The Influence of Distance on Hiring Difficulties: Are Vacancies Located Away from Headquarters More Difficult to Fill?

Debora Abduramanaj, Eren Erkoseoglu, Luis Enrique Kaiser

February 2, 2023

Abstract

Does the distance between the headquarters and the location of a given, advertised job matter for its expected vacancy duration? Triggered by insufficient evidence in the existing literature, we aim to shed some light on the impact that the distance to headquarters for a certain job position may have on its respective duration. We are using Fixed Effects quasi-experimental design, on a granular and extensive dataset with thousands of job openings for German firms, to isolate the effect driven from the labour supply side. In our most stringent specification, a 100 km increase in the distance between the vacancy position and headquarters' locations is associated with a 2.4 % increase in the amount of days it takes to fill it. Noteworthy is also, that the effect is mostly driven by permanent contracts and German-speaking workers. While not claiming a causal relationship, we argue that we found a strong association between the distance of an advertised job position from firm's central offices and its associated hiring difficulties.

Note: This template was initially created by Gaudecker, 2019

List of Figures

1 Locations of Job openings vs. headquarters in Germany 5
2 Duration variation across fixed effects variables 10
3 Distribution of categorical covariates against Log Duration 17
4 The variability of distance effect across selected subgroups 18
5 Comparison of distance and duration distribution across the datasets 20

List of Tables

1 Continuous variables - Overview of summary statistics 6
2 Categorical variables - Overview of summary statistics 6
3 Posting Count's correlation with Duration and Distance 8
4 Main Regression Model Results 11
5 Main Model Extension - Heterogeneity effects 13
6 Table of variables and coefficients used in the study 16
7 Summary statistics of continuous variables, before the data-cleaning process . . . 19
8 Summary statistics of categorical variables, before the data-cleaning process . . . 19
9 Robustness Regression Model Results 21

1 Introduction

We test our hypothesis of how the duration of a vacancy filling is altered depending on whether a vacancy position is located in close proximity to the headquarter or not by measuring the distance between firms' headquarters and their branches. There are findings regarding the impact of distance between vacancy candidates and firms' location on the duration of vacancy filling. However, when the distance is converted to the proximity between the location of the vacancy inside the firm, which is the distance between the branch and the headquarter, a scarcity of studies maintains. In this way, it is important for our results to align with the literature regarding vacancy-filling duration determinants for validity purposes.

1.1 Labor Supply Side Determinants

To exemplify the existing literature, according to Marinescu and Rathelot, job seekers are 35 percent less likely to apply for a job 10 miles away from their residence location. This likelihood of a lower probability of applying accounts for a 5.3 percent loss of hires in the U.S. labor market (2018). This outcome is explained by the time and financial costs of commuting to the workplace. Since the time and financial costs of commuting increase as the distance of required movement increases, utility gain from working decreases. In this way, increasing costs decrease the likelihood of applying for the job which increases the vacancy-filling duration. Furthermore, these findings can be extended further into the movement of employees between cities for the vacancy. In this case, the time cost results from efficiency losses due to the time gap that the employee would be unable to work and the accumulation of job tasks during the movement process. Regarding the financial costs, these costs arise from payments such as furniture, shipment, and accommodation. Moreover, as the distance between the employee's location and the vacancy location increases, the time and financial costs of movement increase as well (Manning and Petrongolo, 2017). Thus, the likelihood of applying for vacancy decreases due to lower present utility by increasing costs of movement. Consequently, as the distance increases, the vacancy-filling duration increases as well.

1.2 Labor Demand Side Determinants

Regarding the firm side determinants of the vacancy-filling duration, firms' employee requirements, firm sizes, and technology account for substantial roles. According to Andrews et al., there is a significantly negative correlation between filling a position in a big firm and the duration to fill that position in the UK. This outcome results from big firms' requirements for a higher productivity level than the market average. Since higher productivity requirements decrease the number of applications, the vacancy-filling duration increases (2008). On the other hand, with the better usage of technological tools and better-designated resources for candidate search, the vacancy-filling duration decreases due to more ease of reaching a higher number of applications (Davis, Faberman, and Haltiwanger, 2013). Thus, more efficient usage of technological tools and higher resource allocation for candidate research, which increase with the firm size, decreases the duration to fill a vacancy. However, employee requirements present the reverse effect on the vacancy-filling duration.

1.3 The Vacancy-Filling Duration and In-firm Vacancy Locations

In the sense of determinants of duration to fill a vacancy, we find that vacancies closely located to headquarters are more likely to be preferred. To explain this further, our estimation results demonstrate that a 100 km increase in the distance between the two locations is associated with a 2.4 % increase in the vacancy-filling duration. These findings are originated by applicants' choices which affect the likelihood of applying for a vacancy. For the simplicity of driving forces, we differentiated the explanations behind these choices into two sources of interest. These are applicant choices in the sense of having an inclination towards working at the headquarter or at a close distance to the headquarter.

Regarding the driven forces of inclination towards working at the headquarter, this inclination arises from the differences between the branch and its headquarter. These differences are wage and working benefit differentials, and prestige respectively. Since a branch has the ability to access almost identical technological tools and resources of the headquarter for reaching potential candidates, the differences regarding the usage of technology and resource allocation for the searching process are negligible (Davis, Faberman, and Haltiwanger, 2013). On the other hand, in the case of wage and working benefit differentials, since a branch receives its budget from the headquarter depending on the branch's performance, it is reasonable to assume that, vacancies located in the headquarters are tended to result in higher wages and working benefits. Furthermore, the naturally bigger size of the headquarter supports this assumption as well (Oosterbeek and Praag, 1995). Thus, the likelihood of applying for a vacancy located in the headquarter is higher compared to its branch. In this way, the vacancy-filling duration for the headquarter is shorter than the branch. Furthermore, in the absence of wage differentials and working benefits differentials, vacancies that are located in the headquarter present higher prestige which yields higher ego-rent. In this way, the higher ego-rent results in developing an inclination towards working at the headquarter compared to its branch (Oosterbeek and Praag, 1995; Rogoff, 1990). Thus, in both of the scenarios, the likelihood of applying for a vacancy located in the headquarter is higher than the likelihood of applying for a job located in the branch which shortens the vacancy-filling duration for the headquarter relative to its branch.

In the case of the inclination towards working at a close distance to the headquarter, this results from future costs of the movement process. According to Miner and Estler, with increasing productivity and tenure, the probability of creating a new position for the employee as a promotion increases (1985, pp. 121-136). In this way, an employee of the branch can be transferred to the firm's headquarter after achieving a certain level of productivity and tenure. Thus, the probability of being transferred to the headquarter results in a higher likelihood of applying for the vacancy located in the branch due to the same driven forces as in the case of inclination towards working at the headquarter (Oosterbeek and Praag, 1995; Rogoff, 1990). However, this transfer process usually requires movement across cities which associates with time and financial costs. Moreover, since these costs increase with the distance between the branch and its headquarter, the future value of working at the headquarter decreases. In this way, due to decreasing expected future value of working at the headquarter by increasing time and financial costs of movement, the likelihood of applying for the vacancy located in the branch decreases as the distance between the headquarter and the branch increases. Consequently, the vacancy-filling duration for the branch located at a close distance to the headquarter becomes lower compared to the branch that is relatively further away from the headquarter.

