Training Specificity and Occupational Mobility: Evidence from German Apprenticeships

Dita Eckardt*

March 30, 2024

Apprenticeships play a key role in enabling successful school-to-work transitions in many countries but, in the presence of imperfect information, the specificity of this type of training may entail important costs for those working outside their training fields. I study this issue in one of the most prominent training settings, the German apprenticeship system. Using administrative data and a broad occupational classification, I find that 40% of individuals work in occupations different from their training. I estimate the cost of mismatch using vacancy instruments and extend methodological approaches in high-dimensional selection settings. Lacking training in one's occupation entails an average wage penalty of 14%, the equivalent of two years of work experience. The penalty increases with the task distance between training and occupation. My findings suggest that retraining is crucial to mitigate the adverse consequences from imperfect information in specialized training settings.

JEL codes: I26, J24, J31, J62

Keywords: *Vocational training, occupational choice, high-dimensional selection, human capital specificity, task content of occupations*

---

*Department of Economics, University of Warwick, Email: Dita.Eckardt@warwick.ac.uk.
A previous version of this paper was circulated under the title "Are Chemists Good Bankers? Returns to the Match between Training and Occupation". I am very grateful to my advisors Alan Manning and Johannes Spinnewijn for their support and guidance on this project. I thank Karun Adusumilli, Joseph Altonji, Manuel Bagues, Camille Landais, Guy Michaels, Steve Pischke, Uta Schönberg, John Van Reenen, and many seminar participants for helpful comments and discussions. This study uses the factually anonymous Sample of Integrated Employment Biographies (version SIAB-R 7510). Data access was provided via a Scientific Use File supplied by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

1 Introduction

The transition of young people into the labor market is a primary concern for policy makers around the world. A key question is whether this transition is best facilitated through specialized vocational or more general academic education. Among high-income countries, vocational training plays a major role in the school-to-work transition in countries like Germany, Austria or Finland, while many others including the United States or the United Kingdom rely more heavily on university-based education which delivers more general knowledge. A prominent type of vocational training that is common in German-speaking countries are so-called dual apprenticeships. These apprenticeships provide occupation-specific skills by combining training in firms with education in vocational schools. It has long been argued that the high-quality provision of occupation-specific marketable skills explains why countries with dual training systems have amongst the lowest youth unemployment rates in the world (e.g., Quintini *et al.* (2007)).

A potentially major concern with training that provides specific skills is the lack of flexibility this specialization entails (e.g., Hanushek *et al.* (2017)). Skills may not be used if workers work in occupations they did not train in following changes to labor market conditions, their preferences or abilities. The lack of transferable skills may also lock workers into the area they trained in, preventing them from taking advantage of opportunities in other fields. In contrast, general education delivers more flexible skills that can be transferred across occupations, although providing it may be more costly (e.g., Goldin (2001)). Consistent with lower levels of skill portability, workers in countries with more prevalent vocational training systems are more likely to report that they work in occupations related to their field-of-study, and they change jobs less frequently.[1] Understanding the importance of this lack of flexibility is crucial to assess the trade-offs inherent in designing vocational education systems, and address potential frictions that workers within these systems face.

The extent to which the specificity of training matters in practice depends on two key measures: the share of workers working in an occupation different from their training and the cost of these moves. An analysis of the latter is particularly challenging, and although the dual apprenticeship system has been termed a role

[1] OECD cross-country statistics and survey data from the Programme for the International Assessment of Adult Competencies (PIAAC) (see Section 2.1).

model by policy makers around the world, quantitative evidence on the effects of the specificity of skills delivered in the system is thin. Importantly, causal identification of the returns to training-occupation matches requires accounting for selection into a training, and subsequent selection into one of many occupations. In light of this challenge, existing studies are descriptive, typically finding small costs to occupational mobility in the dual system (e.g., Clark & Fahr (2001), Göggel & Zwick (2012)).

This paper aims to address this gap in the literature. I consider the specificity of training in the German dual apprenticeship system. Apprenticeships are the main form of post-secondary education in Germany, held by around two thirds of those who continue their education after high school. The training spans a range of occupations, allowing for the analysis of a broad spectrum of tasks and heterogeneity in occupational moves.

To address the empirical challenges, I use administrative panel data for 1975-2010. The data contains information on the occupation workers are trained in and a range of labor market outcomes in subsequent employment spells, including occupations. Given data availability, I consider a system with 13 aggregated occupation categories. Based on these, I show that, at any point in time, an average of 40% of individuals trained in an occupation work in another. Given the broad occupational classification, this rate appears striking in a system that aims to deliver highly specialized skills.

A potential reason for the high rates of occupational mismatch is lack of information at the time of training choice and changing labor markets. Informational problems are prevalent in educational contexts (e.g., Zafar (2011), Saniter *et al.* (2019)), and they may be particularly severe for apprentices as they tend to be younger and from less advantaged backgrounds than university students. As individuals receive more information about their preferences, abilities or the labor market, they may choose employment outside their training. I provide a range of evidence in support of such learning mechanisms. Most importantly, I show that workers become less likely to work in their training occupation with more time spent in the labor market.

Occupational moves may not pose a concern if training received in one occupation is valuable in other occupations. The challenge to identifying the returns to matching trainings with different occupations amounts to identifying Average Treatment Effects (ATEs) in a setting with multiple unordered treatments. To put structure on the selection problem, I set up a generalized Roy (1951) model. In the model, workers choose a training, and subsequently select an occupation in every work period. Training and occupation choices maximize expected payoff, and the latter may affect future payoffs through the accumulation of occupation-specific experience. Consistent with the empirical evidence, labor demand shocks or new information about preferences or abilities may lead individuals to choose employment outside their training.

To identify the returns to training-occupation matches based on this model, I extend the high-dimensional control function approach by Lee (1983) and Dahl (2002), and combine the administrative panel with data on the universe of occupation-specific apprenticeship vacancies posted via employment agencies for 1978-2010. The panel structure of the data allows for individual fixed effects to be included in all regressions. Holding constant time- and occupation-invariant ability differences, identification is based on variation in vacancies in outside options, conditional on own vacancies. I argue that a model in which apprenticeship vacancies are a sufficient statistic for occupation-specific demand justifies the exogeneity assumption, and provide a range of evidence to support this assumption. Since the instruments do not have full support, I additionally rely on parametric assumptions to identify ATEs, but these can be relaxed for certain parameters in the wage equation. I implement the control function approach using a random forest algorithm, where I estimate selection probabilities into trainings and occupations using the instruments. To account for potential endogeneity of *past* occupational selection, I follow Altonji & Shakotko (1987) and use the deviation of occupation-specific experience from its individual mean as instrument.

Implementing this strategy allows me to provide causal estimates of the returns to training-occupation matches under plausible assumptions. On average, I find returns to matching trainings with their corresponding occupation of 14%. The magnitude of this effect is comparable to OLS estimates of the return to two years of apprenticeship training (Krueger & Pischke (1995)). My results thus suggest that 40% of workers face an annual wage penalty of 14% from lacking training specifically in their occupation. Not controlling for selection leads to substantial negative bias in the estimated return so that, descriptively, those who work in their training occupations do not have higher wages. In line with the model, the sign of the bias suggests that only the relatively more able workers work outside their training as their unobserved occupation-specific ability needs to compensate for the lack of training.

The average return masks important heterogeneity across experience levels, trainings and occupations. The return drops by around half from its peak after 10 years, suggesting that workers with training in other occupations partially catch up by learning on the job. Across trainings, I find large differences in the returns to working in the corresponding occupation, and a positive correlation between these returns and the fraction of workers doing so. In line with the model, relative returns thus appear to be a key determinant of occupational selection. Within trainings, there is substantial heterogeneity in returns across occupations. To test how training skills relate to these returns, I use survey data to construct training-occupation task distance measures. Regressing the estimated returns for each match on these measures, I find that a one-standard-deviation higher task distance reduces returns by about 7 percentage points. These findings suggest that workers are trained in a mix of tasks and face higher wage penalties, the less applicable their skills are to their occupation.

To assess the welfare implications of my findings, I combine the selection model with the empirical results and show that the welfare loss from imperfect information at the time of training choices amounts to at least 3% per worker in the system. Back-of-the-envelope calculations suggest that ex-post retraining could effectively address the ex-ante lack of information for a large group of workers.

This paper relates to several strands of literature. Firstly, it contributes to the literature comparing vocational training to more general education. A set of studies argues that general education is less cost effective for individuals who spend their working lives in the same occupation, but it enables workers to adopt new technologies and promotes economic growth in times of technological change (e.g., Goldin (2001), Krueger & Kumar (2004a,b)).[2] At the individual level, an early set of descriptive papers shows that specialized training is associated with economic benefits relative to general education only when workers match their skills to related occupations (e.g., Fredland & Little (1980), Neuman & Ziderman (1991)). In deriving estimates of the occupation-specific returns to apprenticeships under plausible assumptions, this paper is the first to show that the costs of occupational mismatch are large and affect a sizable share of the German population.[3] To explain the latter result, I show that early specialization comes with information problems which lead to substantial mismatch and resulting welfare losses.

This finding is relevant for lower-income countries where workers are likely to face

[2] Testing this hypothesis, Malamud & Pop-Eleches (2010) find similar outcomes for graduates from general and vocational schools during Romania's transition to a market economy.

[3] A related literature studies the institutional environment that incentivizes (German) firms to provide training in marketable skills (e.g., Becker (1964), Acemoglu & Pischke (1998), see Wolter & Ryan (2011) for a review of this literature).

more severe information problems and placement rates of graduates from vocational training programs are particularly low (Bennell (1996)). In these settings, the lack of marketable skills is a key contributor to poor labor market outcomes of young people and training programs are common, but take-up can be low (e.g., Alfonsi *et al.* (2020), Caicedo *et al.* (2022)). My findings show that high rates of occupational mismatch can substantially lower the expected returns to training, and reduce training take-up even when specialized skills are highly valuable if matched to relevant occupations.

Secondly, this paper contributes to the literature on identification of treatment effects in high-dimensional selection models. Lee (1983) and Dahl (2002) develop an estimator for these settings where the control function is a function of a small set of selection probabilities. A recent application of Dahl's approach is Ransom (2021) who studies how the returns to college majors vary across locations and occupations, accounting for the simultaneous selection at the latter two margins while treating major choices as exogenous. I extend the Lee/Dahl approach to a two-stage sequential selection setting and combine it with a novel instrumental variables strategy that uses occupation-specific covariates, similar to Heckman & Sedlacek (1985, 1990) and D'Haultfœuille & Maurel (2013). In relying on instruments, my approach imposes fewer parametric assumptions than Lee (1983) and identifies ATEs of training-occupation combinations while accounting for selection at both margins.

