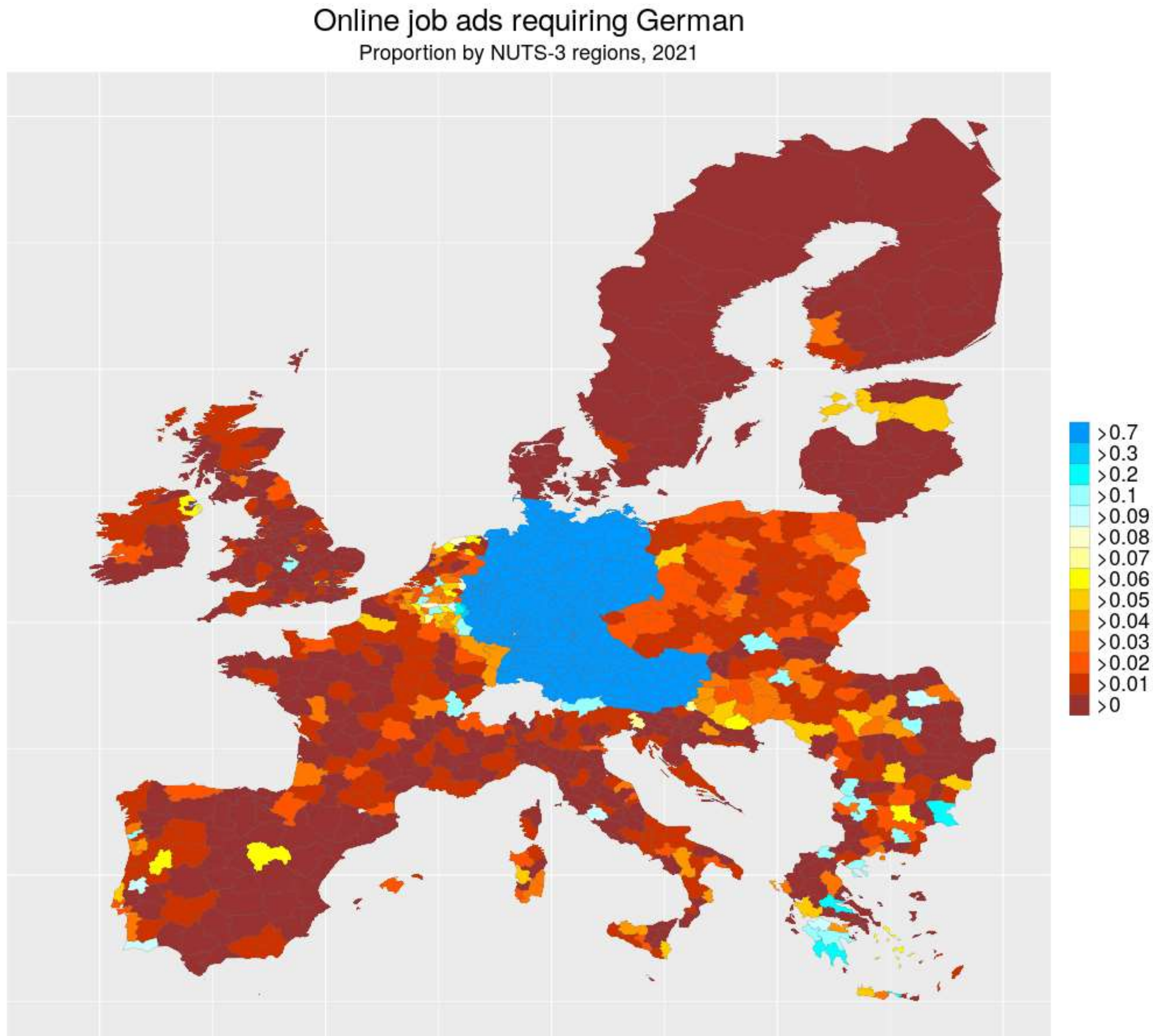
Notes: The "EU cross-country average" excludes the UK and also, for each column, countries with values higher than 80% (i.e. Ireland and Malta for English, Austria and Germany for German, Spain for Spanish and France for French). "English", "German", Spanish" and "French" include both ads explicitly requiring this language and implicitly requiring it (because they are written in that language); "At least one language explicitly required" includes only ads that explicitly require one of the languages listed in Table 1. "Chinese" includes only ads that explicitly require this language.

5. The demand for German language skills

German is required by 33% of ads, on average across European regions, but this proportion drops to just 2% once Austria and Germany are excluded. Outside Germany and Austria, regions with a high proportion of ads requiring German typically reside close to the border with these two countries and in countries where German is an official language like Belgium and Luxembourg (Figure 3). In addition, some touristic regions such as Irakleio (Greece) and Burgas (Bulgaria) also have around one quarter of ads requiring German. This is also visible at the country level (Table 1), where excluding countries with German as an official language, the highest proportion of ads requiring German is observed in Malta (8%) and Portugal (4%).

Figure 3 exemplifies a pattern common to all European languages with the exception of English (see e.g. the charts for Spanish and French in Annex B and Table 2). Demand for the language skills is mostly found implicitly within the countries speaking this language, but it is much lower outside, with the notable exception of some touristic regions. In other words, knowing a language like German or French has the considerable implication of "unlocking" one or more labour markets in which it is widely used, but it is not a common requirement outside them.

Figure 2

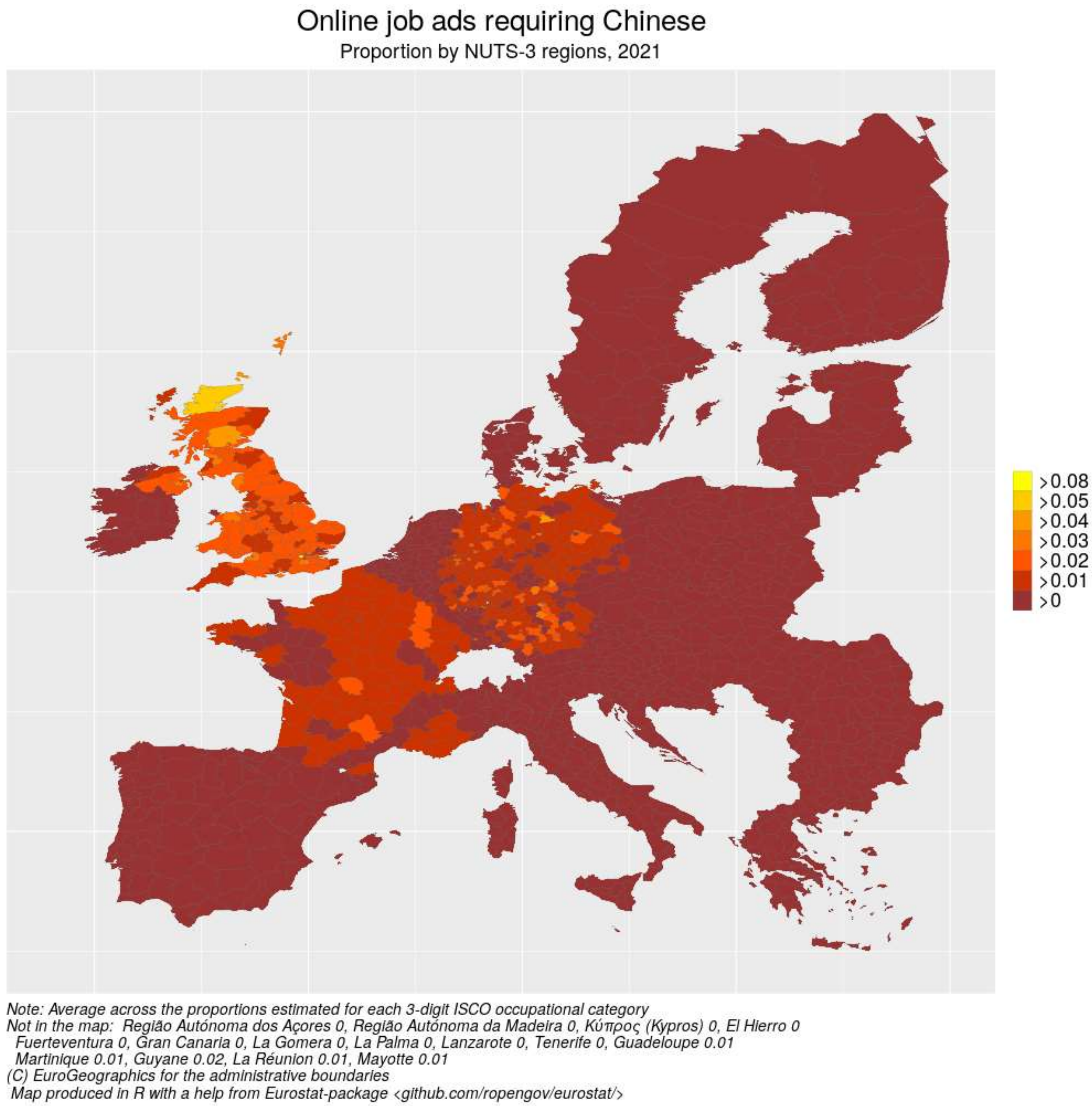


Note: Average across the proportions estimated for each 3-digit ISCO occupational category
Not in the map: Região Autónoma dos Açores 0.01, Região Autónoma da Madeira 0.1, Κύπρος (Kypros) 0.04
El Hierro 0.01, Fuerteventura 0.01, Gran Canaria 0.01, La Gomera 0.04, La Palma 0.01, Lanzarote 0.01, Tenerife 0.01
Guadeloupe 0.01, Martinique 0.01, Guyane 0.02, La Réunion 0.01, Mayotte 0
(C) EuroGeographics for the administrative boundaries
Map produced in R with a help from Eurostat-package <github.com/ropengov/eurostat/>

6. The demand for Chinese language skills

The demand for Chinese (Figure 3) follows a different pattern than large European languages like French and German, for the simple reason that Chinese is not an official language in any European country. The percentage of online job ads requiring some knowledge of Chinese is equal to 1% on average across European regions, and it ranges from 0% in various European regions to 10% in Hounslow and Richmond upon Thames (UK). Among European countries, it is highest in France, Germany and the UK, which are also the only ones where this proportion exceeds 1%, on average across all regions in the country.