2 Data

2.1 Dataset Restrictions

The dataset at hand is web scraped from a data provider. The original dataset contains above 7.5 million observations. Web scraped data is usually considered to be noisy and contains only a relatively small fraction of directly usable observations. In order to customize the dataset to fit the research questions and to make it usable for research purposes, we needed to make the following adjustments:

a) Columns with missing values in the considered covariates are removed. This process removes roughly one million observations, shrinking the dataset to 300 thousand observations. For the essential variables, we provide robustness checks to ensure that the results are unbiased in this process. K-nearest-neighbor and Multivariate Imputation by Chained Equations could also be considered in this case. In the discussion part of this paper, we discuss why these imputation methods were ultimately not used in this paper.

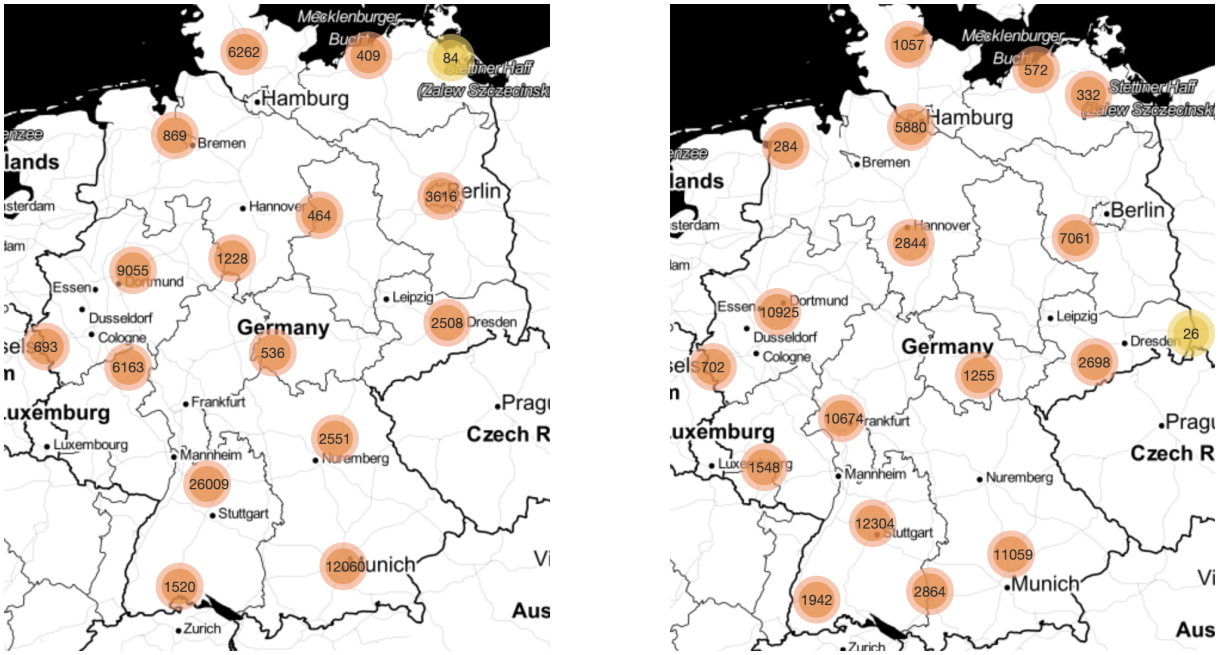
b) Firms with a small number of employees (<500) are prohibited from the dataset. On top of that, only regular working hours are considered.

c) The duration of job postings is top-coded at 365 days. Therefore, all observations above 365 days are counted as 365. There is a simple intuition behind this step: The recruiter in a given firm still needs to delete the job posting, as the job filling was unsuccessful. Since we include firm fixed effects, we only consider cases where firms search directly, not through an intermediary. The job posting count is also top-coded at 20 since a count rate above 20 is not likely and results from falsely scraped data.

d) In order to structure unbalanced variables and make the interpretation of heterogeneous effects more intuitive, we clustered the contract types into a "permanent" and a "non-permanent" class. For a similar case, we clustered the language requirements into "German" and "International" and the education requirements into "Academic" and "Non-Academic".

2.2 Distance measure

Unfortunately, it is not possible to calculate the distance measure between two cities directly from the dataset For the location for which the job is advertised are GPS coordinates present in the dataset. The location of the headquarters of the companies that advertised the jobs in question can only be identified by the name in the dataset. First, these city names have to be mapped onto their GPS coordinates. For that matter, we used a database of all German cities and their GPS coordinates. In order to match the city name in the dataset of a given firm and the city names in the database, we cleaned both columns extensively, based on best practices (Arhur Turrel, 2022), and used a Levenshtein-based comparing algorithm. (This article was helpful for the implementation Adam Cohen, 2011) These algorithms typically use a Levenshtein distance to compute a score for the similarity of strings. Strings which have a high enough similarity score are considered equal. After obtaining the GPS coordinates for both locations, we created the distance by using the distance equation between two points on a sphere. Of course, this measure represents the direct way, not including roads or train routes, and uses a perfect sphere as a proximity for the shape of the earth. Figure 1 shows the locations of the headquarters (a) and the job openings (b). One main takeaway from these maps is that at least some variation in locations is present in the data, which can be exploited.



(a) Locations of headquarters in Germany

(b) Location of job openings in Germany

Figure 1: Locations of Job openings vs. headquarters in Germany

2.3 Data Consistency across the Datasets

As we had to fit the dataset to our research question, we lost a number of observations throughout the process. To prove that our final restricted dataset is representative of the original dataset, we make a comparison across both datasets for our two main variables and provided the results for both vacancy duration and distance from headquarters' location in the appendix. Their distribution metrics resembled each other in both restricted and original versions of the dataset, attributing more reliability to our analysis and the derived findings.

2.4 Descriptive Statistics

Throughout this section, we describe and summarize the basic data features for all variables relevant to our research analysis. We will first provide some statistical measures for continuous variables and then continue with relative shares for categorical variables.

2.4.1 Continuous Variables

Across a sample of 74,027 observations, duration varies from 1 single day to a whole year, with the average vacancy lasting for 79 days. Nevertheless, most of the observations are dispersed from the mean, inducing a high variability across the data. Most of the job positions take 3 months or less to be filled, with 75% of total observations vacancy duration amounting to roughly 98 days. We use the logarithm of duration to overcome the large variation in data and approximate a more plausible normal distribution. By doing so, we can also reduce the influence of outliers and heteroscedasticity to a certain extent. As it can also be seen from Table 1, the data range is now shorter with a lower standard deviation, implying a reduction of the previous right-skewness in the data.