Thirdly, this paper relates to the literature on occupational choice under uncertainty (e.g., Miller (1984), Siow (1984), Keane & Wolpin (1997), Nicholson (2002), Antonovics & Golan (2012), Arcidiacono *et al.* (2020)). Similar to the present paper, several of these studies use a Roy-type selection mechanism to model choices. But while the focus in these papers is to understand the drivers of occupational choices, the present paper is primarily interested in choices as a means to control for selection. As a result, it imposes a more limited amount of structure. A subset of studies incorporates the idea that human capital is not fully transferable across occupations without considering matches between education field and occupations (e.g., Keane & Wolpin (1997), Sullivan (2010), Todd & Zhang (2020)). A notable exception is Kinsler & Pavan (2015) who set up a model of occupational choice to estimate the returns to working in an occupation related to one's college major.[4] I ask a similar question in a different context, exploiting the availability of a training-occupation

[4] A set of other papers studies the same question, taking a more descriptive approach (e.g., Robst (2007), Nordin *et al.* (2010)).

matrix to estimate high-dimensional returns to training-occupation combinations.

Finally, this paper contributes to the literature on human capital specificity. This idea was proposed by Becker (1962, 1964) and extended by Lazear (2009) in the context of the firm, and has been taken to the data to explore specificity along several dimensions such as industry (Neal (1995), Parent (2000)), occupations (Shaw (1987), Kambourov & Manovskii (2009)) and skills (Poletaev & Robinson (2008), Guvenen *et al.* (2020)). Most recently, a strand of this literature suggests that human capital is partly task-specific, and thus more easily transferable across occupations that require a similar mix of tasks (Gathmann & Schönberg (2010), Yamaguchi (2012), Cortes & Gallipoli (2018)). The present paper contributes to this literature by systematically linking wages across occupations to training received in the same set of occupations. To the best of my knowledge, it is the first to provide such estimates.

The remainder of this paper is organized as follows. Section 2 outlines the setting and data. Section 3 provides descriptives on occupational mobility and discusses reasons for occupational mismatch. Section 4 sets up the generalized Roy model. Section 5 discusses identification and explains the estimation using control functions. Section 6 discusses the results. Section 7 relates my findings to task distances. Section 8 discusses welfare and policy implications. Section 9 concludes.

2 Setting and Data

2.1 The German Apprenticeship System

The German apprenticeship system is a dual system where apprentices work in firms for three to four days a week and go to vocational school for one to two days. Franz & Soskice (1995) provide a detailed account of the institutional setting. While the training in firms delivers practical skills, vocational schools teach theoretical skills in different subjects. The total apprenticeship length varies between two and three and a half years, but the majority of apprenticeships last three years.

Dual apprenticeships are the main form of education beyond the lower-secondary level in Germany and, in 2010, about two thirds of those with this education level had completed an apprenticeship in the dual system.[5] The dual system is regulated under

[5]Source: *Statistisches Bundesamt, Bildungsstand der Bevölkerung - Ergebnisse des Mikrozensus 2018.* Education beyond lower-secondary level corresponds to ISCED levels 3 and above, excluding high school qualifications.

a federal vocational training law which implies a large degree of standardization. It is often regarded as the key pillar of the German education system, supporting low youth unemployment rates by facilitating the transition into the labor market.

The dual system trains apprentices in most non-university occupations, with only a small number of exceptions in the medical and care occupations. To start an apprenticeship, high school graduates must apply to and be offered an apprenticeship position with a firm. The firm is then in charge of providing the practical training. The state government is responsible for providing a place at the local vocational school. The curriculum is centrally determined for each apprenticeship occupation and consists of general and specialized subjects. All dual apprenticeships are completed through a final examination which is organized and monitored by industry-specific boards. After completing their apprenticeship, apprentices often continue to be employed at the same firm as full-time employees.[6]

In comparison to other high-income countries, Germany's education system relies heavily on vocational training. Based on survey data from the OECD Programme for the International Assessment of Adult Competencies (PIAAC), the left panel of Figure 1 shows that the cross-country prevalence of vocational training correlates with workers reporting that they work in occupations related to their field-of-study.[7] The right panel shows that workers in countries with more widespread vocational training systems also change jobs less frequently, implying higher average levels of job tenure.

2.2 Data

This paper uses two main datasets: an administrative employment panel, and a dataset containing the universe of occupation-specific apprenticeship vacancies posted through local employment agencies.

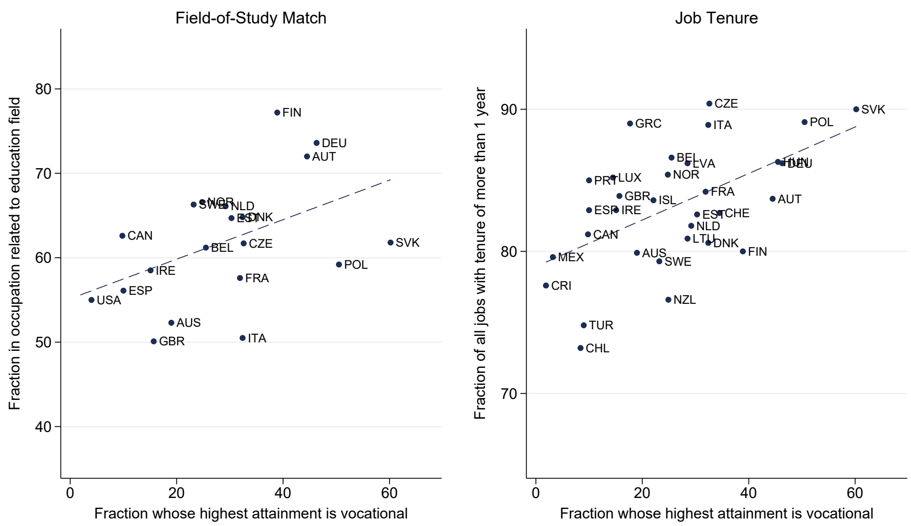
The employment panel consists of a 2% sample of German social security records between 1975-2010.[8] These records are based on all workers employed in that time period, with the exception of civil servants, self-employed and military workers ($\sim$ 80% of the workforce). Workers are followed for the entire sampling period. The data

[6] Around 60% in 2010. Source: *BiBB, Datenreport zum Berufsbildungsbericht 2010.*

[7] The field-of-study measure is taken from Montt (2015) and classifies PIAAC self-reports on one out of nine education fields and on 3-digit occupations into a binary match measure.

[8] Sample of Integrated Employment Biographies (version SIAB-R 7510). The data was provided by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Figure 1: Vocational Training and Field-of-Study Match/Job Tenure Across Countries



Notes: The figure shows latest data for OECD countries where available. Vocational attainment is at upper- or post-secondary non-tertiary level as fraction of all 25-64 year olds. US attainment corresponds to the fraction with vocational qualification at short-cycle tertiary level. Field-of-study measure for GBR (BEL) is based on England/Northern Ireland (Flanders) only. Sources: *OECD, 2024, Education GPS dataset: Educational attainment and labour-force status; OECD.Stat dataset: Employment by job tenure intervals - frequency* and Montt (2015).

is recorded at the spell level and spells can vary in length, but employers are required to report each employee at the beginning of a calendar year, so that spells last at most one year. Shorter spells may be recorded due to job changes during the calendar year or temporary unemployment. The data includes demographic information as well as daily information for each (un)employment spell including the occupation, industry, location and wage. Reported wages are capped at a time-varying threshold defined within the statutory pension scheme. This threshold only affects a small fraction of the sample (see Section 2.4). Importantly, since apprentices work in firms, they pay social security contributions and their apprenticeship spells are contained in the employment panel. I therefore observe the occupation that apprentices are employed in during their apprenticeship which I refer to as their *training*.

The second dataset contains the universe of apprenticeship vacancies posted through local employment agencies between 1978-2010.[9] Recorded vacancies include those

[9]This data combines datasets provided by the German Federal Employment Agency (BA). Sources: *Amtliche Nachrichten der BA, Arbeitsstatistik - Jahreszahlen, 1978-1993; Statistik der*

filled and those not filled and aggregate information is available by year, training and location. Yearly data is measured as a flow of vacancies posted between 1 October and 30 September, but most vacancies are posted to line up with schooling leaving dates in late summer. The institutional setting implies that apprenticeship vacancies likely reflect labor demand (see Section 5.3). A particular advantage of using data on apprenticeship vacancies is the high degree of involvement of employment agencies. In 2013, 71% of firms publicized their apprenticeship vacancies through an agency, while the same figure only amounted to 43% for non-apprenticeship vacancies in 2010.[10]

2.3 Field-Based Occupational Classification

Occupations in both datasets are coded using the same classification called *Klassifikation der Berufe 1988 (KldB88)*. This former German classification system was replaced by the current system in 2010. For the purpose of this paper, the *KldB88* has a key advantage over the newer and other international systems in that it is field-based. In particular, other systems generally contain a category for *managers* and as a result, promotions can imply occupation changes in the classification. It would be impossible to translate hierarchical categories of these classifications into a field-based system, and these measurement problems would be a major concern in the present analysis where the combination of training and occupation is of key interest.

Within the field-based classification system, the number of occupation categories varies with the granularity of the classification used. I use the finest occupational classification level for which the vacancy data is available, implying 13 categories and 169 cells in the training-occupation matrix. Restricting the number of categories also ensures that the estimation remains feasible. At the same time, the number of cells provides sufficient variability for the task analysis presented in Section 7.[11]

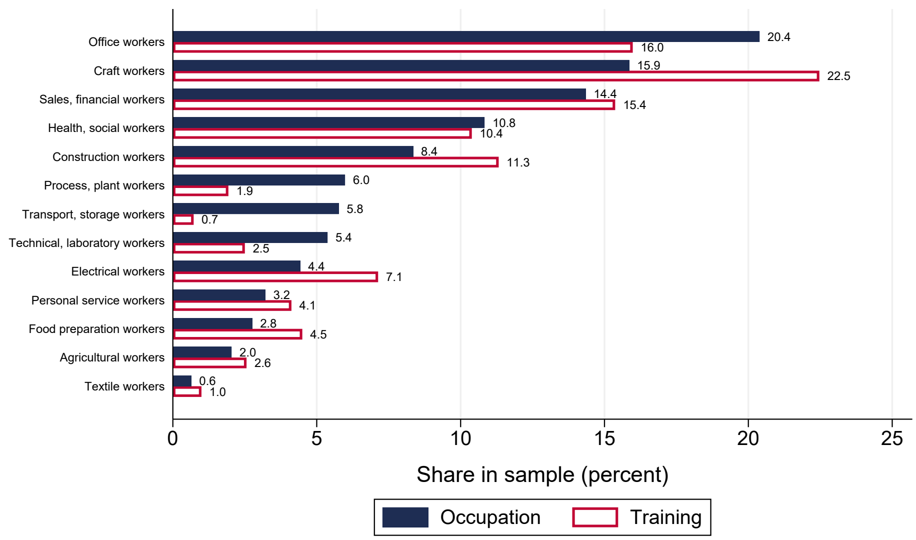
The list of occupations is exhaustive and dual apprenticeships train workers in all

BA, Seit Beginn des Berichtsjahres gemeldete Berufsausbildungsstellen, provided in May 2017.