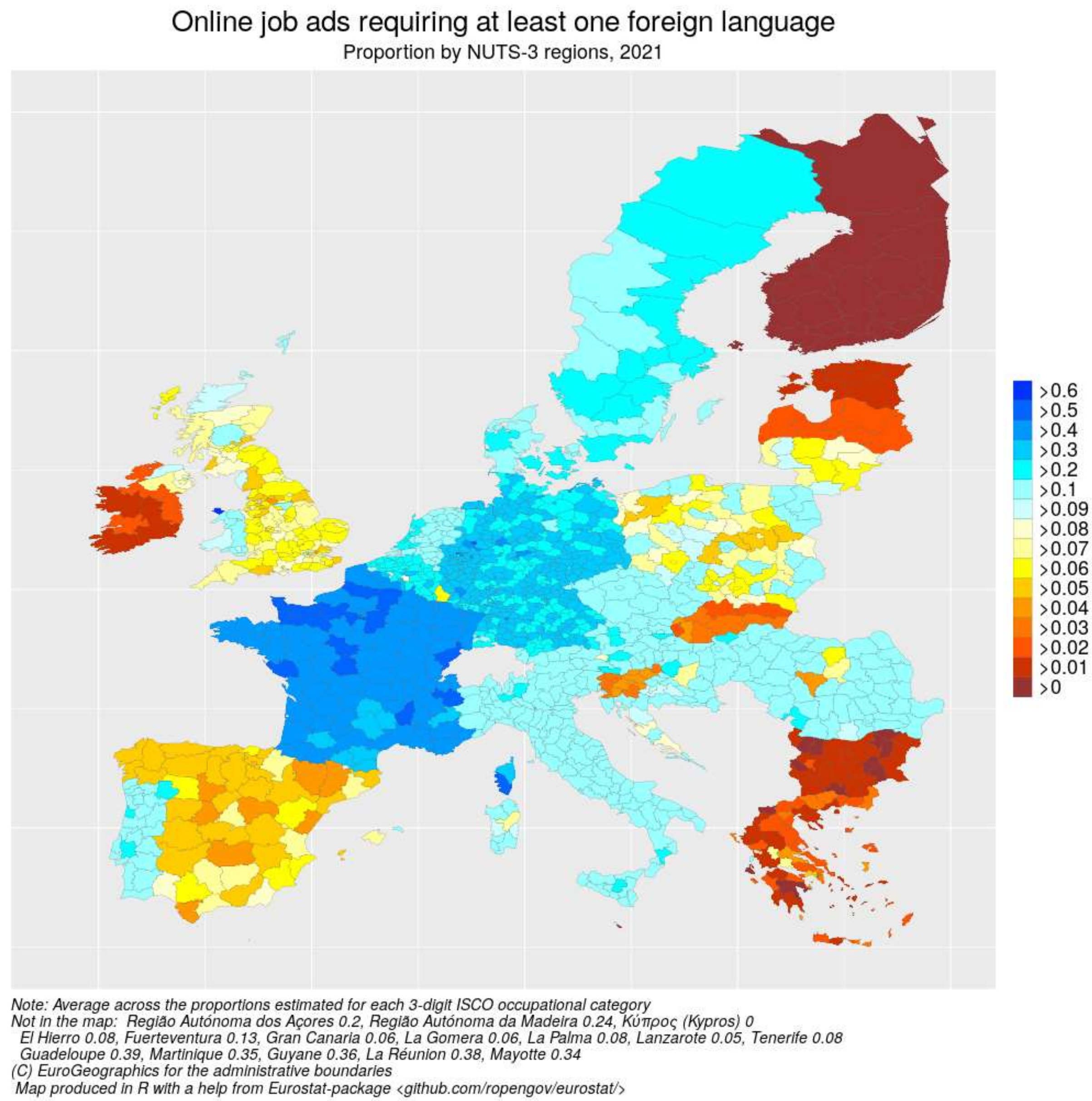
Figure 3



7. **Ads explicitly requiring at least one language**

Figure 4 shows the proportion of online job ads explicitly requiring at least one language in European regions. This is the closest proxy to requiring at least one foreign language, because this explicit requirement implies that the required language cannot be taken for granted in the ad's labour market. Across European regions, the average proportion of online job ads explicitly requiring at least one language is 20%. This proportion ranges from close to 0% in various European regions to 67% in the Isle of Anglesey (UK).

Figure 4



However, the patterns depicted in Figure 4 may also reflect some difficulties inherent in identifying what qualifies as a "foreign language" in an ad. Countries like Finland, Cyprus and Malta have close to 0% of ads requiring a foreign language (Table 1), according to this measure. This could be because, for example, in these countries many ads requiring English are directly written in English; Or because the knowledge of multiple languages can be assumed in some context, such as multilingual Malta.

8. Association of foreign language skills with other skill requirements

Figure 5 shows the results from three decision tree models (James et al., 2021) using skill requirement information to predict whether an ad explicitly requires English, German or Chinese. Decision tree models are machine learning models that split a sample based on some variables to predict an outcome variable (in this case, foreign language skill requirements). In this paper, the purpose of applying the machine learning model is descriptive, rather than predictive: the goal is to show what are the main predictors of observing English, German or Chinese as a skill requirement in an ad. Therefore, the models have been trained on the large majority (90%) of ads with information on skill requirements posted in 2021, i.e. just over 40 million ads. A smaller proportion of ads (10%) has been used to evaluate the models.

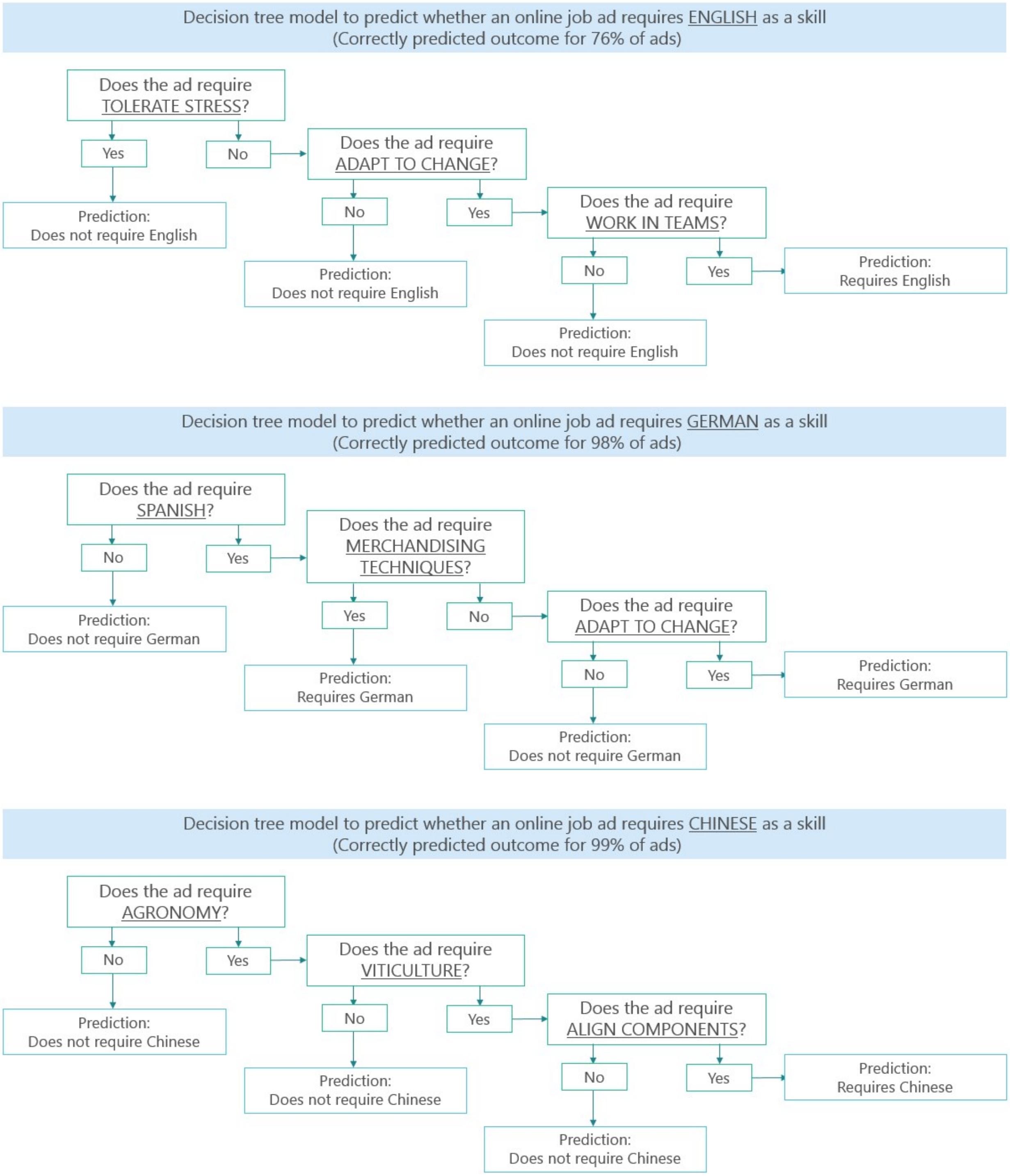
The decision tree models trained for the explorative analysis of this section have three nodes, implying that three skills are used as predictors for the outcome. At each node, the model browses through all skills in the ESCO classification and chooses the best predictor of the outcome based on the resulting weighted Gini impurity index recorded in the two resulting subsamples. Among the two resulting subsamples in which the data has been split, the model chooses the one with highest level of the Gini impurity index and splits its further.

For example, tolerate stress is chosen as the best predictor for English based on the fact that the weighted impurity index of the two resulting subsamples is 0.356 (on a scale from 0 to 1), the lowest among all potential predictors. The association in this case is negative: ads requiring tolerate stress as a skill are less likely to require English than other ads, and the prediction of the model goes accordingly. The model goes on splitting the subsample of ads that do not require tolerate stress as a skill, because the Gini impurity index of this sample is higher (roughly meaning that this subsample contains proportions of ads requiring and not requiring English more in line with the whole dataset). Next, the model finds a positive association between adapt to change and English, in the sense that ads that do not require adapt to change are less likely to require English than other ads (conditional on not requiring tolerate stress). Finally, the model splits the subsample of ads that require adapt to change as a skill, and finds work in teams as the best predictor for English (ads requiring work in teams are more likely to also require English).

The model for English has an accuracy rate of 76%. This means that 76% of the ads are either true positives (the model correctly predicts that the ad requires English) or true negatives (it is correctly predicted that the ad does NOT require English). This compares to an accuracy rate of 63% for a random prediction model classifying an ad as requiring English with a probability equal to the observed proportion in the sample. Focusing on the ads requiring English, we find that the model correctly classifies 39% of them (recall rate).

Overall, the results from the decision tree model confirm that English is associated with the presence (or absence) in an ad of widely demanded transversal skills (see Goggin et al., 2019 and Osmani et al., 2019 for a discussion of transversal skills and their role in the labour market). Combined with its position among the most required skills in the dataset (Section 3), this suggests that English finds its position among the highly-demanded transversal skills in the European labour market.

Figure 5 - Decision tree prediction model for foreign language requirements



Note: The model has been trained on a random sample of 90% of the 46'976'223 online job ads posted in 2021 across all countries in the dataset with information on skill requirements; and it has been evaluated on 10% of these ads.

The case of German is quite different from that of English. As shown in Figure 5, the most important predictor of German is Spanish. This shows that German is often demanded in jobs that require a variety of language skills, for example because of frequent communication with foreigners of different origin. Even though it is not shown in Figure 5, it is relevant to notice that if ads requiring Spanish are excluded, then the best predictor becomes French; and if ads requiring French are also excluded from the analysis, then the best predictor becomes Dutch. This fact confirms the conclusion that German is relatively likely to be observed in jobs requiring other languages; and it suggests that the same conclusion holds for French and Spanish.

The second predictor identified by the model for German is merchandising techniques, a relatively specific skill compared to those that best predicted English. Only in the third node we find a transversal skill (which is adapt to change, the most widely demanded skill in the dataset).

The accuracy rate for German is 98%. However, this high level of accuracy is somehow artificial because German is required only by a small proportion of ads (see previous sections), so it is easy to predict negative outcomes (i.e., ads not requiring German). Indeed, a random prediction model already has an accuracy of 96% (the differences in this analysis are always significant at any confidence level, because of the large number of observations). The recall rate (the proportion of correctly predicted positive outcomes) is probably a better performance measure in this case. This is equal to 35%, close to the rate observed for English.