	Duration (in days)	Log(Duration)	Posting Count	Distance
Mean	78.9	3.8	3.3	162.0
Std. Deviation	86.6	1.2	4.6	183.9
Min	1.0	0.0	1.0	0.0
Max	365	5.9	20.0	834.1
25% Percentile	22.0	3.1	1.0	4.0
50%	49.0	3.9	1.0	81.5
75%	98.0	4.6	3.0	292.7

Table 1: Continuous variables - Overview of summary statistics

As per postings count variable, firms seem to post between once and twenty times, with the average firm posting the same job position no more than 3 times. As it is also displayed in Table 1, only 25% out of total observations tend to go beyond 3 postings per job position. However, there is still a considerable variation across the data, with frequent observations in the low posting values.
Regarding distance, the average job position in our sample is located around 160 km away from the headquarters' location. The minimum distance observed is 0 km, representing all vacancies within the headquarters. The maximum value is almost 835 km, corresponding to the maximum possible distance in Germany's north-south direction. However only an insignificant percentage of positions share this distance, as 75% of all vacancies fall below the distance of 293 km.

2.4.2 Categorical Variables

In Table 2, we statistically show that vacancy observations do not share the same characteristics among them, but rather differ in many aspects, where the most relevant ones for our research analysis are (a) salary inclusion in the job posting, (b) job language, (c) education level, and (d) type of the job contract.

Variable	Category	Relative Share
Salary Posted	False	90.7%
Salary Posted	True	9.3%
Applicant Language	German	96.5%
Applicant Language	International	3.5%
Education Level required	No University degree	53.1%
Education Level required	University Degree	46.9%
Contract Type	Permanent	75.9%
Contract Type	Non Permanent	24.1%

Table 2: Categorical variables - Overview of summary statistics

Only 9.3% out of 74,027 job postings include the salary level in their description. Nevertheless, these limited salary observations are often times inherently flawed and not usable, thus making the computation of salary effect on duration almost impossible. Regarding the language required, around 96% of job vacancies in our sample are intended for applicants with native or fluent German. The rest are more suitable for internationals with the most common required languages being English, French, and Spanish. For the education feature, the relative shares for both "university degree" and "no university degree" are almost equal, implying that both blue collar and white collar jobs are frequent in the sample. Lastly, regarding the contract type, nearly 76%

of all job vacancies offered a permanent contract, while the others offered fixed-term contracts for positions like internship and apprenticeship among others.

3 Empirical Methodology

3.1 Relationship between Distance and Vacancy Duration

It is reasonable to argue that people do not always behave rationally and what happens nowadays in the real world does not necessarily follow old models of economic behavior. One relevant example could be job seekers' search behavior in the labor market. Job search behavior is often driven by people's own beliefs, perceptions, and values. It is not always the classical utility maximization that guides the final decision-making, but instead, the job seeker's attitude and beliefs toward a job position and its specifications.

For this specific reason, we start with the heuristic that the demand for job positions that are far away from firms' main location, is considerably lower compared to similar job positions located in the headquarters. First, job seekers prefer working near headquarters because they believe that it generally offers better wages, benefits, prestige, and promotion opportunities. Therefore they tend to neglect other vacancies located away from where the firm is based. This leads to a decrease in the total number of applications, which consequently affects the time it takes to fill such vacancies. Second, a lower demand arises when people who live in the area where the job position is located, oftentimes lack the knowledge about middle-big sized firms, thus the radiance as an employer of bigger firms suffers. They choose not to apply for such positions because they are not convinced by the insufficient and possibly wrong information they have.

3.2 Baseline Regression and Identification Strategy

$$log(dur_{i,firm,q_t,loc}) = \alpha + \beta dist_{hq,loc_i} + covar_{i,firm} + \mu_{q_t} + \eta_{firm} + \theta_{prof_i} + \omega_{state} + \epsilon_{i,firm,q_t,loc}$$

Being far from a randomized assignment, we have to account for the omitted variable bias that could arise from not controlling for other potential covariates or unobserved confounders. To approximate the causal relationship we aim to establish between distance and duration, we include additional relevant variables as well as use the quasi-experimental design of fixed effects to control for omitted variables that are unobserved and constant over time or constant across units at a certain point in time (Angrist and Pischke, 2008).

3.2.1 Potential Additional Covariates

One potential variable could be the number of postings. We believe that the vacancy duration is positively correlated with the firm's search intensity and its resource designation for candidate search. If firms advertise their open positions more frequently, they may reach a larger audience and find the right candidate faster. Moreover, search intensity is presumably negatively correlated with distance from the headquarters. As firms have bigger recruiting teams in the headquarters, more resources are invested in filling vacancies near that location, resulting in more postings per position.
Table 3 shows that there is a negative correlation between the number of postings and the distance in our sample. However, the contrary is observed between the number of postings and duration.

An underlying reason behind the observed positive correlation can be that job positions that are posted more frequently, are the ones facing a lower demand due to skill shortage.

Correlation with	**Posting Count**
Duration	0.135
Log(Duration)	0.226
Distance	-0.035

Table 3: Posting Count's correlation with Duration and Distance

Additionally, the causality we are striving to come closer to could be overestimated if we leave out certain job postings' features which are generally associated with longer or shorter duration. For example, job positions, meant for applicants with a specific educational background, are believed to experience hiring difficulties to a larger extent. For that reason, a job position requiring a university degree may generally take more time to be filled than a fully comparable position without a university degree.
We can also argue that vacancy duration may vary depending on the job position's language, as every language has its own separate demand. Based on a simple supply and demand logic, a German state with a low share of foreigners in its population will face longer vacancy duration for international job positions and shorter for German ones.
Another characteristic of a job position is salary. Recruiters have been historically excluding the salary level from job postings until recently, aiming to reduce applicants' negotiating power and avoid competition between new and old employees. If the salary is posted together with the position, the applicant can start comparing with other colleagues, companies, or industries. If his analysis is not convincing enough, he won't apply at all and continue with better and more lucrative opportunities. If many applicants behave the same, it will be more difficult for the firm to fill their positions, especially in case of unattractive salaries
We also argue that the contract type may affect vacancy duration, by inducing different effort levels from both recruiters and applicants. Applicants tend to negotiate more and take more time to make their decision when applying for a permanent position. On the other side, recruiters also invest more time and resources in filling permanent positions, because of the expected higher added value those can bring to the company compared to the temporary ones.

In Figure 3 placed in the appendix, we provide plots with the distribution of each variable against vacancy duration, highlighting how their quartiles, median, and outliers change across both categories.