[10] Source: *BiBB, Report 3/2014, Betriebe auf der Suche nach Ausbildungsplatzbewerberinnen und -bewerbern: Instrumente und Strategien; IAB, Brief Report 26/2011, Neueinstellungen gelingen am besten über persönliche Kontakte.* Posting rates across firm size and industries suggest that vacancies posted through employment agencies are representative of all vacancies (see first report for details).

[11] The classification level does not have intrinsic meaning, but welfare results are invariant to changes in the classification if there is an inverse relationship between occupational mobility and the average penalty of working in occupations different from one's training. I provide suggestive evidence that the findings are robust to using a more granular classification level (see Section 6.1).

Figure 2: Occupations and Trainings with Sample Shares



Notes: The figure plots the baseline sample shares by occupation and training. A detailed list of sub-categories contained in each occupation group is provided in Table A.1.

13 occupations.[12] A list of the categories and sample shares is shown in Figure 2.

2.4 Sample Selection

For administrative reasons and because workers can hold more than one employment relationship, spells can overlap in the data. I start by defining primary spells as highest wage spells, and only keep those in the sample (87% of sample days).[13] Starting with all individuals who went through apprenticeship training in 1975-2010, I then restrict the sample to spells of individuals who enrolled in *one* dual apprenticeship (85%), and who were classified as having completed their training (86%).[14] Finally, I only keep individuals whose training occupation and location are known (92%).

For the remaining individuals, I restrict the employment spells to full-time spells (83%), and exclude spells with missing location, occupation, and missing or zero wages, due to e.g. unpaid maternity leave (3%). Finally, I only keep spells that started after the end of the apprenticeship and for which employers recorded voca-

[12]The categories may contain occupation sub-categories that do not require training in the dual system, either because they require no formal training or because they require a university degree.

[13]Of the excluded secondary spells, less than 10% are employment spells.

[14]I exclude individuals not classified as having completed their training in any employment spell.

Table 1: Summary Statistics

	N	Mean	Min	Max	P^{10}	P^{50}	P^{90}
Observations/spells	4,012,034						
Individuals	291,098						
Female (% of indiv.)	45.4						
Female (% of spells)	37.3						
Individuals ever off diag. (%)	47.6						
Occ. switchers (%)	37.7						
Occ. switches per individual		0.7	0	38	0	0	2
Distinct occ. per individual		1.5	1	10	1	1	3
Age		30.6	17	62	21	29	43

Notes: The table reports summary statistics for the baseline sample.

tional training as highest level of education. This excludes spells with lower education levels (apprenticeship is not recorded as completed, 8%), and higher education levels (additional university or technical college degree, 5%), to ensure that education levels as measured by years of schooling are comparable across the sample. The resulting baseline sample contains 291,098 individuals and 4,012,034 employment spells.

Table 1 provides summary statistics for the baseline sample. About 48% of individuals work outside their training occupation for at least one spell. Since apprenticeship spells need to fall within the sampling period for all workers, the average age is only 31. As a result, wages are relatively low and less than 3% of wages exceed the upper earnings limit and are capped in the sample.

3 Occupational Mobility

A key input into an assessment of the effects of training specificity is the degree of occupational mobility. Section 3.1 presents descriptives on mobility of workers with apprenticeship training and provides comparisons to workers with other levels of education. Section 3.2 discusses evidence suggesting that lack of information at the time of training choice is the key driver of occupational mismatch.

3.1 Descriptives

Figure 3 summarizes the degree of occupational mobility by plotting the fraction of individuals working on the diagonal by work experience. While 75% of workers start their career in their training occupation, this fraction drops to 55% after 25 years. Figure B.1 shows that this is not due to compositional effects by plotting the fraction of on-diagonal workers over time for different experience levels. The likelihood of currently working off the diagonal, averaged across experience levels, is 40%.[15]

To put the rates of mobility into context, Figure B.2 offers a comparison between the mobility of apprenticeship, high school and university graduates by plotting the fraction working in their first post-education occupation by work experience. It shows that apprentices are least mobile, followed by university and high school graduates. In line with common intuition, this suggests that apprenticeship training is relatively specific. Apprenticeship graduates also have flatter wage-experience profiles than university graduates, suggesting lower returns to experience relative to initial skills and larger adjustment costs when changing occupations (see Figure B.3).

To give a sense of the distribution of workers across training-occupation cells, Table 2 reports the percentage of spells in each cell as a fraction of the training (first row) or the occupation (second row, in italics) for the largest trainings.[16] Spells are restricted to workers with ten years of experience. The on-diagonal shares display considerable heterogeneity across trainings and occupations, ranging between $55-85\%$.

3.2 Reasons for Occupational Mismatch

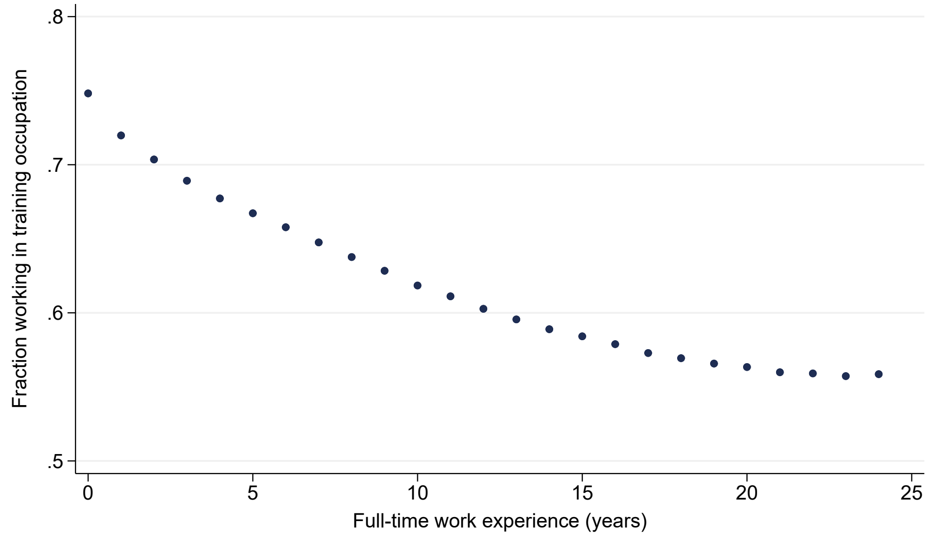
Section 3.1 provides evidence consistent with apprenticeships delivering relatively specialized skills. In light of this, the rates of occupational mismatch appear striking. A potential reason is imperfect information. If workers lack information on their abilities and preferences (types) or future labor demand when choosing a training, learning may lead them to work in occupations different from their training. In this case, training-occupation mismatch is caused by ex-post suboptimal training choices.

A number of papers find that information plays a key role in driving career choices

[15] The exact figure is 39.4%. This figure is derived by first averaging an indicator for off-diagonal work across sample spells for each annual work experience bin, and then averaging the experience-level averages across the 35 years of experience observed in the sample. It can therefore be interpreted as average likelihood of currently working off the diagonal across a worker's career.

[16] Table A.2 contains equivalent figures for all trainings/occupations.

Figure 3: Fraction On Diagonal by Work Experience



Notes: The figure plots the fraction of individuals working in their training occupation by full-time work experience for the baseline sample.

for young adults, showing that students update the information they hold about their abilities while studying (e.g., Zafar (2011), Stinebrickner & Stinebrickner (2014)), or that information on opportunities and labor market outcomes affects expected education choices (e.g., Oreopoulos & Dunn (2013), Wiswall & Zafar (2015)). Information problems tend to be more pronounced for students from less advantaged backgrounds (e.g., Hastings *et al.* (2016), Peter & Zambre (2017)), suggesting they could be particularly severe for apprentices. Apprentices are also younger than university students when choosing a training which could further exacerbate information problems. Most directly relevant to occupational mismatch in the German setting, Saniter *et al.* (2019) exploit quasi-random variation in the opening of job information centers, and show that providing high school students with information about occupational requirements improves labor market outcomes related to match quality.

Several patterns in the administrative data are consistent with workers holding ex-ante imperfect information. Most importantly, Figure 3 shows that workers become less likely to work in their training occupation with more time spent in the labor market. The convex pattern is consistent with most learning happening early in workers' careers. In addition, I find that less than 8% of individuals who move to

Table 2: Spells as Percentage of Trainings/Occupations - Largest Categories

		Occupation				
		Office workers	Craft workers	Sales, financ. workers	Health workers	Constr. workers
Training	Office workers	**80.6**	0.6	12.5	1.6	0.1
		59.4	*0.7*	*14.4*	*2.8*	*0.3*
	Craft workers	4.8	**55.3**	3.9	2.4	2.5
		5.0	***84.3***	*6.2*	*5.7*	*7.5*
	Sales, fin. w.	26.5	1.6	**60.6**	2.1	0.3
		18.2	*1.7*	***64.8***	*3.4*	*0.6*
	Health, soc. w.	12.2	0.7	4.3	**79.0**	0.2
		5.1	*0.4*	*2.7*	***76.4***	*0.3*
	Construction w.	3.5	5.7	3.1	2.9	**60.2**
		1.7	*4.0*	*2.3*	*3.2*	***85.1***

Notes: The table reports the number of spells in training-occupation cells as a percentage in the training (the occupation, in italics) for the baseline sample. Spells are restricted to those with ten years of work experience. Only the five largest occupations are reported.

an occupation different from their training move back within a year, consistent with learning but not with a temporary lack of job offers in the training occupation.[17]

Consistent with imperfect information about own types, I find that workers who were younger when choosing their apprenticeship are more likely to later work off the diagonal.[18] Workers may also learn about labor demand shocks, a particular source being automation (e.g., Acemoglu & Restrepo (2019, 2020)). In Germany, Dustmann *et al.* (2009) provide evidence in line with technology substituting for routine and complementing non-routine tasks in the 1980s and 1990s. Consistent with learning about these changes, I show that workers trained in occupations requiring a higher share of routine relative to abstract tasks saw larger declines in their on-diagonal shares between 1980 and 2000 (see Figure B.4).

[17]Specifically, both patterns are *inconsistent* with a basic search model explanation for mismatch where offers in the training occupation would arrive over time and lead to more suitable matches.