The results for the Chinese model bring to different conclusions than both English and German. The three best predictors are very specific skills (agronomy, viticulture and align components), but they are all found by the model to be negatively associated with Chinese. Since the predicting skills are quite rare, this means that the model predicts that virtually no ads require Chinese. The accuracy rate of 99% is, like in the case of German, artificially high because a random prediction model already scores 97%. The recall rate, in contrast, is extremely low: 0.06%, compared to over 30% for English and German. Therefore, the decision tree model's results suggests that no skill a good predictor of which ads require Chinese.

9. The demand for foreign languages across occupations

In this section we show the demand for foreign languages within the occupational groups from the ISCO classification (ILO 2012). The major groups (at the 1-digit level) of the ISCO classification are defined as follows: managers; professionals; technicians and associate professionals; clerical support workers; service and sales workers; skilled agricultural, forestry and fishery workers; craft and related trades workers; plant and machine operators, and assemblers; elementary occupations. In this and the following sections, we will focus on English and Chinese, while the results for the other languages are reported in Appendix C. The focus on these two languages is motivated by the fact that, as highlighted in the previous sections, they have very interesting profiles: English as the most foreign language in highest demand, which should be even more widely requested in high skills occupations (i.e., managers, professionals and technicians); and Chinese as a language with a distinct demand pattern across Europe, perhaps linked with international trade or with migration patterns.

Figure 6 shows the percentage of ads requiring English language skills in the high-skill occupational groups in the ISCO classification (i.e. managers, professionals, technicians and associate professionals, which require skill level 3 and 4; ILO, 2012). Figure 7 reports the same indicator for the low-skill occupations (requiring skill levels 1 or 2). In these figures, for each country the bar shows the percentage of online ads that require English in the specific occupation and a red dot that represents the national average across all occupations. The first result is that English is even more widely required for the top occupations. For example, English is required in more than 50% of the online ads in 13 countries for managerial occupations, in 10 countries for professionals and in 6 countries for technicians. If we shift the focus on the medium and low skill occupations (Figure 7), the percentage of ads requiring English is close to the average across occupations in most countries only for clerical support workers. Among the other occupations, the knowledge of English is a much less relevant skill (with the exception of service and sales workers in a few countries). This result is coherent with the findings of Isphording (2015) and Fabo et al. (2017). In fact, they stress that the command of a foreign language is linked to labour market success in terms of wage and employability, and we look at the request of English in the high skill occupations that ensure material and symbolic (i.e., prestige) rewards.

For managers and professionals, the highest percentage of online ads requiring English is found in Luxembourg (about 80%), while for technicians the leading countries are Croatia (66.6%), Cyprus (65.1%) and Luxembourg again (61.7%).[1] Other countries that are found at the top for the high skill occupations are France, Romania and Portugal[2], while at the bottom of the ranking are found the Baltic Republics.

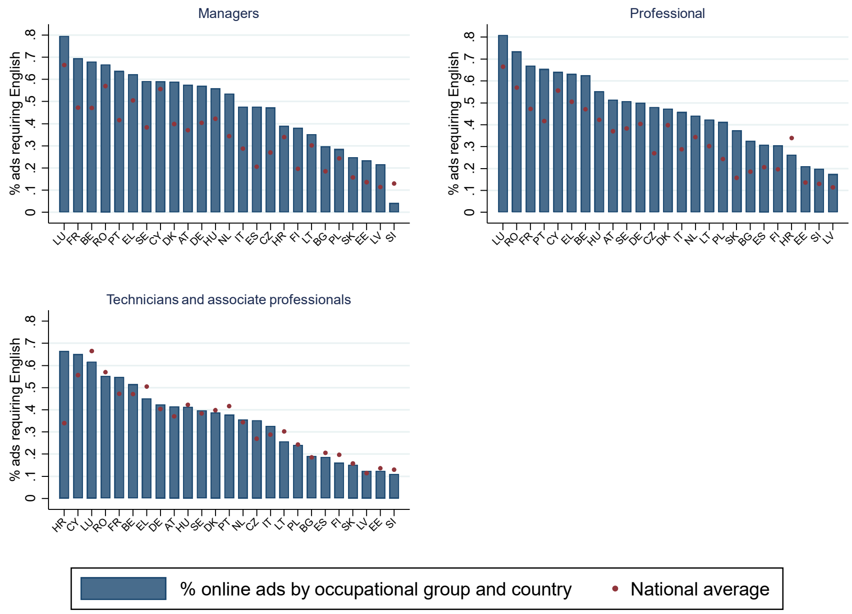
Figures 8 and 9 report similar results for the percentage of online ads requiring Chinese. In this case the picture is almost reversed compared to the English case. In fact, Chinese is requested above all for the medium and low skill occupations. This means that the request of specific skills for Chinese is unlikely to be linked to the requirements of international trade, and that they are most likely linked to the Chinese immigration network in Europe. The country with the highest demand for Chinese in practically all the occupation groups is the United Kingdom, followed by France, and Germany. In general, Chinese is requested in very few online ads: 2.5% in the United Kingdom, 1.6% in France, 1.4% in Germany and about 1% In Ireland and Italy. These countries experienced in the last decades a huge inflow of Chinese nationals (Plewa and Stermšek 2017) that can explain the demand for the knowledge of Chinese in particular for the medium and low occupations. It has to be noted that also Spain is an important destination of the Chinese migration, but this is not mirrored in an increase in the online ads requiring Chinese.

The differences between these countries should be deeper analysed in future research for understanding if there are some relationships between the demand for the knowledge of Chinese and the characteristics of the national (or local) labour markets.

[1] In Luxembourg there is also the highest percentage of online ads requiring also French and German (see Appendix C). This is not entirely surprising, being French and German official languages in Luxembourg.

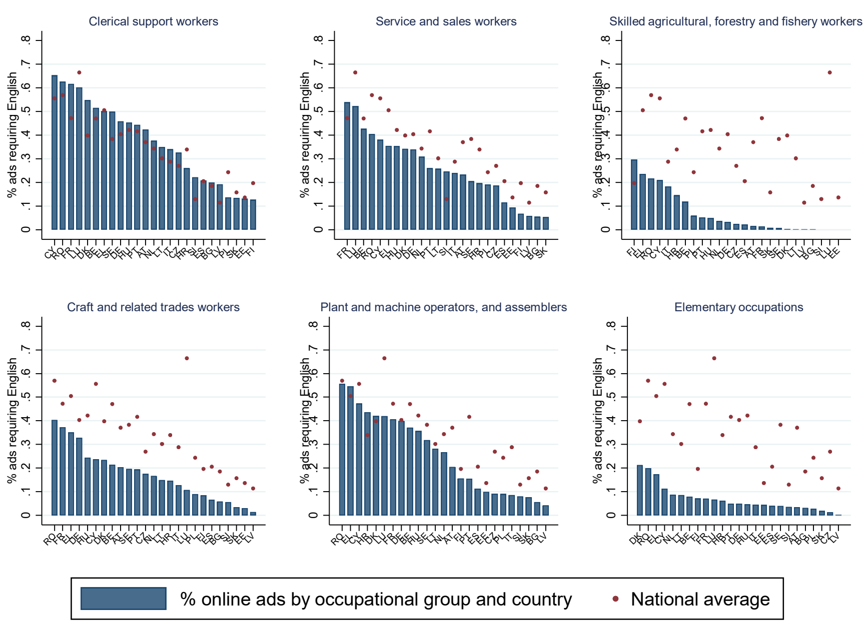
[2] It is not surprising to notice that in Portugal, the knowledge of Spanish is also widely required across all the occupations (see Appendix C).

Figure 6 - Percentage of online ads requiring English in high skill occupational groups, by country



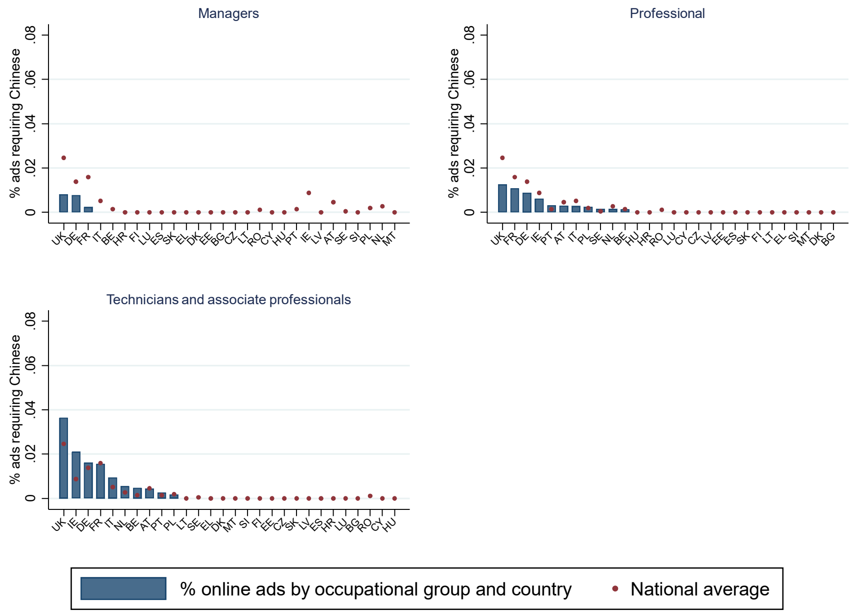
Note: UK, IE and MT are excluded. “National average” refers to the average across all occupational groups within a country.

Figure 7 - Percentage of online ads requiring English in medium-low skill occupational groups, by country



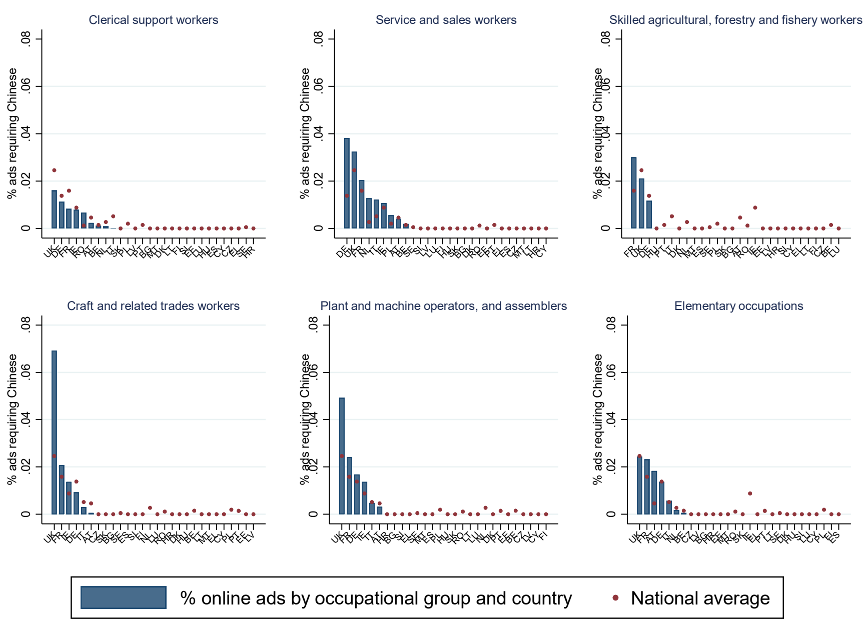
Note: UK, IE and MT are excluded. “National average” refers to the average across all occupational groups within a country.