3.2.2 Fixed Effects Design

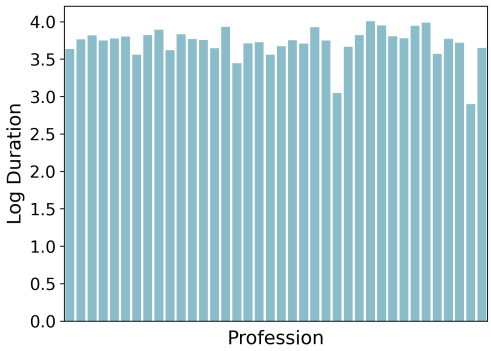
Adding all possible covariates into the regression cannot overcome omitted variable bias completely. In order to make a casual inference, we have to also account for the intrinsic characteristics of our individual observations (Angrist and Pischke, 2008). We are using a dataset with thousands of job vacancies across Germany during 2018. One single vacancy in our sample is located at a certain region, related to a certain profession, owned by a certain firm, and posted at a certain time. We can derive generalized and sound results only if we control for all unobserved time-invariant confounders. In the following paragraphs, we will justify our choice of fixed effects design as the sole identification strategy, by elaborating how this approach can contribute to our research analysis.

One factor that may encumber the settlement of our casual relationship is the vacancy position itself. We exploited the fact that there is a corresponding profession for each respective vacancy in our sample, to test for a possible intrinsic variation. The information in Plot 1 is the result of our computation. The large variation observed across different professions throughout the sample highlights the need for the inclusion of profession fixed effects to avoid the arising potential bias.

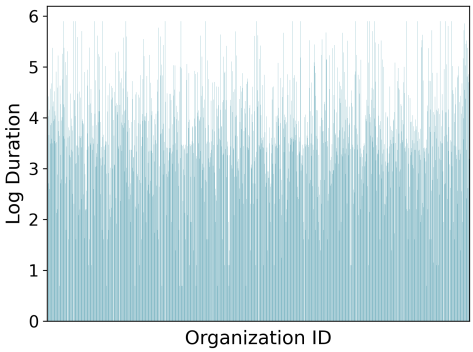
A job vacancy in a public firm listed on DAX 30 cannot be fully comparable to a job vacancy in a regional private firm. These two types of firms differ in size, reputation, and human resources. These underlying differences are eliminated once we include firm-fixed effects and control for all time-invariant and firm-specific factors. This allows us to capture within-firm variation only, by comparing vacancies within the same firm. Plot 2 below displays the high variation in average vacancy duration across the firms in our sample, supporting the inclusion of such fixed effects in our main model.

The same argument is valid also for job vacancies located in different states. A vacancy for a job position located in Bavaria is completely different from a vacancy located in Saxony. Differences in their working population, average education level, housing costs, and living conditions, can induce a high variation in the duration of job vacancies. In Plot 3, we notice that vacancies across German states do experience different duration. This implies that by leaving out state-fixed effects from the regression, we can falsely attribute this variation to the wrong variable.

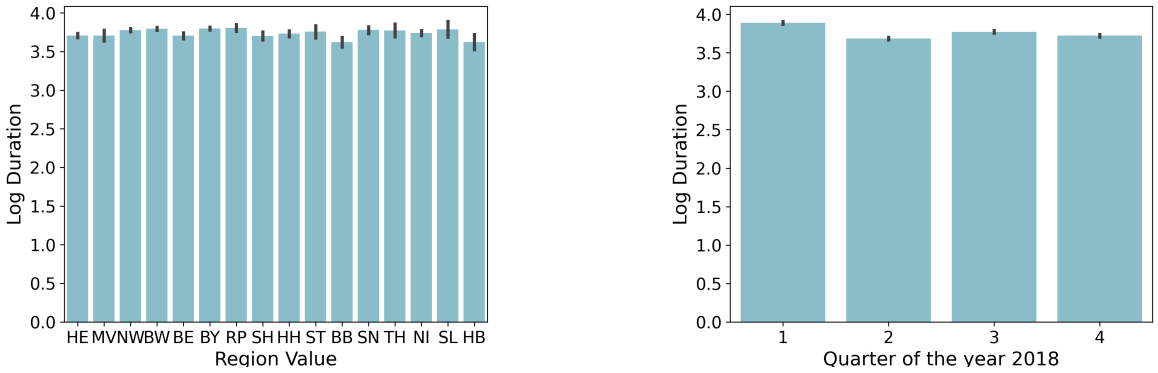
Lastly, as we analyze observations that are spread throughout the year 2018, we need to be cautious of potential shocks that may affect all vacancies at a specific time. We calculated the average duration per quarter and the output is displayed in Plot 4 below. We decided to include quarter-fixed effects to account for any time-varying factors common to all observations, albeit there is little variation across quarters with respect to job vacancy duration.



(a) Plot 1



(b) Plot 2



(a) Plot 3 (b) Plot 4

Figure 2: Duration variation across fixed effects variables

3.2.3 Standard errors

Our linear regression model may not adhere to the assumption of normally distributed errors. The presence of outliers and measurement error in our dependent variable can result in heteroscedasticity, which in turn, leads to small standard errors and inaccurate inferences. To address this issue, we employ robust standard errors, which are less sensitive to this violation, and provide more precise standard error estimates in such scenarios.

We have also considered clustering at a certain stage of our analysis. In the absence of randomization and with a sample of thousands of vacancies over time, the issue of correlation between error terms becomes prominent. In our setting, vacancy residuals can potentially be correlated within the same firm over time, leading to potentially inaccurate inferences. There is a risk that different vacancies from a certain firm may share identical characteristics of both their originating firm and the time they have been posted. To account for this potential correlation, we cluster standard errors at the firm-time level, aiming for more accurate standard errors and p-values.

4 Analysis and Results

4.1 The effect of distance on vacancy duration

In this section, we empirically analyze the relationship between a job position's distance from the headquarters and the duration of its vacancy. Through our analysis, we evaluate how the impact and significance of this relationship evolve when additional variables are taken into consideration and stricter conditions are applied.

Variables	(1)	(2)	(3)	(4)
Dist.	0.00030***	0.00028***	0.00024***	0.00023**
	(0.00002)	(0.00003)	(0.00003)	(0.00010)
Post count		0.06638***	0.06710***	0.06710***
		(0.0008)	(0.00085)	(0.00209)
Salary posted - True		0.26336***	0.27284***	0.27284**
		(0.02006)	(0.0225)	(0.13656)
Uni. degree - True			-0.00688	-0.00688
			(0.011)	(0.01816)
Contract - perm.			0.04770***	0.04770
			(0.013)	(0.04831)
Language - German			0.06357*	0.06357
			(0.029)	(0.0547)
Constant	3.71681***	3.47746***	3.38718***	3.38718***
	(0.0058)	(0.0078)	(0.032)	(0.084)
Fixed effects				
Firm FE		✓	✓	✓
Quarter FE		✓	✓	✓
Profession FE			✓	✓
State FE			✓	✓
Standard errors	robust	robust	robust	Cluster Firm-Time

Table 4: Main Regression Model Results

The analysis progressed from a less restrictive specification to a more stringent one, incorporating a variety of potential covariates, fixed effects, and clustering. The initial specification, a simple OLS regression, revealed a statistically significant and positive relationship between distance from headquarters and vacancy duration, though it was deemed inconclusive. In the second specification, we controlled for posting counts and salary inclusion, as well as for time-invariant firm-specific variables and any time-specific event. In the third specification, we accounted for job language, university degree, and contract type, as well as addressed all time-invariant factors specific to the profession and state. The final specification included all of the aforementioned controls and also firm-time clustering to account for the potential correlation within firms over time.