[18]On average, the fraction off the diagonal is 41.9% (36.5%) for below (above) median entry age.

4 Selection Model

I model the selection into trainings and occupations using a generalized two-stage Roy (1951) model. This section lays out the model, before I discuss the assumptions required to identify wage returns in Section 5.

4.1 Setup and Wages

Training and occupation choices are modeled as a two-stage selection problem. In $t = t_0$ (stage 1), individual i selects into a training $j \in \mathbb{J}$. In $t = t_0 + 1, ..., t_0 + T$ (stage 2), individual i selects into an occupation $k \in \mathbb{K}$. Note that the set of training and occupation options is identical, $\mathbb{J} = \mathbb{K}$.

In stage 2, if i works in occupation k with training j, their log wages follow

$$ln(w_{ijkrt}) = \delta_r + \delta_t + f(ln(vac_{krt})) + \delta_i + \tau_{j(i)k} + \beta' X_{ikt} + \epsilon_{ikrt}, \tag{1}$$

where δ_r, δ_t, δ_i denote region, time and individual fixed effects. Equation (1) is the main empirical equation of interest. The term $f(ln(vac_{krt}))$ denotes a function in log vacancies posted for occupation k in region r at time t. X_{ikt} includes full-time general and occupation-specific work experience, exp_{it} and exp_{ikt}, and their squares, and ϵ_{ikrt} is an individual error that varies across occupations but not trainings. This captures the idea that unobserved occupation-specific abilities affect wages, but there is no unobserved heterogeneity in the ability to productively use a training in an occupation. The fixed effects $\tau_{j(i)k}$ are the parameters of interest, and they capture the log wage effect from a training-occupation match jk. The specification in Equation (1) may be derived from a standard exponential human capital production model (Griliches (1977)), where log wages are the sum of a log skill price and log human capital (see Appendix C for details). In the following, I mostly suppress the dependence of j on i, taking as implicit that trainings only vary across individuals.

4.2 Occupation Choice

In period $t > t_0$, occupation choices are based on an underlying latent utility $U_{i(k|j)rt}$. The latent utility contains current period payoffs that are comprised of an observed component, $\hat{u}_{i(k|j)rt}$, and an error term, e_{ikrt}, and the expected discounted sum of future payoffs, conditional on optimal occupation choices in future periods:

$$\begin{aligned} U_{i(k|j)rt} &= \hat{u}_{i(k|j)rt} + e_{ikrt} + E_t[\sum_{z>t} \beta^{z-t} u_{i(k^*|j)rz}] \\ &= \tilde{U}_{i(k|j)rt} + e_{ikrt}, \end{aligned} \tag{2}$$

where $\tilde{U}_{i(k|j)rt} = \hat{u}_{i(k|j)rt} + E_t[\sum_{z>t} \beta^{z-t} u_{i(k^*|j)rz}]$ is the sub-utility function, and $u_{i(k^*|j)rz}$ is the maximal expected utility in period z, conditional on occupation choice k in t. Since workers accumulate occupation-specific experience, future utilities depend on current choices. The sub-utility captures the part of utility that depends on observables. The contemporaneous part $\hat{u}_{i(k|j)rt}$ is additive in observed parts of log wages and log preferences. The error term e_{ikrt} includes the error of log wages ϵ_{ikrt} and unobserved preferences. Note that workers have unobserved preferences across occupations, but not across training-occupation matches.

Individual i chooses occupation k to maximize the latent utility $U_{i(k|j)rt}$.[19] Using the above notation, i chooses k if and only if

$$(e_{ikrt} - e_{ik'rt}) \geq (\tilde{U}_{i(k'|j)rt} - \tilde{U}_{i(k|j)rt}), \quad \forall k' \neq k. \tag{3}$$

I define a corresponding occupation dummy variable:

$$occ_{i(k|j)rt} = \begin{cases} 1 & \text{if } U_{i(k|j)rt} \geq U_{i(k'|j)rt}, \quad \forall k' \neq k, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

4.3 Training Choice

In period t_0, training choices are based on a period-t_0 training utility, $\hat{u}_{ijr_0t_0} + e_{ijr_0t_0}$, and the expected future utility of choosing training j. Define the utility of choosing j as the sum of these two components:

$$V_{ijr_0t_0} = \hat{u}_{ijr_0t_0} + e_{ijr_0t_0} + E_{t_0}[\sum_{t>t_0} \beta^{t-t_0} u^*_{i(k^*|j)rt}], \tag{5}$$

$$= \tilde{V}_{ijr_0t_0} + e_{ijr_0t_0}, \tag{6}$$

[19] I abstract from regional choices as regional mobility is very low empirically (see Section 5.3).

where $\tilde{V}_{ijr_0t_0} = \hat{u}_{ijr_0t_0} + E_{t_0}[\sum_{t>t_0} \beta^{t-t_0} u^*_{i(k^*|j)rt}]$ is the sub-utility, $e_{ijr_0t_0}$ is an unobserved error, and $E_{t_0}[\sum_{t>t_0} \beta^{t-t_0} u^*_{i(k^*|j)rt}]$ is i's maximal expected future reward, conditional on training choice j in $t = t_0$. The maximal expected reward depends on the probability of choosing different occupations in the future. Since individuals may hold imperfect information, the maximal expected reward can differ from the discounted stream of realized utilities (see Appendix C for details on workers' information sets). Individual i chooses training j if and only if

$$(e_{ijr_0t_0} - e_{ij'r_0t_0}) \geq (\tilde{V}_{ij'r_0t_0} - \tilde{V}_{ijr_0t_0}), \quad \forall j' \neq j. \tag{7}$$

As before, I define a corresponding training dummy variable:

$$train_{ij} = \begin{cases} 1 & \text{if } V_{ijr_0t_0} \geq V_{ij'r_0t_0}, \quad \forall j' \neq j, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

where subscripts r_0 and t_0 for $train_{ij}$ are omitted for expositional clarity.

5 Identification and Estimation

This section explores potential biases in the returns to training-occupation matches. To identify ATEs in the given model with multiple unordered treatments, I then extend the control function approach proposed by Lee (1983) and Dahl (2002) to reduce the dimensionality of the selection problem, and combine it with an instrumental variables strategy. I describe this strategy and present identification conditions in Section 5.2. Section 5.3 defines the instruments, discusses the assumptions and presents a range of supporting evidence. Section 5.4 describes further estimation details.

5.1 Selection Biases

Log wages in occupation k with training j are only observed if individual i selects into jk. Based on the definition of the training and occupation dummies from Sections 4.2 and 4.3, the selection problem in outcome Equation (1) may be written as

$$E[\epsilon_{ikrt}|train_{ij} = 1, occ_{i(k|j)rt} = 1] \neq 0. \tag{9}$$

To illustrate the importance of selection at the training and occupation stages, suppose workers select based on (expected) wages only. At the training stage, they select to be trained in an occupation they have high expected relative ability in. This mechanism in itself may imply a *positive* bias in an estimate of on-diagonal returns. At the occupation stage, some workers may learn they have high ability in an occupation different from their training and select to work off the diagonal if their ability is sufficient to compensate for the lack of training. The latter may imply a *negative* bias in an estimate of on-diagonal returns. Note that these examples are for illustrative purpose only, and actual biases will empirically depend on the exact nature of worker learning and the correlations across occupation-specific error terms.

5.2 Identification with Multiple Unordered Treatments

The model in Section 4 is one of multiple unordered treatments, indicated by a set of dummy variables $train_{ij} \times occ_{i(k|j)rt}$. It also features essential heterogeneity as individuals select into treatment based on idiosyncratic returns. To identify the ATEs of different training-occupation combinations τ_{jk}, I extend the control function approach by Lee (1983) and Dahl (2002) to reduce the dimensionality in the given two-stage selection setting, and combine it with an instrumental variables strategy.

To understand this approach, define $M_{ijkrt} = train_{ij} \times occ_{i(k|j)rt}$. Lee (1983) points out that selection problems may be written in terms of maximum order statistics. In the given setting, this implies:

$$M_{ijkrt} = 1 \quad \text{iff} \quad \max_{j'}(V_{ij'r_0t_0} - V_{ijr_0t_0}) \leq 0, \quad \max_{k'}(U_{i(k'|j)rt} - U_{i(k|j)rt}) \leq 0. \tag{10}$$

Denote the joint distribution of the two maximum order statistics by $H_{jk}(\cdot)$. Using this distribution, Lee (1983) then argues that it is possible to create new standard normal random variables by transforming the maximum order statistics as follows:

$$\zeta_{ijkrt} = \Phi^{-1}\{H_{jk}(\max_{j'}(V_{ij'r_0t_0} - V_{ijr_0t_0}), \max_{k'}(U_{i(k'|j)rt} - U_{i(k|j)rt})|...)\}, \tag{11}$$

where the conditioning is on all sub-utility differences and $\Phi(\cdot)$ denotes the standard normal cumulative density function. The transformation can be chosen such that ζ_{ijkrt} are iid across jk. Since the function $\Phi^{-1}\{H_{jk}(\cdot)\}$ is increasing in both arguments, selection into jk implies the following conditions on the new random variables:

$$M_{ijkrt} = 1 \quad \Rightarrow \quad \zeta_{ijkrt} \leq \Phi^{-1}\{L_{jk}(\tilde{V}_{ijr_0t_0} - \tilde{V}_{i1r_0t_0}, ..., \tilde{V}_{ijr_0t_0} - \tilde{V}_{iJr_0t_0}, \tilde{U}_{i(k|j)rt} - \tilde{U}_{i(1|j)rt}, ..., \tilde{U}_{i(k|j)rt} - \tilde{U}_{i(K|j)rt})\}, \quad (12)$$

where $L_{jk}(\cdot)$ is the joint distribution of selection errors evaluated at the observed sub-utility differences which is equal to the probability of selection into jk. Denote this selection probability by $p_{ijkrt} = p_{ijr_0t_0} \times p_{i(k|j)rt}$, where $p_{ijr_0t_0}$ and $p_{i(k|j)rt}$ are the probabilities of selecting into training j and into occupation k conditional on j, respectively. Lee's (1983) identification approach involves a parametric assumption on the joint distribution of outcome errors and the newly created random variables. To understand how this may be combined with an instrumental variables strategy, let Z_j and Z_k denote the vectors of observed variables affecting the sub-utilities in Equations (2) and (5), i.e. $\tilde{V}_{ijr_0t_0} = \tilde{V}_{ijr_0t_0}(Z_j)$ and $\tilde{U}_{i(k|j)rt} = \tilde{U}_{i(k|j)rt}(Z_k)$, where I leave it as implicit that Z_j and Z_k vary across i, r and t. Denote the vector containing all unique elements of Z_j and Z_k by Z. Equivalently, denote the vector of all variables affecting potential outcomes in Equation (1) by X_k, and the vector containing all unique elements of X_k by X. Finally, denote by $X^{[-A]}$ and $Z^{[-A]}$ all elements of X and Z except the components in some vector of variables A.