Figure 8 - Percentage of online ads requiring Chinese in high skill occupational groups, by country.



Note: "National average" refers to the average across all occupational groups within a country.

Figure 9 - Percentage of online ads requiring Chinese in medium-low skill occupational groups, by country



Note: "National average" refers to the average across all occupational groups within a country.

10. The demand for foreign languages according to NUTS characteristics

In this section, we look at the possible macro factors that could predict the demand for English and Chinese. We start with a descriptive bivariate analysis, and then turn to a multivariate spatial autoregressive econometric model. Given the explorative nature of this exercise and the difficulty to find data series with good coverage of regions at regional level, we look only at the following characteristics (Eurostat, 2022):[3]

i) GDP per capita at purchase power parity;
ii) population density (population per squared kilometre);
iii) participation rate of 25-64 year-olds in education and training in the last 4 weeks;
iv) proportion of the population with a tertiary degree (ISCED 5-8);
v) proportion of the workforce employed in the high-tech sector;
vi) net occupancy rate of bed-places and bedrooms in hotels and similar accommodation.

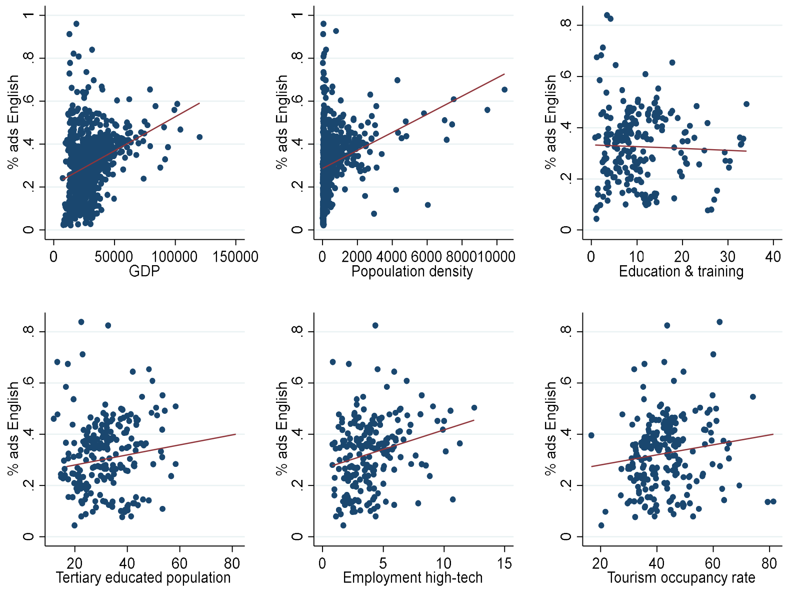
The first two variables are measured at NUTS-3 level, while the others at NUTS-2 level. GDP per capita and population density are intended to measure the level of economic prosperity of provinces and regions. The participation in education and training and proportion of the population with a tertiary degree are seen as a (crude) indicator of the human capital, while the last two variables take into account the nature of the local labour market, supposing that the high-tech and touristic sectors would require a workforce with English skills. Also in this section, we focus the attention on the demand for English and Chinese, reporting the results for the other languages in Appendix D.

Figure 10 reports the scatter plot of the relationship between the percentage of online ads requiring English and the six variables above defined. It emerges a positive and quite strong relationship with GPD, population density and the share of employment in the high-tech sector, the relationship is surprisingly weaker with the employment in the touristic sector and slightly negative with the percentage of people involved in education and training. The relationship turns positive again once the proportion of people with a tertiary degree is considered. These results are in line with the findings of Melitz (2008) and Fidrmuc (2016). They show that English is particularly relevant for enhancing international trade and, for this reason, it is more requested in the wealthiest European areas. Figure 11 shows a similar set of graphs considering the percentage of online ads requiring Chinese. In this case, all the relationships are positive. Similar to what we find for the English case, GDP, the percentage of population with a tertiary degree and the employment in the high-tech sector seem to be potential strong predictors also for the requirement of Chinese.

To sum up, the preliminary evidence reported in this section highlights a stronger demand for English and Chinese in European provinces and regions characterised by a high level of economic development and with a well-established high-tech sector. It is likely that in these geographical areas, the firms are in search of a high-skilled workforce. It is apparent that the relationships highlighted in the figures are non-linear, with clusters of outliers that could hide specific geographical patterns. Future analysis should attempt to look more closely at these patterns.

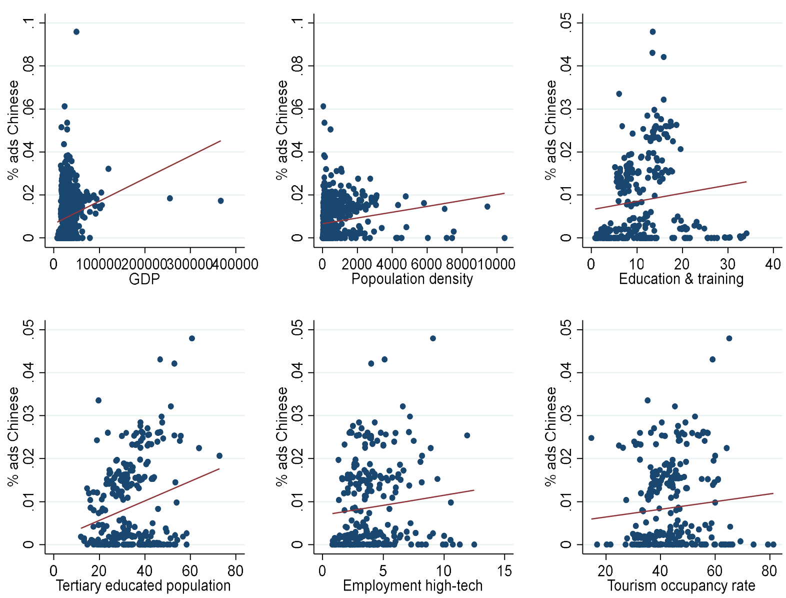
[3] To avoid possible effects of data fluctuations and to minimize missing information we consider the average on the three last available years that are for all the indicators 2019, 2020 and 2021. The only exceptions are population density and net occupancy rate for which the available years are respectively 2017, 2018, 2019 and 2016, 2017, 2018. The information regarding GDP for UK come from the UK Office for National Statistics.

Figure 10 - Scatter plot for the relationship between the percentage of online ads requiring English and GDP; population density; tertiary educated population; education & training; employment in the high-tech sector and tourism occupancy rate



Note: UK, IE and MT are excluded. The areas "Wolfsburg, Kreisfreie Stadt" and "Paris" are excluded respectively in the GDP and in the population density graphs since they represent extreme outliers.

Figure 11 - Scatter plot for the relationship between the percentage of online ads requiring Chinese and GDP; population density; tertiary educated population; education & training; employment in the high-tech sector and tourism occupancy rate.



Note: The areas "Wolfsburg, Kreisfreie Stadt" and "Paris" are excluded respectively in the GDP and in the population density graphs since they represent extreme outliers.

The next step of the analysis consists in the implementation of a spatial autoregressive model (see Drukker et al., 2013, for additional details) to properly taking into account spillover effects and spatial heterogeneity. This relaxes the assumption of independence of the units of analysis (i.e., the regions) which implicitly underlie the analysis of the previous part of this section. This assumption is unrealistic since two areas that are close geographically tend to be more similar to each other with respect to a specific outcome than areas that are spatially distant. The estimated models have the following form:

$$\boldsymbol{y} = \rho\mathbf{W}\boldsymbol{y} + \mathbf{X}\beta + \mathbf{u}$$

$$\mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \epsilon$$

Where $\mathbf{y}$ is the vector of observations of the dependent variable; $\mathbf{X}$ is the matrix of observations of the independent variables (those listed at the beginning of the section); $\mathbf{W}$ is a spatial-weighting matrix that parametrize the distance between the NUTS-2 units. This specification accounts for the spatial dependence of the outcome and of the disturbance process. For this explorative analysis, the spatial-weighting matrix has been calculated according to the contiguity criterion, according to which only neighbouring regions are allowed to affect each other. The analysis is reported for English and Chinese in this section, while the results for German, French and Spanish are reported in Appendix E.

Table 3 - Generalized spatial two-stage least-squares (GS2SLS) estimation for the percentage of online ads requiring **English** according to GDP; population density; employment in the high-tech sector; tourism occupancy rate; tertiary educated population; education and training.

	Model 1		Model 2		Model 3		Model 4	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Level of economic prosperity								
GDP per capita	0.003***	0.001					0.004***	0.001
Population density	0.030***	0.010					0.023**	0.011
Local labour market								
Employment in the high-tech sector			0.013***	0.003			0.006	0.005
Tourism occupancy rate			0.002*	0.001			0.001	0.001
Human capital								
Population with a tertiary degree					0.003***	0.001	-0.002	0.001
Participation in education and training					-0.001	0.002	-0.001	0.002
Constant	0.203***	0.029	0.182***	0.048	0.243***	0.038	0.196***	0.053
Wy	0.084	0.102	0.031	0.113	-0.017	0.112	0.113	0.100
Wu	0.822***	0.094	0.858***	0.097	0.861***	0.095	0.821***	0.090
N	214		214		214		214	
R^2	0.149		0.063		0.021		0.148	

Note: UK, IE and MT are excluded. GDP per capita and population density have been rescaled (divided by 1,000).
* p < 0.10, ** p < 0.05, *** p < 0.01.

Table 3 reports the models for the percentage of online ads requiring English.[4] Within the type of models presented in this section, the interpretation of the coefficients is not straightforward as they represent a combination of direct and indirect effects. The former are the effects of the spatial unit on itself, while the latter are the spillover effects across regions. In other words, the β and ρ parameters are not interpretable as the direct and spillover effects since the latter are computed recursively: a variation of X produces a change in y by β and that change in y spills over to produce a further reduction in y of ρW and that reduction spills over to produce yet another reduction in y, and so on. In our case, the spillover effects result to be marginal[5] so that the coefficient reported in Table 3 can be interpreted as direct effects. We estimate four different models. The first one considers only the variables related to the economic prosperity, the second one the variables related to the local labour market, the third one the variables connected to the human capital dotation, while in the final model all these variables are jointly considered. We can notice from the first three models how the variables considered tend to be positively correlated with the outcome and the only exception is the participation in education and training for which no effect is detected. The fourth model show that, once all the factors are jointly considered, only the variables connected to economic prosperity maintain a significant coefficient. This analysis confirms the results of the scatter plots in which the relationship between the percentage of online ads requiring English and GDP and population density appear to be stronger than the other factors considered.

Table 4 reports the model results for the percentage of online ads requiring Chinese, following the same logics used for English. The picture emerging from the spatial models differs substantially from what we found with the scatter plots that display a positive relationship between the online ads requiring Chinese and all the covariates considered and in particular for GDP and the proportion of population with a tertiary degree. The findings emerging from the spatial model show instead that the GDP is (weakly) negatively correlated in model 1, while no effect is detected in the last model. The other interesting and divergent results are about the percentage of workforce employed in the high-tech sector. In fact, it appears to be negatively correlated with the ads requiring Chinese. All in all, it seems to emerge that the knowledge of Chinese is required in densely populated areas[6] characterised by a medium-low level of economic development in terms of workforce qualification. This result is consistent with the discussion done in section 9 about the fact that Chinese is required for medium and low skill occupations.