The distance from the headquarters has a significant and positive impact on the duration of job vacancies, but the effect becomes less pronounced as the specifications become more stringent. An increase of one standard deviation in distance is associated with a 4.4% increase in vacancy duration, while holding other factors constant. This relationship provides plausible evidence for a positive association between distance from the headquarters, suggesting that job positions further away from the headquarters take longer to fill.

Both numbers of postings and the inclusion of salary information have a positive and statistically significant impact on the duration of a vacancy. Specifically, an additional posting per position is associated with a 6.7% increase in vacancy duration, and the inclusion of salary information can lead to a 27.2% increase in vacancy duration, ceteris paribus. These results do not support the notion that increased search intensity shortens the hiring process. Rather, they suggest that job postings that are posted more frequently may take longer to fill in general. Furthermore, our results highlight the importance of salary as a crucial factor in determining the duration of

a vacancy, with the mere inclusion of salary information having a significant impact on vacancy duration, regardless of the specific salary amount offered.

Surprisingly, the coefficient of university degree is negative and statistically insignificant across all specifications. This contradicts our initial expectation that a requirement for a specific university degree would lead to an increased duration of a vacancy. These findings suggest that further research is needed to better understand the relationship between education level and vacancy duration. On the other hand, both job language and contract type offered have a positive statistically significant impact on vacancy duration, after controlling for any state, time, firm, and profession-specific factors. Job positions requiring knowledge of the German language tend to have a 6.3% longer duration of vacancy compared to positions requiring knowledge of an international language. Additionally, job positions offering a permanent contract have a 4.7% longer duration of vacancy compared to similar positions offering a non-permanent contract. However, when clustering at the firm level, both job language and contract type lose their significance, leading to inconclusive statements.

4.2 Heterogeneity across subgroups

Up until now, we have only discussed the unique effect of distance. But we haven't considered the fact that distance may depend on another variable and consequently, its impact may vary among selected subgroups. The relationship that we are aiming to settle in this paper, can possibly be influenced by other factors that distinguish one vacancy from the other.

Our research provides shreds of evidence that being farther away from the headquarters is positively associated with longer vacancy duration. The extent of this association may, however, vary depending on vacancies' specific characteristics such as whether it offers a permanent contract, requires German fluency, or a particular university degree. In Figure 4, we indeed observe a considerable variability in the distance effect on vacancy duration, based on certain values of all three above-mentioned variables. To statistically test this hypothesis, we extended our baseline regression by including interaction terms in the analysis.

In the second specification in Table 5, we can see that the coefficient of distance changes its sign, once we interact it with the job language of the vacancy. The positive association is passed to the interaction term, suggesting that German-speaking people are the main reason behind the increase in duration. This may point to the direction that German applicants are often reluctant to move and work in different locations other than headquarters. The third specification shows that the impact of distance on vacancy duration can decrease by 1% when the job position of that vacancy requires a specific university degree. This concludes that a job position located 100 km away from the headquarters location has a heterogeneous effect on duration depending on what educational background is required. Lastly, in the fourth specification, when we interact distance with the contract type of vacancy, the significance of the distance disappears, and the positive association is entirely driven by the type of contract offered at the end of the hiring process.

Variables	(1)	(2)	(3)	(4)
Dist.	0.00030***	-0.00054***	0.00029***	0.00002
	(0.00002)	(0.00014)	(0.00004)	(0.00006)
Post count		0.06706***	0.06705***	0.06714***
		(0.0008)	(0.0008)	(0.0008)
Salary posted - True		0.27140***	0.27295***	0.28158***
		(0.0225)	(0.022)	(0.022)
Uni. degree - True		-0.00629	0.01097	-0.00460
		(0.0112)	(0.0133)	(0.011)
Language - int.		-0.04903	0.06403*	0.06313*
		(0.034)	(0.029)	(0.029)
Contract - perm.		0.04801***	0.04722***	-0.00001
		(0.01283)	(0.0128)	(0.0158)
Language - German x Dist.		0.00080***		
		(0.0001)		
Uni. degree - True x Dist.			-0.00012*	
			(0.00005)	
Contract - perm. x Dist.				0.00028***
				(0.00006)
Constant	3.71681***	3.49385***	3.37792***	3.42395***
	(0.0058)	(0.035)	(0.032)	(0.033)
Fixed effects				
Firm FE		✓	✓	✓
Quarter FE		✓	✓	✓
Profession FE		✓	✓	✓
State FE		✓	✓	✓

Table 5: Main Model Extension - Heterogeneity effects

4.3 Robustness Checks

The ultimate step in our research analysis is the confirmation of our main findings' validity. In Table 9 in the appendix, we modify the third regression specification several times to test our results' sensitivity and consistency. We first replicate it, by replacing distance with a binary variable that simply indicates whether the vacancy is located in the same city as the firm's headquarters or not, while not changing anything else. The regression output, attached in the appendix, shows that vacancies located in the same city as the firm's headquarters experience a lower duration than comparable vacancies but located in another different city, aligning with the positive association we derived between distance and vacancy duration. Surprisingly, once we include both distance and the variable indicating the same city in the specification, we observe that the distance variable becomes insignificant, implying that the positive association with duration is driven by the reluctance to move across cities rather than the actual amount of proximity to headquarters.

Secondly, we replicate the third specification again, adding two numerical variables as representative of the salary level. The effect of both resulted insignificant, despite our attempts to create a quantitative and meaningful measure of salary. This emphasizes once again the limitations we experience with this particular covariate as well as the need for further research which can expand on the importance of salary level as a determinant of vacancy duration.

5 Discussion

In our paper we aim to explore various factors of firms' hiring difficulties. We focus on disentangling the effect of the distance of a job position from the firm's headquarters on its vacancy duration. While we use a granular dataset, and therefore implement a strict empirical method, without losing on significance, we want to address some remaining shortcomings here. The dataset contains mostly qualitative and categorical data, which makes it difficult to address the most obvious flaw in our empirical design: The absence of reliable salary data. While we argue that most of the variation associated with the salary is covered through profession and firm fixed-effects, the usage of the former can only be considered as second best. We did address this issue by running robustness checks with a simple mean-extrapolated version of the salary, but more advanced imputation techniques like the K-nearest-neighbor or MICE could not be implemented because of various reasons: The KNN algorithm needs to use the distance between observations. To compute a classical distance, numerical variables are needed. Since almost all of the variables in the dataset are categorical, this algorithm could not be considered. The MICE algorithm would need more data preparation for its usage, which was out of scope for our purposes and is memory and computationally intensive, which forbids us from using this particular method. As the distribution of individual variables are not changed in any meaningful manner, we used this pragmatic way.
Apart from the missing salary, other possible limitations of our research are the problem of measurement errors and the absence of other possible relevant variables, especially a variable which indicates the success of a particular job search. While regression dilution potentially biases our results towards zero, this would only further emphasize the final results. While not claiming to establish a causal relationship between the distance of the headquarter and the location where the job needs to be filled, we indeed find a strong association between these two variables. Our findings are in line with the existing theoretical and applied findings of people preferring working in the headquarter rather than working in a branch.