I make the following identification assumptions:

A 1 *For each i, k, r, t,* $(e_{ikrt}, \epsilon_{ikrt})$ *is independent of* Z *conditional on* X_k*.*

A 2 *For each* jk*,* $j \in \mathbb{J}$*,* $k \in \mathbb{K}$*, there exists an instrument for* j*,* $I^j \in Z^{[-Z_j]}$*, and an instrument for* k*,* $I^k \in Z^{[-Z_k]}$*, such that* I^j *and* I^k *are* <u>*not*</u> *an element of* X_k*, and such that the distribution of* $p_{ijr_0t_0}$ *conditional on* $(X^{[-I^j]}, Z^{[-I^j]})$ *and the distribution of* $p_{i(k|j)rt}$ *conditional on* $(X^{[-I^k]}, Z^{[-I^k]})$ *are non-degenerate and continuous.*

A 3 *The distribution of* $(\{e_{ikrt}\}_{k\in\mathbb{K}})$ *is absolutely continuous with respect to Lebesgue measure on* $\Pi_{k\in\mathbb{K}}\mathbb{R}$*.*

A 4 *For each i, j, k, r, t,* $E(|ln(w_{ijkrt})|) < \infty$*.*

A 5 *For each i, j, k, r, t,* $P(M_{ijkrt} = 1|X) > 0$*.*

A 6 *For each i, j, k, r, t,* $(\epsilon_{ikrt}, \zeta_{ijkrt})$ *follow a bivariate normal distribution with variances equal to one and covariance between* ϵ_{ikrt} *and* ζ_{ijkrt} *equal to* ρ_{jk}*.*

Assumptions A1 and A3 imply that any two trainings and occupations have different sub-utilities with probability one. Assumption A4 is required for parameters τ_{jk} to be well-defined. Assumption A5 requires that at least some workers choose each training-occupation cell for all X. Imposing A6 on the joint distribution of outcome errors and Lee's transformed random variable implies that the conditional expectation $E[\epsilon_{ikrt}|\zeta_{ijkrt} \leq \Phi^{-1}(p_{ijkrt})]$ can be written as a well-known function of the inverse Mill's ratio (Heckman (1976, 1979)). In the given setting with many selection cells, this can be used to approximate the selection term $E[\epsilon_{ikrt}|M_{ijkrt} = 1]$.[20] Equation (1) of an individual i observed in training-occupation cell jk may then be written as

$$ln(w_{ijkrt}) = \delta_r + \delta_t + f(ln(vac_{krt})) + \delta_i + \tau_{j(i)k} + \beta' X_{ikt} - \rho_{jk} \frac{\phi[\Phi^{-1}(p_{ijkrt})]}{p_{ijkrt}} + u_{ikrt}, \quad (13)$$

where ρ_{jk} is the correlation between the outcome error ϵ_{ikrt} and the transformed variable ζ_{ijkrt}, u_{ikrt} is a mean-zero error, and $\phi(.)$ and $\Phi(.)$ denote the standard normal probability and cumulative density functions, respectively. Note that A2 is not required for Equation (13) to hold. Lee's Assumption A6 specifies the marginal distribution of outcome errors to be normal, but it does not restrict the correlation of the outcome errors across k. Given the marginal distributions of ϵ_{ikrt} and of ζ_{ijkrt}, A6 also specifies a flexible class for their joint distribution that preserves these marginals. This is equivalent to imposing a class of joint distributions for the outcome errors and the error term differences in the utility functions. But importantly, as noted by Dahl (2002), the marginal distribution of error term differences in the utility functions $L_{jk}(\cdot)$ does *not* dictate the form of the selectivity bias correction in Equation (13) since one can make a transformation consistent with assumed joint normality.

Implementation of Lee's approach requires specifying $L_{jk}(\cdot)$ to derive estimates of the selection probabilities. To avoid further parametric assumptions, I instead impose assumption A2 which implies that variation in the instruments allows for identification of $p_{ijr_0t_0}$ and $p_{i(k|j)rt}$ (and thereby p_{ijkrt}) separately from other variables in the outcome equation, without invoking parametric assumptions on the distribution of the selection errors.[21] This approach follows Dahl (2002) who also imposes exclusion restrictions.[22] The specific restrictions imposed in A2 rely on sector-specific regressors

[20] Note that this is an approximation since the condition on ζ_{ijkrt} in (12) is necessary but not sufficient for selection (see Appendix D for a discussion).

[21] Note that identification of τ_{jk} relies on both instruments and their interactions. The estimation approach takes this into account in the prediction of selection probabilities (see Section 5.4).

[22] A related literature uses instruments for *non-parametric* identification in settings with multiple

that generate variation in the utility of outside options (Heckman & Sedlacek (1985, 1990), D'Haultfœuille & Maurel (2013)), a strategy that is also common in the empirical industrial organization literature (Berry *et al.* (1995)). I define the instruments in Section 5.3 where I also discuss A1 and A2 for the given setting.

Implementation may proceed in a two-step control function procedure, by evaluating the inverse Mill's ratio at consistent probability estimates derived using the instruments, and including its interaction with selected jk-cells in Equation (13).[23]

Lee's parametric assumption A6 implies the specific expression for the selection term in Equation (13) and is necessary to identify parameters τ_{jk} separately from intercepts in the control functions. In contrast, Dahl (2002) shows that parameters on variables that exhibit variation within selection cells are identified non-parametrically, based on a weaker index sufficiency assumption (see Appendix D for details). This is the approach taken in a recent application by Ransom (2021) who considers heterogeneity in the returns to college majors accounting for selection into locations and occupations, treating education choices as exogenous. I make use of the differences in identification conditions and show that the parametric and non-parametric control function estimators give very similar results for parameters identified using either method (see Section 6.2.1). This lends support to the index sufficiency assumption, a necessary condition for A6. Note, however, that it is difficult to extrapolate this robustness to the ATEs τ_{jk} which are not identified non-parametrically.

5.3 Instrumental Variables Assumptions

This paper uses apprenticeship vacancies (henceforth *vacancies*) in outside options as instruments for the observed training and occupation choices to implement the control function approach described in Section 5.2. Instruments are needed for each of the possible training and occupation choice alternatives. I address this challenge by splitting vacancies into expectations and shocks, with the idea that expected vacancies affect the training choice, and shocks to these expectations affect the occupation

unordered treatments. Heckman *et al.* (2006, 2008) and Heckman & Vytlacil (2007) consider identification of LATEs/ATEs using large support conditions on instruments. Heckman & Pinto (2018) show that these conditions can be relaxed under unordered monotonicity. Kirkebøen *et al.* (2016) rely on additional data for identification, and estimate LATEs of education fields over next-best alternatives. Mountjoy (2022) uses instrument shifts that induce overlapping complier flows to identify marginal treatment effects of community colleges. Lee & Salanié (2018) consider identification with instruments in a general class of models where treatment is assigned based on multiple cutoffs.

[23] The selected cells include those where $j = k$, and cells for each occupation where $j \neq k$.

choice. Section 5.4 discusses how vacancies are split empirically.

For now, as before, let vac_{krt} denote vacancies in occupation k, and let vac_{jrt} denote vacancies in the occupation that j is training in. Define the expected log vacancies for occupation k at time t of an individual choosing a training in region r_0 at time t_0 as $E[ln(vac_{kt})|\Omega_{r_0t_0}]$, where $\Omega_{r_0t_0}$ summarizes the information set.[24] Log vacancies at time t are given by $ln(vac_{krt}) = E[ln(vac_{kt})|\Omega_{r_0t_0}] + (ln(vac_{krt}) - E[ln(vac_{kt})|\Omega_{r_0t_0}])$, where the second term is the shock to vacancies relative to the expectation formed in region r_0 at time t_0. Using the above, I define the set of instruments for a particular training and occupation choice jk as

$$I^j_{r_0t_0,j'(t_0+\tau)} = E[ln(vac_{j'(t_0+\tau)})|\Omega_{r_0t_0}] \ \forall j' \neq k, \ \forall \tau = 0, ..., 30, \tag{14}$$

$$I^k_{r_0t_0,k'rt} = (ln(vac_{k'rt}) - E[ln(vac_{k't})|\Omega_{r_0t_0}]) \ \forall k' \neq k. \tag{15}$$

The instruments for a training choice j are the predictions up to 30 years ahead of vacancies in occupations other than k. The instruments for occupation choice k at time t are the shocks to vacancies in occupations other than k, relative to the expectation in t_0.[25] Note that in cell jk where $j = k$, the training and occupation instruments sum to log vacancies in the outside options, $ln(vac_{k'rt})$. As discussed in Sections 5.2, identification relies on $vac_{k'rt}$ satisfying Assumptions A1-A6.

A1-A2 Conditional independence and exclusion. The key identifying assumption equivalent to A1 and A2 is that occupation-specific vacancies in occupation k' are excluded from the wage equation for occupation k and that, conditional on vacancies in occupation k, vacancies in k' are uncorrelated with unobserved components of wages in k. For example, conditional on vacancies for *craft workers*, vacancies for *electrical workers* should be excluded from and otherwise uncorrelated with wages for *craft workers*.

An economic model underlying this assumption is one in which posted occupation-specific vacancies at time t are a sufficient statistic for current occupation-specific labor demand, and changes to vacancies in other occupations are random conditional on own vacancies. Since occupation-specific labor supply is linked through worker self-selection, if vacancies reflected supply changes, these would not be random, potentially confounding the wage equation. Conditioning on own vacancies will reduce such

[24] Note that individuals do not form different expectations across regions.

[25] Given the 13 training/occupation categories, there are $12 \times 31 = 372$ instruments for each training choice and 12 instruments for each occupation choice.

confounders in the empirical analysis, but may not fully address them. Appendix C sets out a simple theoretical framework where variation in vacancies reflects changes to static labor demand. Empirically, vacancies correspond to posted apprenticeship vacancies, regardless of whether they become filled. Labor supply can therefore not affect the number of vacancies mechanically. Many vacancies do not get filled and filling rates vary over time, suggesting that firms do not base the number of posted vacancies on accurate expectations about the number of applicants. Since firms post most apprenticeship vacancies around the school leaving date, posting decisions are also unlikely to be affected by fluctuations in apprentice supply in the same year. To provide evidence that vacancies reflect contemporaneous demand changes (rather than past hiring), I show in a robustness check that the results from Equation (13) are quantitatively robust to the inclusion of occupation *times* time fixed effects.