[4] It has to be stressed that we are not in the position to identify causal effects. In the main text we will use the terms "correlation", "relationship" and "effect" as synonymous, while not implying any statement on causality.
[5] We computed the direct and indirect effects and the latter are always non-statistically significant (see Appendix E).
[6] The direct effect for the population density is equal to 0.0010 and the spillover effect to neighbouring regions is almost as large (0.0012) percentage point expected increase in online ads requiring Chinese (see Appendix E).

Table 4 - Generalized spatial two-stage least-squares (GS2SLS) estimation for the percentage of online ads requiring **Chinese** according to GDP; population density; employment in the high-tech sector; tourism occupancy rate; tertiary educated population; education and training.

	Model 1		Model 2		Model 3		Model 4	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Level of economic prosperity								
GDP per capita	-0.0001*	0.0000					0.0000	0.0000
Population density	0.0008**	0.0004					0.0009**	0.0004
Local labour market								
Employment in the high-tech sector			-0.0003**	0.0002			-0.0005**	0.0002
Tourism occupancy rate			0.0000	0.0000			0.0000	0.0000
Human capital								
Population with a tertiary degree					0.000	0.000	0.0001	0.0001
Participation in education and training					0.000	0.000	0.0000	0.0001
Constant	0.0067***	0.0011	0.005***	0.002	0.007***	0.002	0.005**	0.002
Wy	0.6636***	0.1031	0.694***	0.097	0.576***	0.127	0.713***	0.093
Wu	0.7757***	0.1365	0.781***	0.152	0.856***	0.140	0.738***	0.148
N	254		254		254		254	
R^2	0.056		0.005		0.003		0.071	

Note: GDP per capita and population density have been rescaled (divided by 1,000).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

11. Conclusions

In this paper we show a set of possible analysis related to skill demand that can be carried out with the Web Intelligence Hub's OJA database. The aim is to illustrate the regional distribution of foreign language skill demand in the European labour market, and the association of this demand with occupational groups, other skills, and macroeconomic factors.

The first step of the analysis identified the most demanded languages among online job ads in Europe. It is not surprising to find that English is by far the most requested language in the European labour market, followed by German, Spanish, Chinese and French.

The second step has been the analysis of the geographical distribution in the demand of the foreign languages focusing the attention on selected languages. For what concerns English, even though there are visible differences across European regions, no obvious regional pattern emerges at a first glance. The demand for German is much less strong outside Austria and Germany. In this case (as well as French and Spanish, shown in Appendix B), a regional pattern seems to emerge for the residual demand which is focused in the regions on the border with these two countries (particularly Belgium and Luxembourg, where German is an official language) and in some touristic regions. The demand for Chinese is higher than 1% only in France, Germany and the UK, while it is lower in other European countries.

As a third step, we looked at the association between demand for foreign languages and other skills (at the European level), and its distribution across occupational groups (at country level). This analysis showed that English is associated with the other widely requested transversal skills in the labour market (e.g. adapt to change), and that it is required above all among the high skill occupations. In contrast, Chinese is particularly demanded in medium and low skill occupations and it has no clear predictor among other skills. The prevalence of low-medium skill occupations suggests that demand for Chinese is not linked to international trade jobs, but mostly to the Chinese immigration network in Europe. Finally, German is relatively likely to be found in the same ads requesting Spanish and French, suggesting that these languages are often requested for jobs requiring communication with people from various language backgrounds.

The final step of the analysis looked at the relationship between the demand for foreign languages and a set of macro characteristics at NUTS-3 and NUTS-2 level: i) GDP; ii) population density; iii) participation rate in education and training in the last 4 weeks; iv) percentage of population with a tertiary degree; v) percentage of people employed in the high-tech sector; and vi) net occupancy rate of bed-places and bedrooms in hotels and similar accommodation. Our preliminary evidence points out that the demand for both English and Chinese tend to be higher in more densely populated area, but while English demand is positively associated to regional economic development, Chinese demand is negatively associated with the percentage of people employed in the high-tech sector.

References

Antonietti, R. and Loi, M. (2014). Firm Internationalization and Demand for Foreign Language Skills: A Firm-level Investigation. Economia Politica, 2, 173-194, http://dx.doi.org/10.1428/77436

Araújo, L., Flisi, S., Dinis da Costa, P. and soto-Clavo, E.. (2015). Languages and Employability, Publications Office of the European Union, Luxembourg, http://dx.doi.org/10.2788/860807

Ascheri, A., Kiss-Nagy, A., Marconi, G., Meszaros, M., Paulino, R. and Reis, F. (2021). Competition in urban hiring markets: evidence from online job advertisements. Eurostat Statistical Working Papers, http://dx.doi.org/10.2785/667004

Ascheri, A., Marconi, G., Meszaros, M. and Reis, F. (2022). Online Job Advertisements for Labour Market Statistics using R. Romanian Statistical Review, 2022(1), 3-26. https://www.revistadestatistica.ro/2022/03/online-job-advertisements-for-labour-market-statistics-using-r/

Bak, T., Long, M., Vega-Mendoza, M. and Sorace, A.. (2016). "Novelty, challenge, and practice: the impact of intensive language learning on attentional functions", PLoS ONE, 11, 4. http://dx.doi.org/e0153485

Cedefop (2019). Online job vacancies and skills analysis: a Cedefop pan-European approach. Luxembourg: Publications Office. Retrieved from https://www.cedefop.europa.eu/files/4172_en.pdf

Cedefop (2022). Skills in Occupation. Accessed on 01 July 2022 at: https://www.cedefop.europa.eu/en/tools/skills-online-vacancies/skills/occupations

Council of the european Union (2014). Conclusions on multilingualism and the development of language competences. Council of the European Union, https://github.com/gabrimarconi/langageskills/blob/Literature/2014_Conclusions%20on%20multilingualism%20and%20the%20development%20of%20language%20competences.pdf

Della Chiesa, B., Scott, J. and Hinton, C. (2012). Languages in a GlobalWorld: Learning for Better Cultural Understanding, OECD Publishing, https://www.oecdilibrary.org/docserver/9789264123557-

en.pdf?expires=1559551110&id=id&accname=ocid84004878&checksum=77AEA46C12225DA61CDF25B5B7389A43.

Drukker, D. M., Prucha, I. R., & Raciborski, R. (2013). Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. The Stata Journal, 13(2), 221-241, https://doi.org/10.1177/1536867X1301300.

Eurostat (2022). Regions and Cities database. Accessed on 01 June 2022 at: https://ec.europa.eu/eurostat/web/regions-and-cities

Fabo, B., Beblavý, M. and Lenaerts, K. (2017). The importance of foreign language skills in the labour markets of Central and Eastern Europe: assessment based on data from online job portals. Empirica, 44, 487–508. https://doi.org/10.1007/s10663-017-9374-6

Fidrmuc, J. and Fidrmuc, J. (2016). Foreign languages and trade: evidence from a natural experiment. Empirical Economics, 50(1), 31-49. https://doi.org/10.1007/s00181-015-0999-7

Goggin, D., Sheridan, I., Lárusdóttir, F., and Guðmundsdóttir, G. (2019). Towards the identification and assessment of transversal skills. INTED2019 Proceedings, 2513-2519. https://doi.org/10.21125/inted.2019.0686

Gudykunst, W. (2003), Cross-cultural and intercultural communication. Sage Publications, https://uk.sagepub.com/en-gb/eur/cross-cultural-and-intercultural-communication/book226099

ILO (2012). International Standard Classification of Occupations. Geneve: ILO, https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_172572.pdf

Isphording, I. (2015). “Language and Labor Market Success”, in International Encyclopedia of the Social & Behavioral Sciences (Wright, D. ed.), 260-265. Elsevier, http://dx.doi.org/10.1016/b978-0-08-097086-8.94033-9

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning (Vol. 112, p. 18). Springer. http://dx.doi.org/10.1007/978-1-0716-1418-1

Jones, D. (2021). noctua: Connect to 'AWS Athena' using R 'AWS SDK' 'paws' ('DBI' Interface). Retrieved from https://cran.rstudio.com/web/packages/noctua/index.html

Lahti, L., Huovari, J., Kainu, M., and Biecek, P. (2017). Retrieval and Analysis of Eurostat Open Data with the eurostat Package. R Journal, 9(1), 385-392. https://journal.r-project.org/archive/2017/RJ-2017-019/RJ-2017-019.pdf

Marconi, G. and Vergolini, L. (2022). Langageskills. Github repository available at: https://github.com/gabrimarconi/langageskills

Melitz, J. (2008). Language and foreign trade. European Economic Review, 52(4), 667-699. https://doi.org/10.1016/j.euroecorev.2007.05.002

OECD (2021). PISA 2025 Foreign Language Assessment Framework. OECD, https://www.oecd.org/pisa/foreign-language/PISA-2025-FLA-Framework.pdf

Osmani, M., Weerakkody, V., Hindi, N., and Eldabi, T. (2019). Graduates employability skills: A review of literature against market demand. Journal of Education for Business, 94(7), 423-432. http://dx.doi.org/10.1080/08832323.2018.1545629

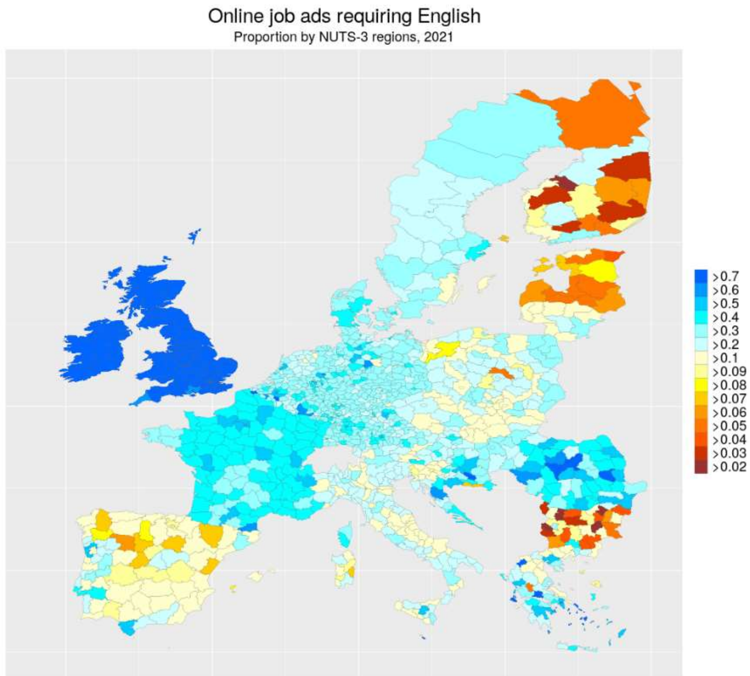
Plewa, P. and Stermšek, M. (2017). Labour Migration from China to Europe: Scope and Potential. International Labour Organization: Geneva, Switzerland, 1-136. http://www.ilo.int/wcmsp5/groups/public/---asia/---ro-bangkok/---ilo-beijing/documents/publication/wcms_573470.pdf