6 Conclusion

In this paper, we investigate the impact of the distance between a firm's branch and its headquarter on the vacancy-filling duration in Germany. To specify, our estimation results indicate that a 100 km increase in the distance between the branch and its headquarter is associated with a 2.4 % increase in the vacancy-filling duration for the vacancy located in the branch. This outcome has two main contributors which align with the literature. Firstly, potential candidates for a given vacancy are tended to apply with a higher likelihood if the vacancy is located in the headquarter. This results from higher wages, better working benefits, and higher prestige i.e. higher ego-rent presented by the vacancies located at the headquarter (Oosterbeek and Praag, 1995; Rogoff, 1990). Lastly, in the case of applications made to the branch, the likelihood of applying for a vacancy decreases as the vacancy is located further away from the headquarter. This circumstance arises from the time and financial costs of the movement process resulting from the probability of being transferred to the headquarter with increases in tenure and productivity (Miner and Estler, 1985). Since the time and financial costs of movement increase with the distance, the future value gained from working at the headquarter decreases (Manning and Petrongolo, 2017). In this way, the inclination of working at the branch decreases. Thus, as the distance becomes greater, the likelihood of applying for the vacancy decreases which increases the duration to fill vacancies located in the branch. Consequently, overall, our results assert that there exists a positive association between the vacancy-filling duration and the distance of the

vacancy to firm's headquarter due to the decreasing likelihood of applying for the vacancy as the distance increases.

References

Adam Cohen (2011). https://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/. Accessed: 2022-12-17.

Andrews, M. J., S. Bradley, D. Stott, and R. Upward (2008). "Successful Employer Search? An Empirical Analysis of Vacancy Duration Using Micro Data". In: *Economica* 75.299, pp. 455–480. ISSN: 00130427, 14680335. URL: http://www.jstor.org/stable/40071812 (visited on 01/16/2023).

Angrist, Joshua D. and Jörn-Steffen Pischke (Dec. 2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. ISBN: 0691120358.

Arhur Turrel (2022). *Coding for economists*. URL: https://aeturrell.github.io/coding-for-economists/intro.html.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger (2013). "THE ESTABLISHMENT-LEVEL BEHAVIOR OF VACANCIES AND HIRING". In: *The Quarterly Journal of Economics* 128.2, pp. 581–622. ISSN: 00335533, 15314650. URL: https://www.jstor.org/stable/26372507 (visited on 01/16/2023).

Gaudecker, Hans-Martin von (2019). "Templates for Reproducible Research Projects in Economics". https://doi.org/10.5281/zenodo.2533241.

Manning, Alan and Barbara Petrongolo (2017). "How local are labor markets? Evidence from a spatial job search model". In: *American Economic Review* 107.10, pp. 2877–2907.

Marinescu, Ioana and Roland Rathelot (2018). "Mismatch Unemployment and the Geography of Job Search". In: *American Economic Journal: Macroeconomics* 10.3, pp. 42–70. ISSN: 19457707, 19457715. URL: https://www.jstor.org/stable/26528444 (visited on 01/16/2023).

Miner, Anne S. and Suzanne E. Estler (1985). "Accrual Mobility: Job Mobility in Higher Education through Responsibility Accrual". In: *The Journal of Higher Education* 56.2, pp. 121–143. ISSN: 00221546, 15384640. URL: http://www.jstor.org/stable/1981662 (visited on 01/16/2023).

Oosterbeek, Hessel and Mirjam van Praag (1995). "Firm-Size Wage Differentials in the Netherlands". In: *Small Business Economics* 7.3, pp. 173–182. ISSN: 0921898X, 15730913. URL: http://www.jstor.org/stable/40228815 (visited on 01/16/2023).

Rogoff, Kenneth (1990). "Equilibrium Political Budget Cycles". In: *The American Economic Review* 80.1, pp. 21–36. ISSN: 00028282. URL: http://www.jstor.org/stable/2006731 (visited on 01/16/2023).

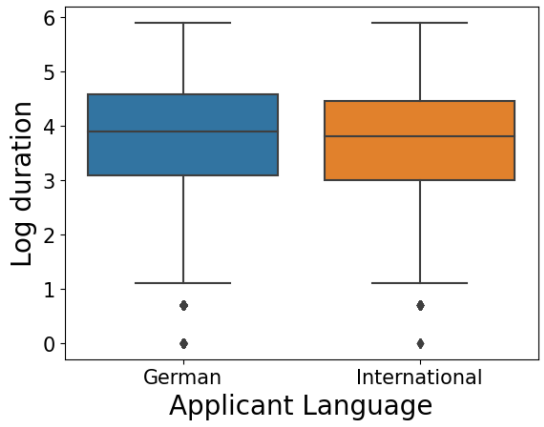
A Appendix

A.1 Variable Explanations

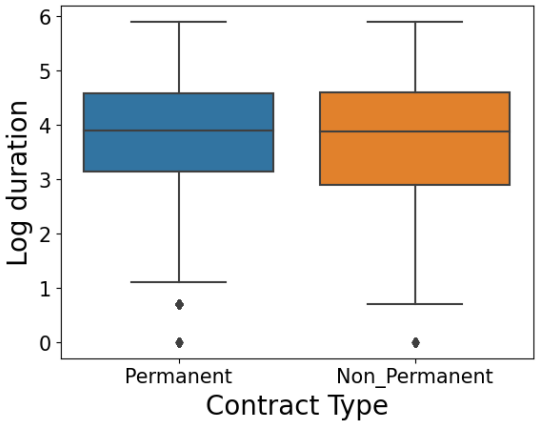
Variable	Coefficient or Variable Name	Explanation
Duration	dur	The duration, in days, that a given vacancy was online.
Distance	β	The distance between the advertised working location and the headquarters of the advertising organization.
Posting Count	Post count	Represents the number of times a given job was posted. We use this variable as a proxy for hiring intensity.
Region	ω	Represents the region where the advertised job is located. Can take the value of each German state.
Profession	θ	An ID that represents the profession of the job advertised. For similar jobs, the first digits are equal.
Salary	Salary posted - true	An integer that shows the yearly salary for the advertised job.
Time	μ	A variable that shows the date when a given job was posted. We clustered these dates into quarters.
Organization	η	An ID that uniquely identifies each firm.
University Degree	Uni. degree - true	A binary variable that indicates if a university degree is mandatory to fill the advertised job position.
Contract type	Contract - perm.	A binary variable that indicates if the advertised job has a permanent contract or a temporary contract.
Language	Language - German	A binary variable that indicates the language requirements (German vs. International) for the advertised job.