Even if vacancies reflect occupation-specific demand changes, they may not be a sufficient statistic for these changes and confounding interactions may arise across occupations. To address time-varying occupation-specific confounders, I include occupation *times* time fixed effects as a robustness check. Another source for confounding correlations are industry-specific shocks. If vacancies were not a sufficient statistic for occupation-specific demand, these shocks could imply that within-industry vacancies for *electrical workers* may be correlated with wages for *craft workers*, even after conditioning on vacancies for *craft workers*. This would violate A1. To address this, I provide estimation results for Equation (13) including industry *times* time fixed effects. To address time-varying confounders at the industry-occupation level, I also provide a robustness check controlling for occupation *times* industry *times* time fixed effects. The fact that the results are quantitatively robust to the inclusion of these three sets of fixed effects suggests that vacancies are a sufficiently close proxy for occupation-specific demand in the given context.

Assumption A1 could also be violated if vacancy changes were non-random with respect to wages. To rule out strategic vacancy setting, firms need to be small relative to the market. This is true empirically where around three quarters of apprentices are trained in small and medium-sized firms. On the worker side, A1 rules out systematic relocation. Empirically, mobility is low. On average, over 93% of individuals start their training in their residence state.[26] Moreover, about 98% of all sample spells occur in the same region as the previous spell, 87% of individuals never more regions,

[26]Weighted average across states. Source: *BiBB, Datenreport zum Berufsbildungsbericht 2016.*

and only about $10 - 15\%$ of occupation changes are also geographical moves. I show that excluding movers from the sample does not affect the results.

Finally, general equilibrium effects through labor supply may violate Assumption A2. While it is difficult to rule out such feedback effects, it seems likely they would occur with a time lag. The fact that my results are quantitatively robust to including occupation *times* time fixed effects therefore suggests that feedback effects are not a major concern in the present setting.

A3 Relevance. To satisfy Assumption A3, the instruments need to be sufficiently strong drivers of the training and occupation choices. In the context of categorical endogenous variables, a natural way of assessing this is through the variation in selection probabilities generated by the instruments. I use a random forest algorithm to predict the selection into trainings and occupations and derive estimated selection probabilities (see Section 5.4 for details on the estimation of the probabilities). Figures B.5 and B.6 show histograms of the selection probabilities into the five largest training and occupation categories. It can be seen that there is considerable variability in the selection probabilities stemming from the instruments, indicating a substantial degree of first stage variation. At the same time, the estimated probabilities do not reach extreme values of 0.9 or above. This suggests that a fully non-parametric estimator will be infeasible in the given context, justifying the additional structure imposed in the estimation (see Section 5.2).

5.4 Estimating the Selection Probabilities

The selection probabilities $p_{ijr_0t_0}$ and $p_{i(k|j)rt}$ depend on the sub-utility differences $(\tilde{U}_{i(k'|j)rt} - \tilde{U}_{i(k|j)rt}), \forall k' \neq k$, $(\tilde{V}_{ij'krt} - \tilde{V}_{ijkrt}), \forall j' \neq j$, which in turn depend on the exogenous variables from Equation (1), the instruments defined in Section 5.3, as well as individual preferences. Due to data availability, I assume that preferences depend on the observables in the outcome equation.

To obtain the instruments, vacancies need to be split into expectations and shocks. To do so, I estimate separate linear time trend models in each region-time cell, where log vacancies for each occupation are explained using five years of previous data (see Appendix D for details).[27] Intuitively, individuals predict vacancies at the time and in the location of their training choice. Subsequent shocks are defined as the difference

[27]This implies that no instruments will be available for the first five sample years. Due to regional classification changes, data are also not available for four regions for 1994-1997.

between realized and predicted vacancies. I use an alternative AR(1) and a quadratic model to predict vacancies and show that the results are quantitatively robust to this (see Appendix D).

In a second step, I use the instruments together with the exogenous variables from the wage equation to predict training and occupation choices, and derive estimates for the selection probabilities.[28] To avoid functional form assumptions, I choose a flexible machine learning approach for this estimation, random forests. Besides avoiding functional form assumptions, random forests have the advantage of allowing for a large number of independent variables and are widely recognised for their accuracy. The algorithm predicts choices using optimal splitting rules on the explanatory variables until a final set of nodes is reached. Estimates for the choice probabilities are obtained as proportion of counts in the final nodes. Note that this method is similar to the non-parametric approach used by Dahl (2002), but instead of discretizing observables to create cells in which selection is assumed to be similar, the algorithm optimally splits the explanatory variables. Most related to my approach, Ransom (2021) uses conditional inference recursive partitioning to predict selection probabilities into occupations and locations in his setting.[29]

To avoid overfitting, I train the forest on a 50% sample of individuals, and use the remaining 50% in the regression analysis. Further details on the algorithm and its implementation can be found in Appendix D. The control function estimation proceeds by replacing the selection probabilities from Section 5.2 with their estimates $\hat{p}_{ijr_0t_0}$ and $\hat{p}_{i(k|j)rt}$ (Dahl (2002)).

6 Results

This section discusses the results for Equation (13), where I parameterize τ_{jk} to estimate on-diagonal returns within occupations (Sections 6.1, 6.2.1) and across occupations for each training (Sections 6.2.2, 6.2.3). Results are based on the baseline sample, excluding observations used to train the forest and years where the instruments are unavailable (see Section 5.4). To account for endogeneity of occupation-specific experience, I follow Altonji & Shakotko (1987) and use the deviation of exp_{ikt} from its individual mean as an instrument. The variable exp_{ikt} is then replaced by

[28] See Appendix D for further information on the explanatory variables in the prediction.

[29] This approach uses hypothesis testing to determine stopping criteria (see Ransom (2021)).

its first stage prediction (see Appendix D). To account for variability from generated regressors, standard errors are bootstrapped using the individual as sampling unit.[30] The function $f(ln(vac_{krt}))$ is approximated with a fourth order polynomial.

6.1 Average Return On versus Off the Diagonal

This section reports results where $\tau_{jk} = \delta_k + \tau D_{j=k}$, with δ_k denoting an occupation fixed effect and the variable of interest $D_{j=k}$ being equal to one if i works on the diagonal. Parameter τ captures the average on- versus off-diagonal return within occupations or, equivalently, the cost of lacking training in one's current occupation.[31] Table 3 shows the results. Columns (1) and (2) report results without controlling for occupation-specific experience, columns (3) and (4) condition on the first-stage prediction of exp_{ikt}.[32] Both models are estimated without selection control (columns (1), (3)), and using the control function estimator (columns (2), (4)).

Columns (1) and (3) show that $D_{j=k}$ is associated with a small wage effect. Using the control function estimator (columns (2) and (4)), the effect becomes positive and significant, implying a sizable negative selection bias of almost 14 percentage points. The coefficients on the control function are highly significant, confirming the importance of the bias. The negative bias leads to an underestimation of the cost of lacking training specifically in one's occupation and points towards selection into occupations in line with the mechanism highlighted in Section 5.1. Intuitively, on-diagonal workers may be negatively selected relative to off-diagonal ones as the latter compensate for the lack of training with higher occupation-specific ability.

Results from column (4) suggest an average cost of lacking training specifically in one's occupation of 14%. The effect is economically meaningful and equivalent to more than two years of occupation-specific work experience. In the robustness and heterogeneity analysis, I focus on the specification from columns (3) and (4) and control for occupation-specific experience throughout.

Table A.4 provides a range of robustness checks for the main result. To account for

[30]For computational reasons, I use 50 bootstrap replications. To confirm that the number of replications is sufficient, I run the main regression in Table 3 column (4) using 200 replications. This only changes standard errors by 11.6%.

[31]Note that the resulting wage formulation is equivalent to a model in which log skill prices per unit of human capital include an occupation-specific constant (see Appendix C). Further note that training fixed effects are absorbed by the individual fixed effects.

[32]Table A.3 reports results without instrumenting for exp_{ikt}.

Table 3: Average On- versus Off-Diagonal Returns

	(1)	(2)	(3)	(4)
$D_{j=k}=1$	-0.0144	0.1248	0.0027	0.1396
	(0.0038)	(0.0274)	(0.0039)	(0.0245)
exp	0.0619	0.0617	0.0527	0.0553
	(0.0006)	(0.0006)	(0.0020)	(0.0020)
exp^2	-0.0010	-0.0010	-0.0002	-0.0003
	(0.0000)	(0.0000)	(0.0001)	(0.0001)
exp_k			0.0117	0.0084
			(0.0020)	(0.0021)
exp_k^2			-0.0012	-0.0010
			(0.0001)	(0.0001)
Parametric cf	no	yes	no	yes
p-value cf		0.000		0.000
N	1,123,574	1,123,574	1,123,574	1,123,574

Notes: The table reports regression results for Equation (1) with $\tau_{jk} = \delta_k + \tau D_{j=k}$. All regressions include individual, occupation, region and time fixed effects. Bootstrap standard errors are reported in parentheses.

potentially non-random location moves which may violate the conditional independence assumption, column (1) excludes location movers; column (2) only considers workers with apprenticeships lasting two and a half to three years; column (3) excludes all individuals with potentially capped wages; column (4) excludes apprenticeship firm switchers; column (5) restricts the sample to years after German re-unification; column (6) controls for firm-specific experience and its square. In all columns, the results show significant on-diagonal returns that are similar to the main estimate of 14%, suggesting that certain sample sub-groups or the fact that on-diagonal workers may have more firm-specific human capital are not driving the main result.

Further robustness checks on the estimation can be found in Table A.5. Columns (1)-(4) show results using a higher-order polynomial in vacancies, additional predictors for the training probabilities, and two alternative models to predict vacancies. Columns (6)-(8) show estimates that only use variation within occupation-time, industry-time and occupation-industry-time cells to address potential identification concerns (see Section 5.3). All specifications lead to estimates that are very similar to the baseline estimate of 14%. Column (5) uses a more granular occupation classification. In line with the fact that this leads to on- and off-diagonal cells that are more similar in terms of their task distance, this leads to a smaller on-diagonal return.[33]

6.2 Heterogeneity

I explore the heterogeneity in the estimated average returns in a number of dimensions including work experience levels, trainings and occupations.

6.2.1 By Occupation-Specific Experience

This section discusses results where $\tau_{jk} = \delta_k + \tau^{exp_k} D_{j=k}$ to explore the heterogeneity in on-diagonal returns across full-time occupation-specific experience. Figure B.7 plots separate coefficient estimates for τ^{exp_k}, where experience levels have been binned into yearly categories. As before, an instrument is used for occupation-specific experience (see Appendix D). Each coefficient compares on-diagonal workers with a specific level of experience in their current occupation to workers with the same level of experience who were not trained in their occupation. As before, Figure B.7 shows that not controlling for selection implies sizable negative biases. The control function estimates show that the cost of lacking training in one's occupation initially increases and then falls from 14% to about 7% after 10 years of experience where it stabilizes. These results suggest that experience may compensate for lack of training and off-diagonal workers partly catch up with their on-diagonal co-workers. However, consistent with the relative importance of initial skills documented in Section 3.1, sizable differences remain after 20 years.