R Core Team. (2021). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

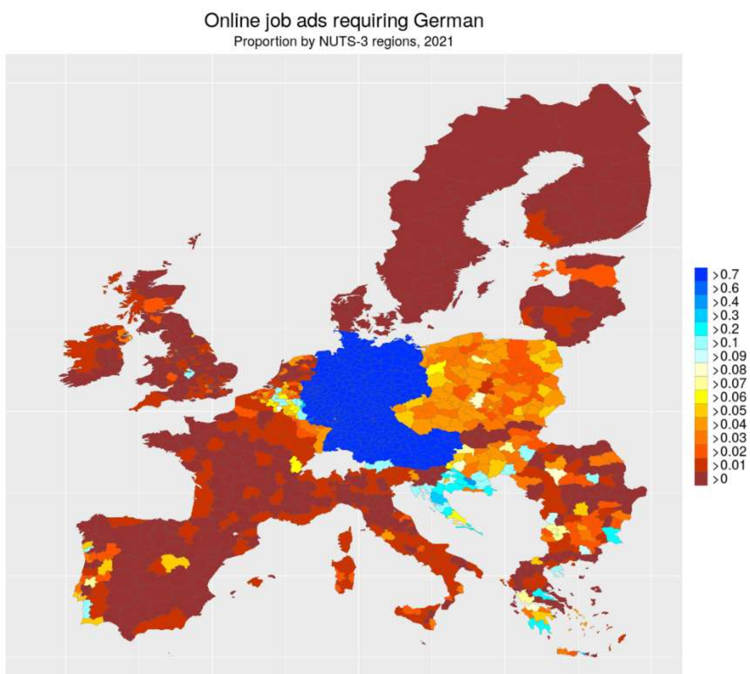
R-SIG-DB R Special Interest Group on Databases, Wickham, H., and Müller, K. (2021). DBI: R Database Interface. https://cran.r-project.org/web/packages/DBI/index.html

Woll, B. and Wei, L. (2019), Cognitive Benefits of Language Learning: Broadening our perspectives. British Academy. https://www.thebritishacademy.ac.uk/sites/default/files/Cognitive-Benefits-LanguageLearning-Final-Report.pdf

Appendix A



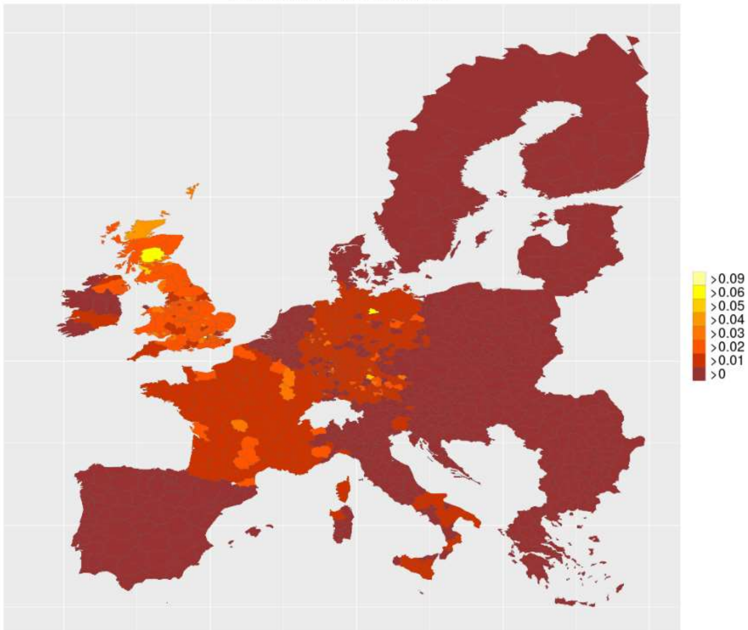
Notes: Average proportion across all ads
Not in the map: Região Autónoma dos Açores 0.3, Região Autónoma da Madeira 0.2, Κύπρος (Kypros) 0.55
El Hierro 0.08, Fuerteventura 0.15, Gran Canaria 0.12, La Gomera 0.03, La Palma 0.07, Lanzarote 0.08, Tenerife 0.16
Guadeloupe 0.36, Martinique 0.38, Guyane 0.37, La Réunion 0.42, Mayotte 0.35
(C) EuroGeographics for the administrative boundaries
Map produced in R with a help from Eurostat-package <github.com/ropengov/eurostat/>



Notes: Average proportion across all ads
Not in the map: Região Autónoma dos Açores 0.03, Região Autónoma da Madeira 0.07, Κύπρος (Kypros) 0.01
El Hierro 0, Fuerteventura 0.02, Gran Canaria 0.01, La Gomera 0.04, La Palma 0.01, Lanzarote 0.01, Tenerife 0.01
Guadeloupe 0.01, Martinique 0.01, Guyane 0.01, La Réunion 0.01, Mayotte 0.01
(C) EuroGeographics for the administrative boundaries
Map produced in R with a help from Eurostat-package <github.com/ropengov/eurostat/>

Online job ads requiring Chinese

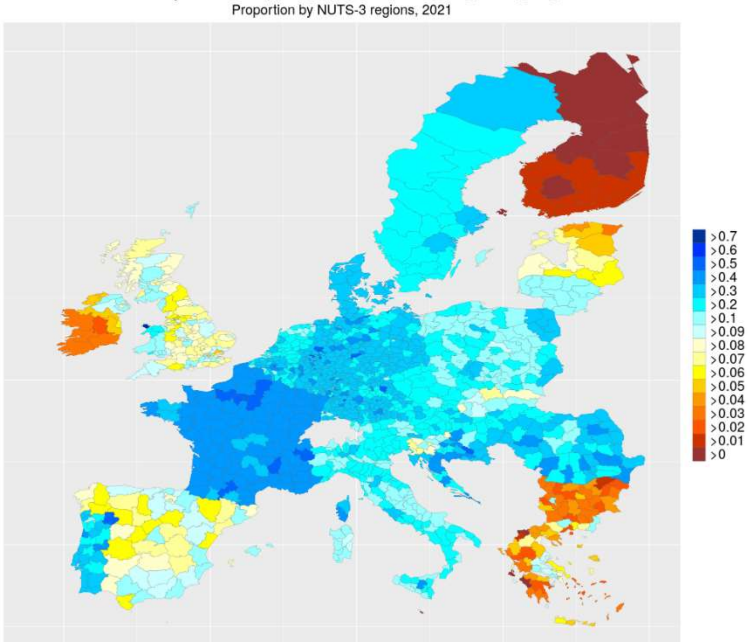
Proportion by NUTS-3 regions, 2021



Notes: Average proportion across all ads
Not in the map: Região Autónoma dos Açores 0, Região Autónoma da Madeira 0, Κύπρος (Kypros) 0, El Hierro 0 Fuerteventura 0, Gran Canaria 0, La Gomera 0, La Palma 0, Lanzarote 0, Tenerife 0, Guadeloupe 0.03 Martinique 0.02, Guyane 0.03, La Réunion 0.02, Mayotte 0.01
(C) EuroGeographics for the administrative boundaries
Map produced in R with a help from Eurostat-package ‹github.com/ropengov/eurostat/›

Online job ads requiring at least one foreign language

Proportion by NUTS-3 regions, 2021

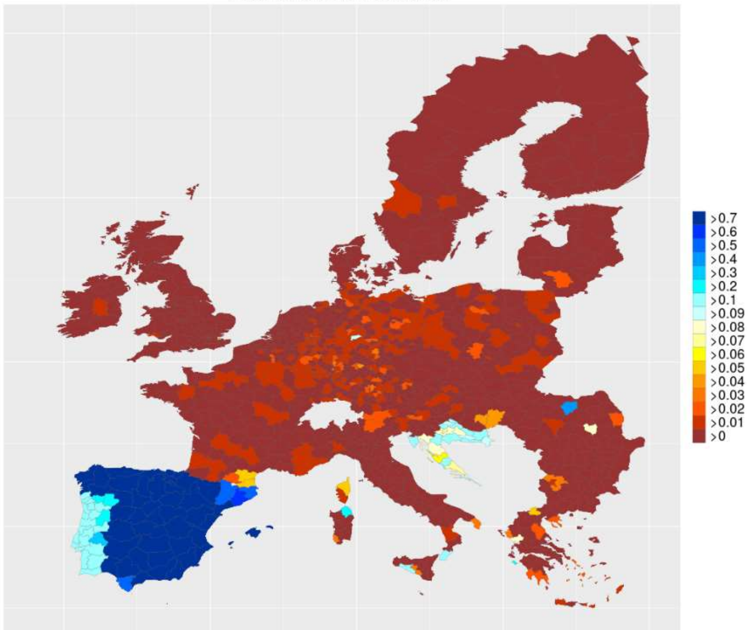


Notes: Average proportion across all ads
Not in the map: Região Autónoma dos Açores 0.39, Região Autónoma da Madeira 0.41, Κύπρος (Kypros) 0 El Hierro 0.06, Fuerteventura 0.13, Gran Canaria 0.1, La Gomera 0.07, La Palma 0.07, Lanzarote 0.07, Tenerife 0.12 Guadeloupe 0.39, Martinique 0.4, Guyane 0.4, La Réunion 0.43, Mayotte 0.37
(C) EuroGeographics for the administrative boundaries
Map produced in R with a help from Eurostat-package ‹github.com/ropengov/eurostat/›

Appendix B

Online job ads requiring Spanish

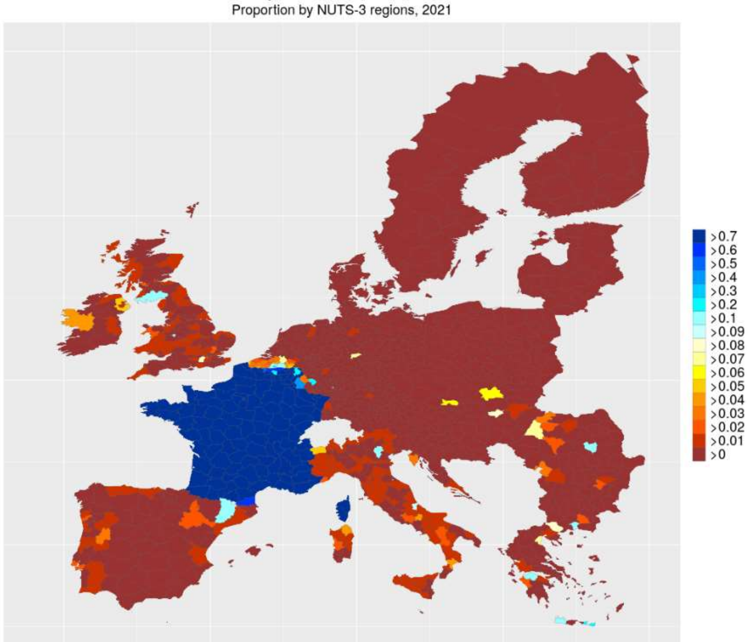
Proportion by NUTS-3 regions, 2021



Note: Average across the proportions estimated for each 3-digit ISCO occupational category
Not in the map: Região Autónoma dos Açores 0.16, Região Autónoma da Madeira 0.2, Κύπρος (Kypros) 0
El Hierro 0.96, Fuerteventura 0.94, Gran Canaria 0.9, La Gomera 0.97, La Palma 0.94, Lanzarote 0.95, Tenerife 0.87
Guadeloupe 0, Martinique 0, Guyane 0.01, La Réunion 0, Mayotte 0
(C) EuroGeographics for the administrative boundaries
Map produced in R with a help from Eurostat-package ‹github.com/ropengov/eurostat/›