Table 6: Table of variables and coefficients used in the study

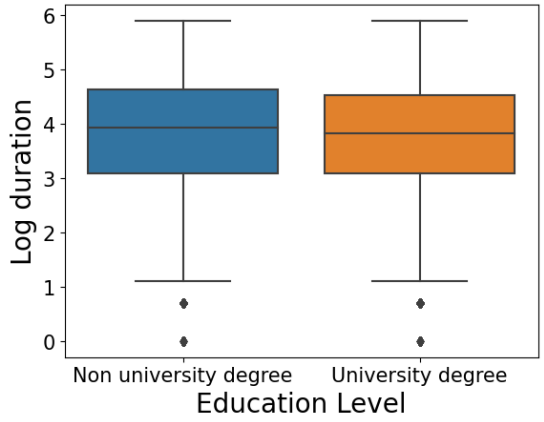
A.2 Potential Covariates: Plots



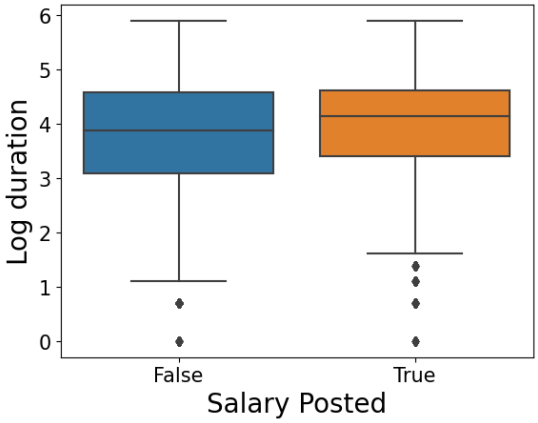
(a) Plot 1



(b) Plot 2



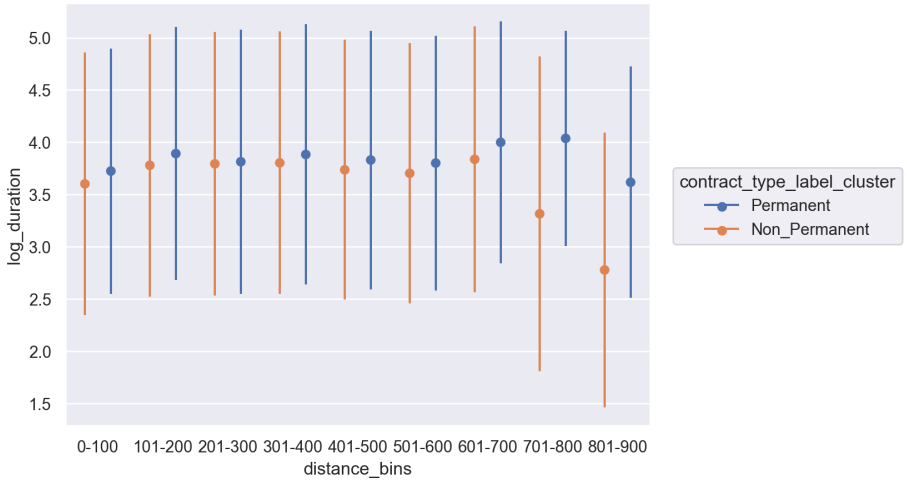
(c) Plot 3



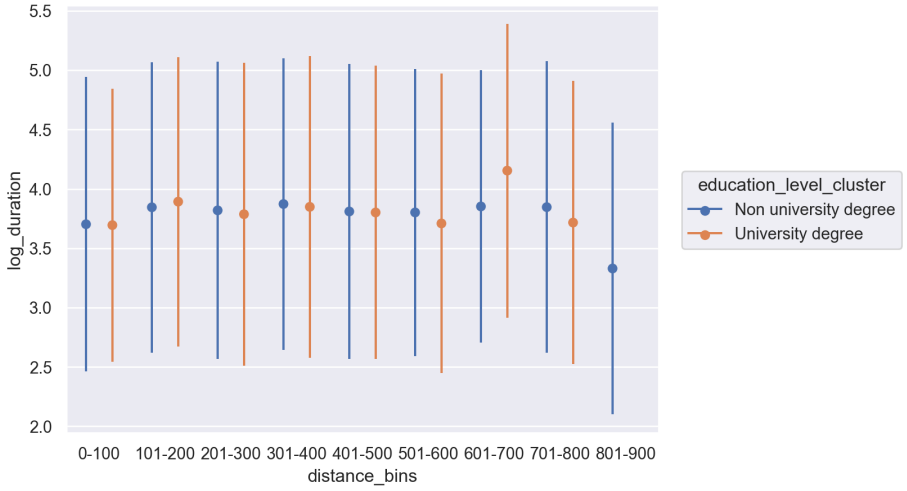
(d) Plot 4

Figure 3: Distribution of categorical covariates against Log Duration

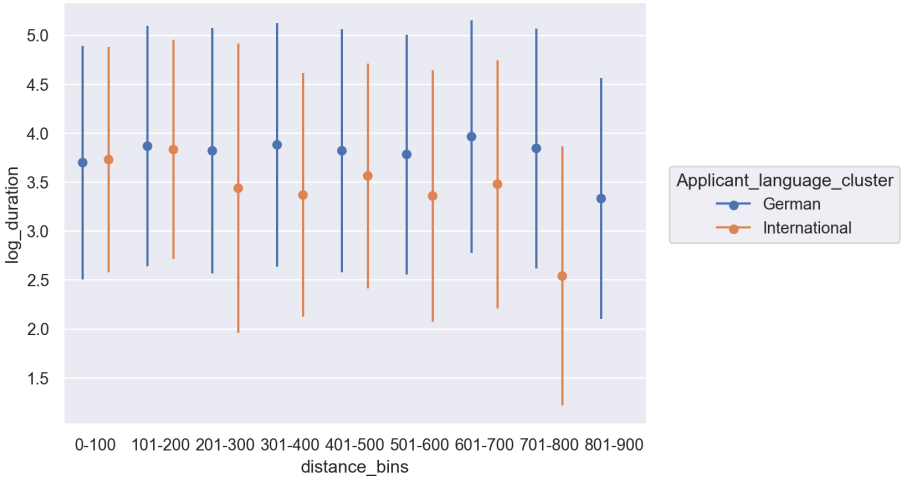
A.3 Potential Interaction terms: Plots



(a) Plot 1



(b) Plot 2



(c) Plot 3

Figure 4: The variability of distance effect across selected subgroups

A.4 Robustness Checks

A.4.1 Cleaning process

In this part of the Appendix, we provide summary statistics and histograms for the most important variables used, before the data cleaning process. This part should ensure that the distributions did not change in any meaningful manner. For the duration variable, the cleaning process did shift the whole distribution away from zero by an approximate amount of 15 %. The standard deviation remains virtually unchanged. The density function keeps the same shape, indicating that the relative values of duration in the dataset remain the same. The distance metric is changed in the inverse way by the cleaning process. The average distance increases by 27.6 km through the cleaning process, with an increase in the standard deviation.We argue that this change in descriptive metrics for the distance only effects our main coefficients negatively, i.e., a bias towards zero. This argumentation is based on the fact that the effect is mostly driven by observations with smaller distances. The posting count is skewed towards the higher values by the cleaning process, showcased by the higher 75 percentile. The mean and standard deviation are similar across versions.