To provide suggestive evidence in support of the distributional assumptions required for the control function estimator, Figure B.8 plots the parametric estimates from Figure B.7, together with estimates from a non-parametric estimator based on the probability of the observed jk-cell (see Section 5.2). Since the latter only identifies slope coefficients, all coefficients are normalized to zero at zero years of experience. Reassuringly, the results show that the parametric and non-parametric estimates are almost identical.[34] Note, however, that it is difficult to extrapolate from the robustness of these estimates to the coefficients that are only parametrically identified.

[33] This result is suggestive as the control function only captures selection into baseline categories.

[34] The coefficient estimates remain very similar when also including on-diagonal probability estimates in the non-parametric estimator as in Dahl (2002).

6.2.2 By Training

This section discusses results where $\tau_{jk} = \tau_j D_{j=k}$ to explore the on-diagonal returns for each training. This model does not contain occupation fixed effects to estimate the parameters relevant for occupational selection conditional on a training. Note that the parameters τ_j therefore reflect both training-occupation match effects and average differences in opportunities across occupations. Figure B.9 plots coefficients τ_j estimated with and without selection control. The estimates are highly heterogeneous, pointing to differences in labor market or promotional opportunities across occupations, or differences in the portability of training skills to other occupations.

Regardless of the causes of heterogeneity, if workers choose occupations based on the return to working in their training occupation, the heterogeneity in these returns should correlate with the fraction of on-diagonal workers. The Roy model predicts that more workers work on the diagonal, the higher the on-diagonal return in that training. Figure B.10 explores this relationship by plotting the returns from Figure B.9 together with the on-diagonal fraction for each training. The positive slope is consistent with the model as outlined above, and suggests that relative returns are an important determinant of occupational selection.

6.2.3 Full Training-Occupation Matrix

To further explore the heterogeneity in returns across occupations, this section reports results for the outcome equation with parameters τ_{jk} for all cells in the training-occupation matrix. Table 4 shows results using the parametric control function estimator for the five largest occupations. Tables A.6 and A.7 contain the full set of coefficients. The inclusion of individual fixed effects implies that all coefficients are relative to the diagonal in the training row. Table 4 shows that most coefficients are negative suggesting that workers incur off-diagonal penalties.

There is considerable heterogeneity in the off-diagonal returns across trainings. This variation presents an opportunity to provide direct evidence that the specificity of training skills is a driver of the estimated returns. In Section 7, I use data on the task content of occupations to derive measures of task distance between trainings and occupations, and use these to explore the heterogeneity in estimated returns.

Table 4: Full Matrix of Returns - Within-Training Comparisons

		Occupation				
		Office workers	Craft workers	Sales, fin. workers	Health workers	Constr. workers
Training	Office w.	0	−0.13	0.00	−0.27	−0.26
			(0.10)	(0.06)	(0.10)	(0.09)
	Craft w.	0.08	0	0.44	0.16	0.27
		(0.04)		(0.06)	(0.09)	(0.08)
	Sales, fin. w.	−0.13	0.20	0	−0.03	0.06
		(0.07)	(0.10)		(0.10)	(0.10)
	Health, soc. w.	−0.80	−0.51	−0.35	0	−0.64
		(0.07)	(0.13)	(0.08)		(0.14)
	Constr. w.	−0.26	0.07	0.17	−0.13	0
		(0.06)	(0.08)	(0.09)	(0.11)	

Notes: The table shows estimates for τ_{jk} in Equation (1), estimated with the parametric control function estimator. Bootstrap standard errors are reported in parentheses.

7 Task Content

I draw on the task approach to occupations to provide evidence that the estimated returns to training-occupation matches can be rationalized by the specificity of skills. The task approach considers tasks as production inputs, and skills as the human capital required to carry out the tasks (e.g., Autor (2013)). Occupations, as discrete units, correspond to vectors of tasks that are carried out by workers. Based on this concept, it is possible to construct task distance measures between occupations. Poletaev & Robinson (2008) and Gathmann & Schönberg (2010) argue that, if human capital is task-specific, it should be more easily transferable across occupations that require similar tasks.[35] In the present context, these findings suggest an intuitive explanation for the heterogeneity in returns in the training-occupation matrix that is based on skill specificity. If workers are trained in a specific mix of tasks, one would expect the penalty in a different occupation to be larger, the more distant the occupation is from the original training.

[35] Using samples of displaced workers, they find that wage penalties are larger the more distant the occupational switch is after displacement. Yamaguchi (2012) sets up a structural model to formalize these findings. Cortes & Gallipoli (2018) estimate a structural model and show that task difference is a significant component of the cost of switching occupations.

7.1 Measuring Task Distance

The measure of task distance is constructed using the German Qualification and Career Survey, a representative survey of around 20.000 individuals conducted by the Federal Institute for Vocational Training and Education (BiBB). This data has been used to study skill requirements across occupations in different contexts (e.g., Spitz-Oener (2006), Gathmann & Schönberg (2010)). I use four survey waves that fall into the sampling period.[36] The survey records information on workers' occupations and asks them to pick from a list of tasks the ones that they perform at work. A summary table of the tasks together with the share of individuals performing these tasks is presented in Table A.8. Following Gathmann & Schönberg (2010), I use the task data to construct a measure of distance between training j and occupation k.

Define a task vector for each occupation k, $q_k = (q_{1k}, ..., q_{Sk})$, where q_{sk} is the fraction of workers performing task s in occupation k. Similarly, define a task vector for each training j, $q_j = (q_{1j}, ..., q_{Sj})$, where q_{sj} is the fraction of workers performing task s when being trained in training j, which is assumed to be equivalent to the fraction when working in $k = j$. The angular separation between training j and occupation k is defined as a measure of similarity using task vectors q_j and q_k:

$$AngSim_{jk} = \frac{\sum_{s=1}^{S} (q_{sj} \times q_{sk})}{[(\sum_{s=1}^{S} q_{sj}^2) \times (\sum_{s=1}^{S} q_{sk}^2)]^{1/2}}. \tag{16}$$

$AngSim_{jk}$ ranges from zero to one, and is increasing in the overlap between task vectors q_j and q_k. Define the *distance* between training j and occupation k as

$$Dist_{jk} = (1 - AngSim_{jk}). \tag{17}$$

The distance measure is decreasing in the overlap between the task vectors with two orthogonal task vectors having distance one. Excluding on-diagonal cells where $Dist_{jk} = 0$, the distance measure varies between 0.02 and 0.59, with a mean of 0.35. When weighting training-occupation cells by their sample fractions, the mean distance drops to 0.28, indicating a negative correlation between training-occupation distance and the fraction of workers in the relevant cell. Tables A.9 and A.10 report distance measures for the five most similar and most distant training-occupation pairs, as well

[36] *BiBB/IAB Erhebungen 1985/86, 1991/92, 1998/99; BiBB/BAuA Erhebung 2005/06.*

as for the five largest trainings and occupations.

7.2 Match Returns and Task Distance

I model the estimated returns to a training-occupation combination, $\hat{\tau}_{jk}$, from Section 6.2.3 using the following simple specification:

$$\hat{\tau}_{jk} = \alpha + \beta Dist_{jk} + \eta_{jk}, \tag{18}$$

where the distance measure $Dist_{jk}$ between training j and occupation k is standardized, and η_{jk} is a match-specific error term.

Table 5 presents results for Equation (18). Column (1) shows that higher task distance is related to lower returns in training-occupation cells. Specifically, it suggests that a one-standard-deviation higher task distance is associated with a fall in $\hat{\tau}_{jk}$ of around 7 percentage points (*pp*).[37] To account for heterogeneity in opportunities across occupations, column (2) includes occupation fixed effects. This slightly reduces the coefficient on $Dist_{jk}$ to around 4*pp* or around 76% of the average $\hat{\tau}_{jk}$. Columns (3) and (4) present equivalent results where the returns τ_{jk} have been estimated without selection control, showing that the effect of $Dist_{jk}$ is smaller and no longer significant.

Table A.11 shows that the findings are robust to excluding on-diagonal observations where $Dist_{jk} = 0$, and to restricting the sample to the five largest occupations. Using a quadratic specification, I find suggestive evidence for decreasing penalties to task distance (see Table A.11). Overall, the results are in line with the hypothesis that apprentices are trained to carry out a mix of tasks and their returns in an occupation are lower, the less applicable their skills are to that occupation.

8 Welfare and Policy

The results from Section 6 suggest that lacking training in one's occupation can be costly. Using the model from Section 4, this section explores the partial equilibrium welfare loss from ex-post suboptimal training choices due to imperfect information. It then considers retraining as potential policy intervention. I focus on off-diagonal

[37]Note that the average $\hat{\tau}_{jk}$ differs from the baseline estimate of $\hat{\tau}$ in Section 6.1 since it is derived without controlling for occupation fixed effects (see Section 6.2.3).

Table 5: Match Returns and Task Distance

τ_{jk} estimated	with parametric control fcn.		without selection control	
	(1)	(2)	(3)	(4)
$Dist_{jk}$	−0.0688	−0.0395	−0.0184	0.0052
	(0.0282)	(0.0253)	(0.0125)	(0.0097)
Occ. FE	no	yes	no	yes
Mean of $\hat{\tau}_{jk}$	−0.0520	−0.0520	−0.0239	−0.0239
R-squared	0.0364	0.2977	0.0160	0.4523
N	169	169	169	169

Notes: The table reports regression results for Equation (18). Observations are weighted by the sample fraction in the relevant training. Robust standard errors are reported in parentheses.

workers, moving a discussion of workers locked into their training to Appendix E. As a result, the below calculations constitute a lower bound on total losses.

8.1 Welfare Losses

Consider the model from Section 4 with homogenous on- versus off-diagonal returns as in the baseline specification in Section 6.1. In this model, occupation choices of off-diagonal workers are first best by revealed preference. As a result, the optimal cell in the training-occupation matrix at any point in time is the on-diagonal cell in the current occupation. Using Equations (1) and (2) in Section 4, the welfare loss relative to this cell corresponds to the within-occupation on- versus off-diagonal return, τ. Note that there are no other losses since the error e_{ikrt} does not vary across trainings (see Section 4.2). Further note that, with heterogeneity in on-diagonal returns or search frictions, τ is a lower bound on losses relative to the optimal cell as observed occupations may not be first best.