Online job ads requiring French

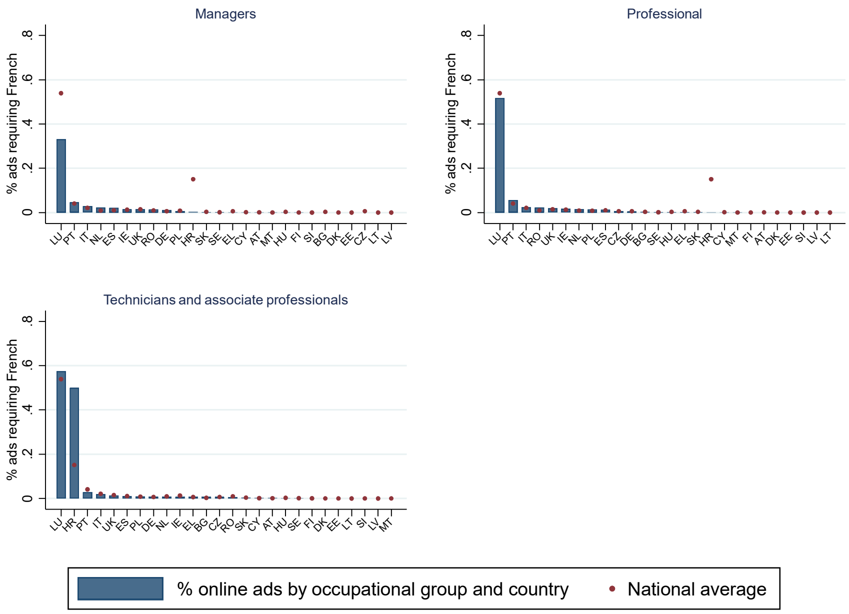
Proportion by NUTS-3 regions, 2021



Note: Average across the proportions estimated for each 3-digit ISCO occupational category
Not in the map: Região Autónoma dos Açores 0.01, Região Autónoma da Madeira 0, Κύπρος (Kypros) 0
El Hierro 0, Fuerteventura 0, Gran Canaria 0, La Gomera 0.01, La Palma 0, Lanzarote 0, Tenerife 0, Guadeloupe 0.99
Martinique 0.97, Guyane 0.96, La Réunion 0.93, Mayotte 0.94
(C) EuroGeographics for the administrative boundaries
Map produced in R with a help from Eurostat-package ‹github.com/ropengov/eurostat/›

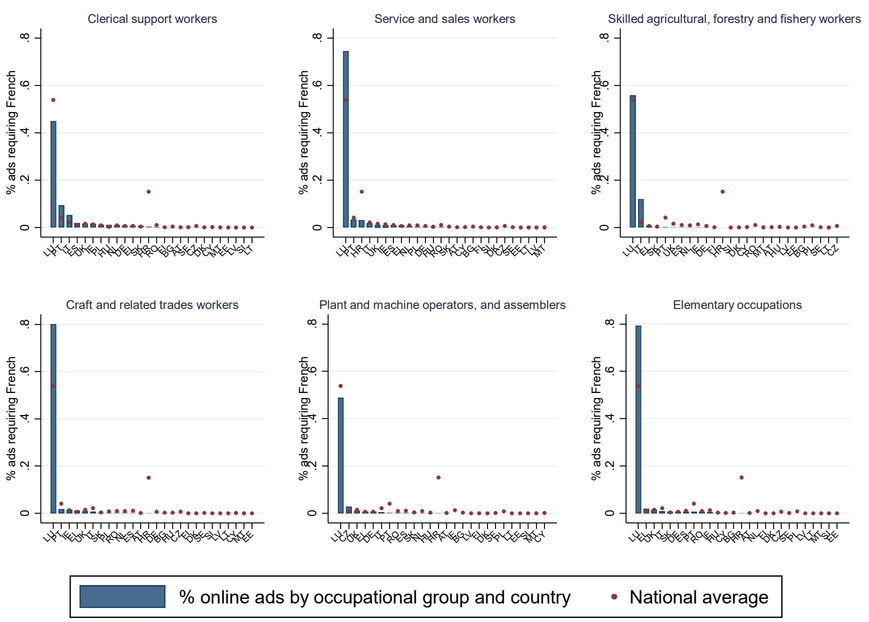
Appendix C

Percentage of online ads requiring French in high skill occupational groups, by country.



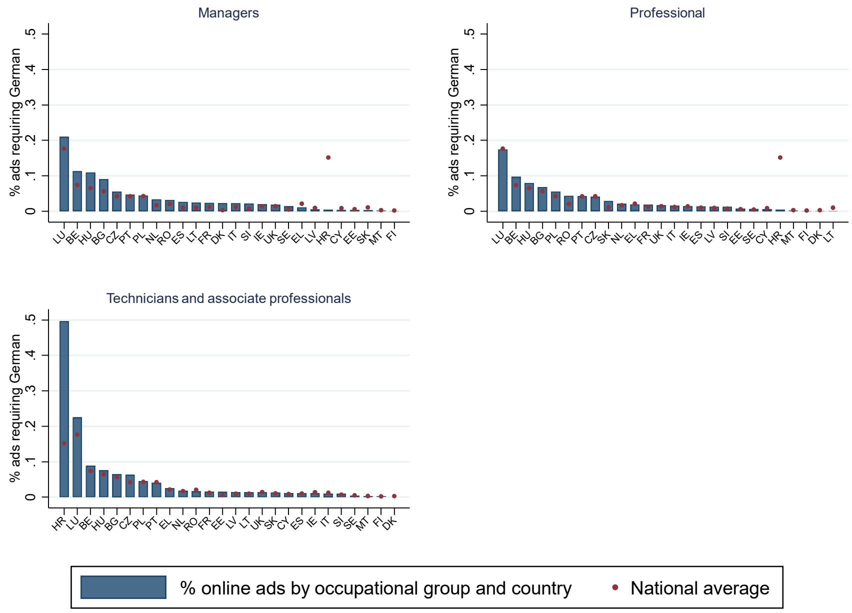
Note: FR and BE are excluded.

Percentage of online ads requiring French in medium and low skill occupational groups, by country.



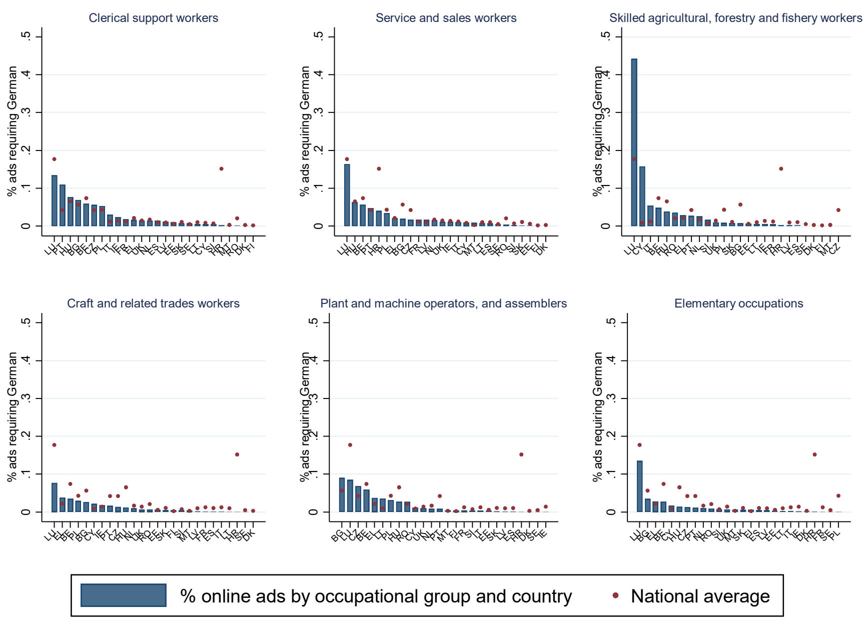
Note: FR and BE are excluded.

Percentage of online ads requiring German in high skill occupational groups, by country.



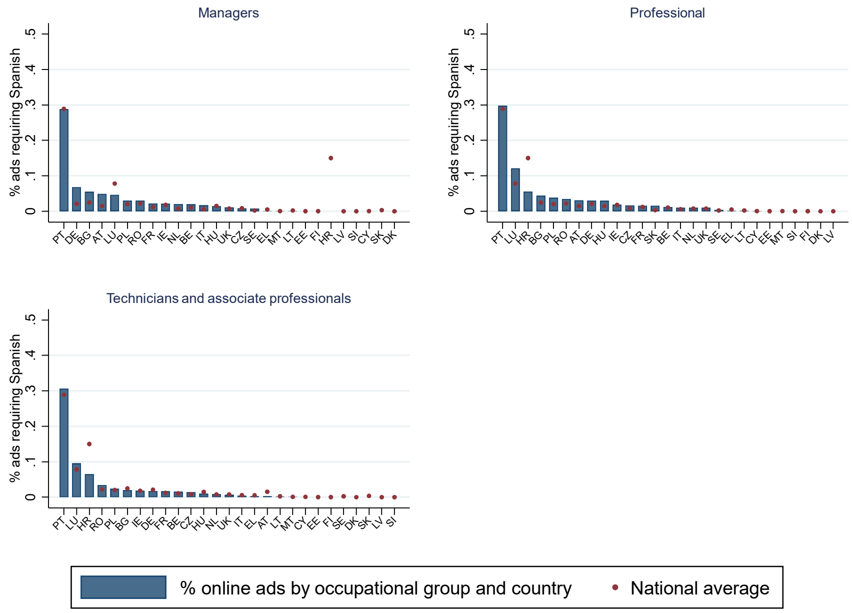
Note: DE and AT are excluded.

Percentage of online ads requiring German in medium and low skill occupational groups, by country.



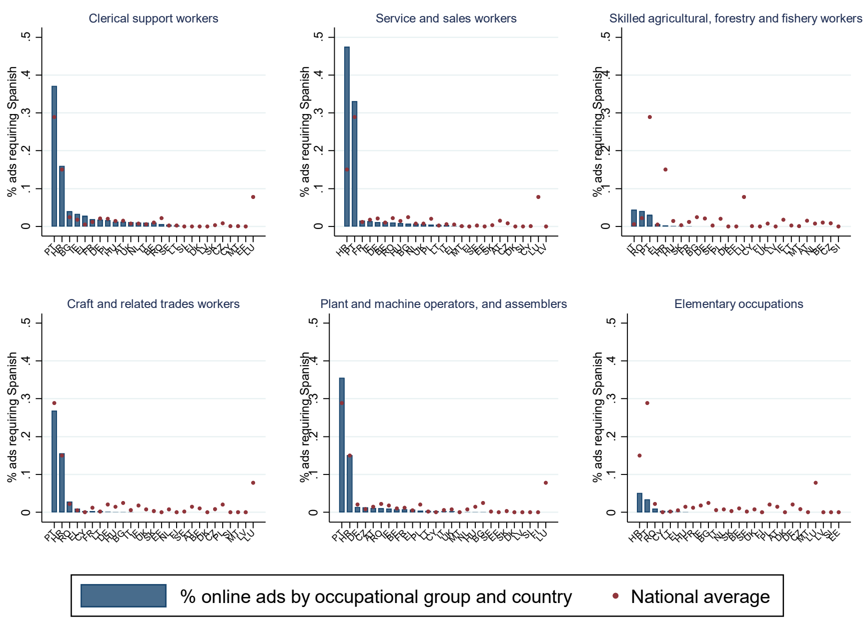
Note: DE and AT are excluded.

Percentage of online ads requiring Spanish in high skill occupational groups, by country.