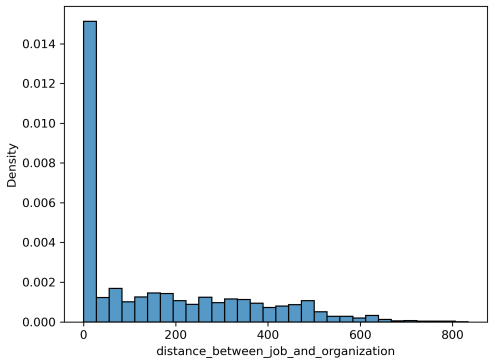
Categorical variables are across not changed by a magnitude higher than 15 %, with the highest change in percentage for the Contract type variable, followed by the education level variable. The change in relative shares for these two variable is driven by non-regular working hours. Non-regular working hours are not part of our research question; Therefore these changes are expected.

	Duration (in days)	**Posting Count**	**Distance**
Mean	67.1	2.6	135.3
Std. Deviation	81.5	3.7	170.6
Min	1.0	1.0	0.0
Max	365.0	20.0	841.6
25% Percentile	14.0	1.0	3.4
50%	39.0	1.0	37.6
75%	80.0	2.0	244.0
Count	1512018	1492609	1127626

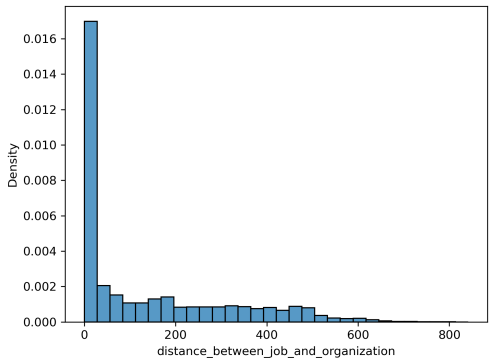
Table 7: Summary statistics of continuous variables, before the data-cleaning process

Variable	**Category**	**Relative Share**	**Count**
Salary Posted	False	88.0%	1527554
Salary Posted	True	12.0%	1527554
Applicant Language	German	95.9%	1490624
Applicant Language	International	4.1%	1490624
Education Level required	No University degree	63.7%	713919
Education Level required	University Degree	36.3%	713919
Contract Type	Permanent	88.4%	1149272
Contract Type	Non Permanent	11.6%	1149272

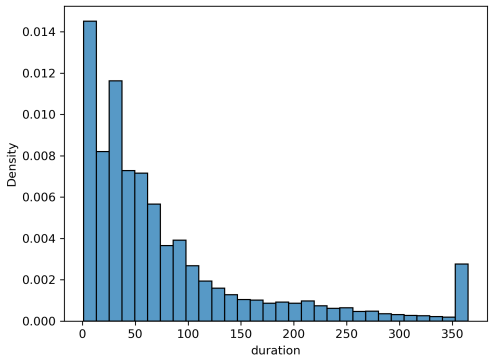
Table 8: Summary statistics of categorical variables, before the data-cleaning process



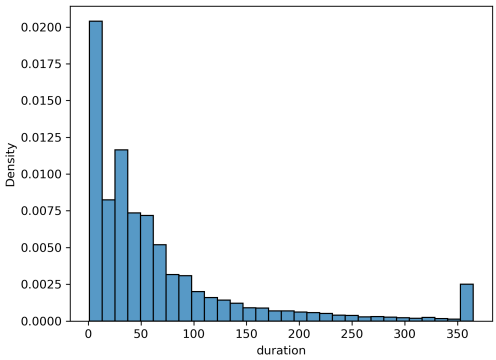
(a) Distance cleaned dataset



(b) Distance uncleaned dataset



(a) Duration cleaned dataset



(b) Duration uncleaned dataset

Figure 5: Comparison of distance and duration distribution across the datasets

A.4.2 Regressions

As a first robustness check, we examine the magnitude of the effect when using a dummy variable, which indicates if the location of both, the headquarter and the job to be filled, are the same city instead of the distance metric. As a heuristic, we used a distance of 30 km as the threshold for the same city. All distances above that threshold are counted as a different city. This particular approach was chosen to overcome the following obstacle: Even in the same city, names can differ widely. As an example, "Bonn" and "Bad Godesberg" are virtually the same city, but a fuzzy matching algorithm cannot match these two cities as one. The empirical specification is otherwise unchanged compared to the third specification of the main analysis. We include all four fixed effects and all considered covariates. As expected, the coefficient for the same city covariate is negative and significant, indicating that jobs located within the same city as the headquarter are filled faster than outside the city. Jobs within the same city are associated with a 13% decrease in duration to fill a given job.

As a last robustness check, we again consider our third specification, but instead of using a salary dummy, we use extrapolated versions of the salary. This extrapolation does not impute missing values but is an attempt to make the salary variable more meaningful to our research purposes. To construct the salary, we used two simple extrapolation methods: 1) We multiplied all salary variables, which are below a threshold of 20000 Euros with the value of 12. This process should

account for values which are below the minimum yearly wage in Germany for a full-time position. 2) We replaced these values with the mean salary in the dataset. Both of these methods lead to parameters which are virtually zero and not significant. While this procedure could, in theory, account for one specific source of measurement error, more advanced methods or an ensemble of other methods are needed, to make this variable useful in a meaningful manner.

Variables	**(1)**	**(2)**	**(3)**
Dist.		0.00068***	0.00069***
		(0.00012)	(0.00001)
Same city	-0.134***		
	(0.013)		
Post count	0.068***	0.071***	0.071***
	(0.0008)	(0.0041)	(0.0040)
Uni. degree - True	-.0007	-0.21*	-0.23**
	(0.011)	(0.077)	(0.077)
Contract - perm.	0.049***	0.19***	0.020***
	(0.013)	(0.051)	(0.055)
Language - German	-0.063**	-0.43*	-0.43*
	(0.029)	(0.25)	(0.24)
Salary posted - True	0.267***		
	(0.022)		
Salary extrapolated mean		$-1.6 * 10^{-6}$	
		$1.5 * 10^{-6}$	
Salary extrapolated multipl.			$-7.2 * 10^{-7}$
			$5.3 * 10^{-7}$
Constant	3.478***	4.0***	4.06***
	(0.032)	(0.25)	(0.25)
Fixed effects			
Firm FE	✓	✓	✓
Quarter FE	✓	✓	✓
Profession FE	✓	✓	✓
State FE	✓	✓	✓
Standard errors	robust	robust	robust

Table 9: Robustness Regression Model Results