Note: ES is excluded.

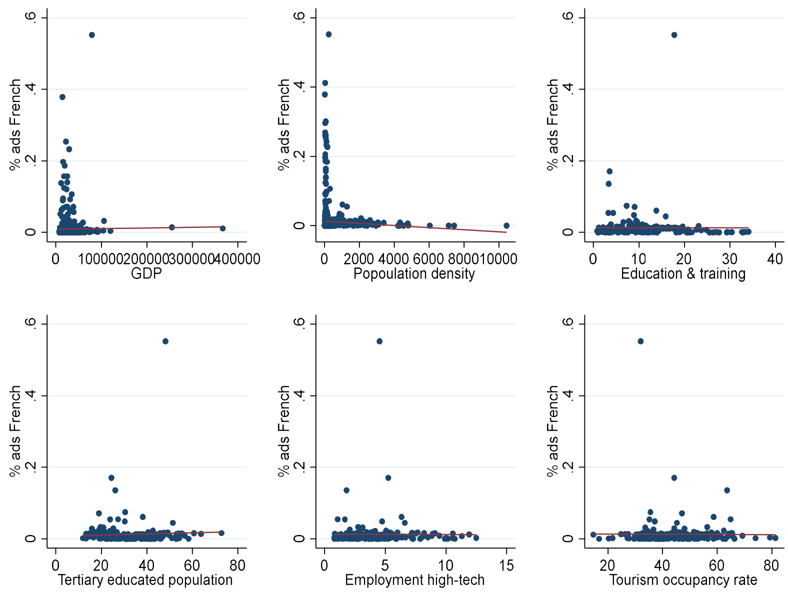
Percentage of online ads requiring Spanish in medium and low skill occupational groups, by country.



Note: ES is excluded.

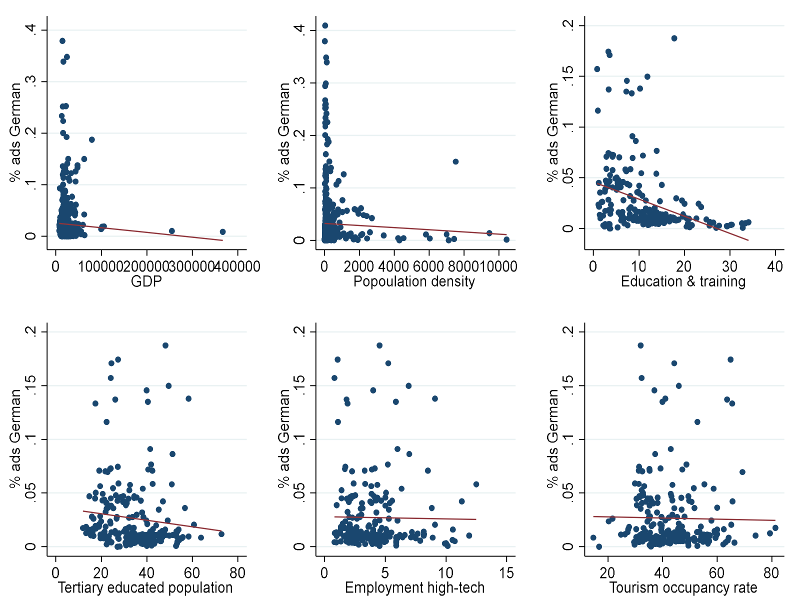
Appendix D

Scatter plot for the relationship between the percentage of online ads requiring French and GDP; population density; tertiary educated population; education & training; employment in the high-tech sector and employment in touristic sector.



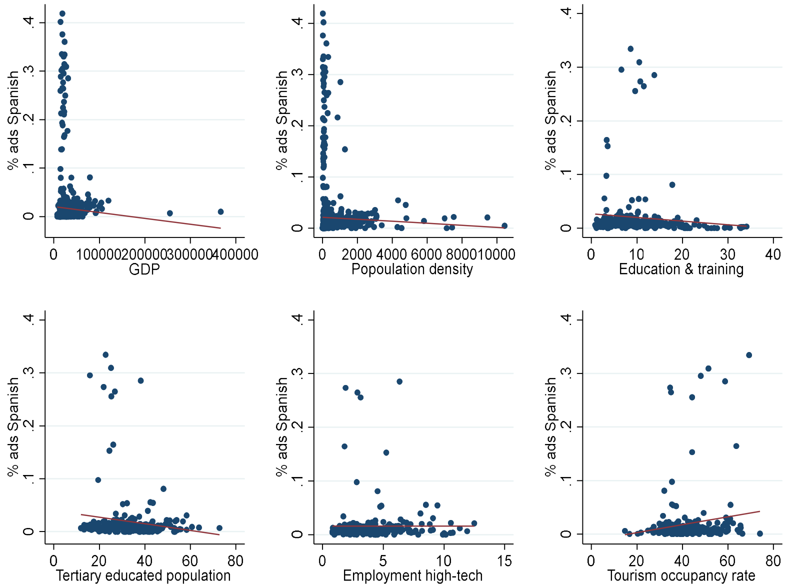
Note: FR and BE are excluded.

Scatter plot for the relationship between the percentage of online ads requiring German and GDP; population density; tertiary educated population; education & training; employment in the high-tech sector and employment in touristic sector.



Note: DE and AT are excluded.

Scatter plot for the relationship between the percentage of online ads requiring Spanish and GDP; population density; tertiary educated population; education & training; employment in the high-tech sector and employment in touristic sector.



Note: ES is excluded.

Appendix E

Generalized spatial two-stage least-squares (GS2SLS) estimation for the percentage of online ads requiring French according to GDP; population density; employment in the high-tech sector; tourism occupancy rate; tertiary educated population; education and training.

	Model 1		Model 2		Model 3		Model 4	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Level of economic prosperity								
GDP per capita	-0.003	0.003					-0.001	0.003
Population density	0.048	0.030					0.033	0.029
Local labour market								
Employment in the high-tech sector			-0.023*	0.014			-0.030*	0.017
Tourism occupancy rate			0.005	0.003			0.006*	0.003
Human capital								
Population with a tertiary degree					0.002	0.004	0.003	0.004
Participation in education and training					-0.003	0.006	-0.001	0.005
Constant	0.389***	0.094	0.264*	0.149	0.253**	0.108	0.028	0.148
$\boldsymbol{\rho}$	0.390**	0.163	0.265	0.182	0.531***	0.138	0.762***	0.762
$\boldsymbol{\lambda}$	0.497***	0.186	0.659***	0.197	0.319*	0.182	-0.104	-0.104
N	221		221		221		221	
R^2	0.006		0.000		0.000		0.000	

Note: FR and BE are excluded. * p < 0.10, ** p < 0.05, *** p < 0.01.

Generalized spatial two-stage least-squares (GS2SLS) estimation for the percentage of online ads requiring German according to GDP; population density; employment in the high-tech sector; tourism occupancy rate; tertiary educated population; education and training.

	Model 1		Model 2		Model 3		Model 4	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Level of economic prosperity								
GDP per capita	-0.003	0.003					0.002	0.003
Population density	0.011	0.028					-0.006	0.025
Local labour market								
Employment in the high-tech sector			-0.008	0.012			0.009	0.016
Tourism occupancy rate			0.001	0.003			0.003	0.003
Human capital								
Population with a tertiary degree					0.000	0.003	-0.003	0.004
Participation in education and training					-0.023***	0.006	-0.017***	0.005
Constant	0.334***	0.095	0.147	0.153	0.569***	0.124	0.281*	0.159
$\boldsymbol{\rho}$	0.510***	0.149	0.787***	0.108	0.410***	0.133	0.702***	0.101
$\boldsymbol{\lambda}$	0.020	0.203	-0.429**	0.183	0.031	0.179	-0.391**	0.168
N	208		208		208		208	
R^2	0.027		0.009		0.248		0.216	

Note: DE and AT are excluded. * p < 0.10, ** p < 0.05, *** p < 0.01.

Generalized spatial two-stage least-squares (GS2SLS) estimation for the percentage of online ads requiring Spanish according to GDP; population density; employment in the high-tech sector; tourism occupancy rate; tertiary educated population; education and training.

	Model 1		Model 2		Model 3		Model 4	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Level of economic prosperity								
GDP per capita	0.003	0.002					0.003	0.003
Population density	0.004	0.027					-0.014	0.027
Local labour market								
Employment in the high-tech sector			0.011	0.013			0.003	0.018
Tourism occupancy rate			0.005	0.003			0.006^{*}	0.003
Human capital								
Population with a tertiary degree					0.006^{*}	0.004	0.002	0.004
Participation in education and training					-0.015^{**}	0.006	-0.012^{**}	0.005
Constant	0.175^{**}	0.079	-0.015	0.153	0.284^{**}	0.131	-0.061	0.159
$\boldsymbol{\rho}$	0.554^{***}	0.121	0.617^{***}	0.122	0.379^{**}	0.154	0.675^{***}	0.106
$\boldsymbol{\lambda}$	0.195	0.185	0.152	0.183	0.368^{**}	0.186	-0.019	0.174
N	237		237		237		237	
R^2	0.038		0.024		0.083		0.109	

Note: ES is excluded. * p < 0.10, ** p < 0.05, *** p < 0.01.

Direct, indirect and total effects derived from the generalized spatial two-stage least-squares (GS2SLS) model for the percentage of online ads requiring English.

	Models 1-3						Model 4					
	Direct		Indirect		Total		Direct		Indirect		Total	
GDP per capita	0.003***	0.001	0.000	0.000	0.003***	0.001	0.004***	0.001	0.000	0.000	0.004***	0.001
Population density	0.030***	0.010	0.002	0.003	0.032***	0.012	0.024**	0.010	0.002	0.003	0.026**	0.012
Employment in the high-tech sector	0.013***	0.003	0.000	0.001	0.013***	0.004	0.007	0.005	0.001	0.001	0.007	0.005
Tourism occupancy rate	0.002*	0.001	0.000	0.000	0.002*	0.001	0.001*	0.001	0.000	0.000	0.002	0.001
Population with a tertiary degree	0.003***	0.001	0.000	0.000	0.003***	0.001	-0.002*	0.001	0.000	0.000	-0.003*	0.002
Participation in education and training	-0.001	0.002	0.000	0.000	-0.001	0.002	-0.001	0.002	0.000	0.000	-0.001	0.002

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Direct, indirect and total effects derived from the generalized spatial two-stage least-squares (GS2SLS) model for the percentage of online ads requiring Chinese.

	Models 1-3						Model 4					
	Direct		Indirect		Total		Direct		Indirect		Total	
GDP per capita	-0.0001*	0.0000	-0.0001	0.0000	-0.0001**	0.0001	0.0000	0.0000	-0.0001	0.0001	-0.0001	0.0001
Population density	0.0009**	0.0004	0.0009*	0.0006	0.0018*	0.0009	0.0010**	0.0004	0.0012*	0.0007	0.0022**	0.0010
Employment in the high-tech sector	-0.0004**	0.0002	-0.0005	0.0003	-0.0008*	0.0005	-0.0005**	0.0002	-0.0007*	0.0004	-0.0013**	0.0006
Tourism occupancy rate	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0001
Population with a tertiary degree	0.0000	0.0001	0.0000	0.0000	-0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002
Participation in education and training	0.0000	0.0001	0.0000	0.0001	0.0000	0.0002	0.0000	0.0001	0.0000	0.0001	0.0000	0.0002

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.