Due to the inclusion of individual fixed effects in the wage equation, τ does not capture average return differences across trainings. Empirically, this implies a further reason why τ is a lower bound on losses since off-diagonal workers disproportionately work in occupations that have high-average-return corresponding trainings.[38]

Column (4) in Table 3 suggests that τ is around 14%. This estimate is a meaningful average as it takes into account the distribution of workers across occupations and

[38] I proxy for returns using the fraction of apprentices with advance schooling (Abitur).

levels of experience. On average, the likelihood of working off the diagonal at any point in time is 40% (see Section 3.1). Taken together, this implies that the per-period welfare loss relative to the optimal allocation due to off-diagonal work amounts to 5.6% per worker in the apprenticeship system.

In a two-period model where workers train in the first period and work in the second, this corresponds to the loss from imperfect information at the time of training choice. With T periods, the loss from imperfect information may be smaller as workers can hold multiple occupations, but only one training. Empirically, however, 50% of workers who change occupations only do so once and over 70% never move back to their training occupation. Excluding workers with multiple occupation changes implies a 2.6% conservative lower bound on losses from lack of information.[39]

8.2 Retraining Programs

Section 8.1 suggests that providing applicants with information for instance through pre-training internships could lead to sizable welfare gains, but it is hard to quantify these without knowledge of how much information may be transmitted. An alternative intervention are ex-post retraining programs. I briefly consider these here, moving any details to Appendix E. Since workers are trained in occupation-specific subjects for two thirds of their training, I assume that retraining would last two years.

To estimate the costs from retraining, I consider schooling costs, training costs in firms and foregone earnings. In 2010, these costs amounted to 32,400 Euros. The annual benefit from retraining for off-diagonal workers τ, net of foregone experience, corresponds to around $2,600$ Euros in 2010. Taking into account that costs need to be paid upfront while benefits accrue for every subsequent year spent working on the diagonal, and using a discount factor of 0.98, my calculations suggest that retraining costs would be recovered for workers moving off the diagonal with at most seven years of work experience, or over 88% of workers who only switch occupations once.

These results suggest that retraining programs could be effective in addressing the uncertainty workers face at the time of training choice. Yet, only very few individuals retrain in practice.[40] My findings suggest that retraining opportunities should be

[39] Figure is calculated by multiplying the average likelihood of working off the diagonal and not switching occupations more than once by τ.

[40] The share of those with apprenticeship training who have training spells in more than one occupation is 13%. This is likely an overestimate of retraining as it counts missing training entries as different occupations and includes non-completed apprenticeships.

expanded, and that future research should consider potential barriers to retraining such as liquidity constraints.

9 Conclusion

This paper uses administrative panel data to study training specificity in the German dual apprenticeship system. The dual system is widely praised for facilitating school-to-work transitions through the provision of occupation-specific marketable skills. A potential concern is that highly specific skills, while enabling strong labor market attachment, entail important costs if workers move fields.

Based on a relatively broad occupational classification, I show that 40% of trained individuals work in occupations they did not receive training in. Assessing whether this mobility poses a concern requires credible estimates of the returns to training-occupation matches. To address selection into trainings and occupations, I extend existing control function approaches in high-dimensional selection settings and use plausibly exogenous variation. I find sizable average returns to matching one's training to the corresponding occupation. Using heterogeneity in this return, I provide evidence consistent with apprentices being trained in a mix of tasks and receiving larger wage penalties the less applicable their skills are to their current occupation.

My findings show that, when skills are specific, imperfect information at the time of training choice can lead to important welfare losses. Further work in other settings is required to understand the full set of trade-offs inherent in the provision of more or less specific skills. Systems delivering more general skills may entail lower penalties for individuals working outside their field, but more individuals could end up moving occupations. And not only may general education systems be less cost effective for some workers, but they could also lead to weaker labor market attachment and higher youth unemployment. Ultimately, a more complete picture of the trade-offs will allow for progress on the design of optimal education systems under uncertainty.

References

Acemoglu, Daron, & Pischke, Jörn-Steffen. 1998. Why Do Firms Train? Theory and Evidence. *The Quarterly Journal of Economics*, **113**(1), 79–119.

Acemoglu, Daron, & Restrepo, Pascual. 2019. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, **33**(2), 3–30.

Acemoglu, Daron, & Restrepo, Pascual. 2020. Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, **128**(6), 2019–2452.

Alfonsi, Livia, Bandiera, Oriana, Bassi, Vittorio, Burgess, Robin, Rasul, Imran, Sulaiman, Munshi, & Vitali, Anna. 2020. Tackling Youth Unemployment: Evidence from a Labor Market Experiment in Uganda. *Econometrica*, **88**(6), 2369–2414.

Altonji, Joseph G., & Shakotko, Robert A. 1987. Do Wages Rise with Job Seniority? *The Review of Economic Studies*, **54**(3), 437–459.

Antonovics, Kate, & Golan, Limor. 2012. Experimentation and Job Choice. *Journal of Labor Economics*, **30**(2), 333–366.

Arcidiacono, Peter, Hotz, V. Joseph, Maurel, Arnaud, & Romano, Teresa. 2020. Ex Ante Returns and Occupational Choice. *Journal of Political Economy*, **128**(12), 4475–4522.

Autor, David H. 2013. The "Task Approach" to Labor Markets: An Overview. *Journal for Labour Market Research*, **46**(3), 185–199.

Becker, Gary S. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, **70**(5), 9–49.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* University of Chicago Press, Chicago.

Bennell, Paul. 1996. General Versus Vocational Secondary Education in Developing Countries: A Review of the Rates of Return Evidence. *The Journal of Development Studies*, **33**(2), 230–247.

Berry, Steven, Levinsohn, James, & Pakes, Ariel. 1995. Automobile Prices in Market Equilibrium. *Econometrica*, **63**(4), 841–890.

Breiman, Leo. 2001. Random Forests. *Machine Learning*, **45**, 5–32.

Caicedo, Santiago, Espinosa, Miguel, & Seibold, Arthur. 2022. Unwilling to Train? - Firm Responses to the Colombian Apprenticeship Regulation. *Econometrica*, **90**(2), 507–550.

Clark, Damon, & Fahr, René. 2001. The Promise of Workplace Training for Non-College-Bound Youth: Theory and Evidence from German Apprenticeship. *IZA Discussion Paper, No. 378.*

Cortes, Guido Matias, & Gallipoli, Giovanni. 2018. The Costs of Occupational Mobility: An Aggregate Analysis. *Journal of the European Economic Association*, **16**(2), 275–315.

Dahl, Gordon B. 2002. Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, **70**(6), 2367–2420.

D'Haultfœuille, Xavier, & Maurel, Arnaud. 2013. Inference on an Extended Roy Model, with an Application to Schooling Decisions in France. *Journal of Econometrics*, **174**(2), 95–106.

Dustmann, Christian, Ludsteck, Johannes, & Schönberg, Uta. 2009. Revisiting the German Wage Structure. *The Quarterly Journal of Economics*, **124**(2), 843–881.

Franz, Wolfgang, & Soskice, David. 1995. The German Apprenticeship System. *Chap. 8, pages 208–234 of:* Buttler, Friedrich, Franz, Wolfgang, & Soskice, David (eds), *Institutional Frameworks and Labor Market Performance.* Routledge, London.

Fredland, John Eric, & Little, Roger D. 1980. Long-Term Returns to Vocational Training: Evidence from Military Sources. *The Journal of Human Resources*, **15**(1), 49–66.

Gathmann, Christina, & Schönberg, Uta. 2010. How General is Human Capital? A Task-Based Approach. *Journal of Labor Economics*, **28**(1), 1–49.

Göggel, Kathrin, & Zwick, Thomas. 2012. Heterogeneous Wage Effects of Apprenticeship Training. *The Scandinavian Journal of Economics*, **114**(3), 756–779.

Goldin, Claudia. 2001. The Human-Capital Century and American Leadership: Virtues of the Past. *The Journal of Economic History*, **61**(2), 263–292.

Griliches, Zvi. 1977. Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, **45**(1), 1–22.

Guvenen, Fatih, Kuruscu, Burhan, Tanaka, Satoshi, & Wiczer, David. 2020. Multidimensional Skill Mismatch. *American Economic Journal: Macroeconomics*, **12**(1), 210–244.

Hanushek, Eric A., Schwerdt, Guido, Woessmann, Ludger, & Zhang, Lei. 2017. General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle. *Journal of Human Resources*, **52**(1), 48–87.

Hastie, Trevor, Tibshirani, Robert, & Friedman, Jerome. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd edn. Springer, New York.

Hastings, Justine S., Neilson, Christopher A., Ramirez, Anely, & Zimmerman, Seth D. 2016. (Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data. *Economics of Education Review*, **51**(1), 136–151.

Heckman, James J. 1976. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Pages 475–492 of:* Berg, Sanford V. (ed), *Annals of Economic and Social Measurement*, vol. 5, no. 4. National Bureau of Economic Research.

Heckman, James J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, **47**(1), 153–161.

Heckman, James J., & Pinto, Rodrigo. 2018. Unordered Monotonicity. *Econometrica*, **86**(1), 1–35.

Heckman, James J., & Sedlacek, Guilherme. 1985. Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self- Selection in the Labor Market. *Journal of Political Economy*, **93**(6), 1077–1125.

Heckman, James J., & Sedlacek, Guilherme L. 1990. Self-Selection and the Distribution of Hourly Wages. *Journal of Labor Economics*, **8**(1), S329–S363.

Heckman, James J., & Vytlacil, Edward J. 2007. Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. *Chap. 71, pages 4875–5143 of:* Heckman, James J., & Leamer, Edward E. (eds), *Handbook of Econometrics*, vol. 6, part B. North Holland.

Heckman, James J., Urzua, Sergio, & Vytlacil, Edward. 2006. Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, **88**(3), 389–432.

Heckman, James J., Urzua, Sergio, & Vytlacil, Edward. 2008. Instrumental Variables in Models with Multiple Outcomes: the General Unordered Case. *Annales d'Economie et de Statistique*, **91/92**, 151–174.

Kambourov, Gueorgui, & Manovskii, Iourii. 2009. Occupational Specificity of Human Capital. *International Economic Review*, **50**(1), 63–115.

Keane, Michael P., & Wolpin, Kenneth I. 1997. The Career Decisions of Young Men. *Journal of Political Economy*, **105**(3), 473–522.

Kinsler, Josh, & Pavan, Ronni. 2015. The Specificity of General Human Capital: Evidence from College Major Choice. *Journal of Labor Economics*, **33**(4), 933–972.

Kirkebøen, Lars J., Leuven, Edwin, & Mogstad, Magne. 2016. Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics*, **131**(3), 1057–1111.

Krueger, Alan B., & Pischke, Jörn-Steffen. 1995. A Comparative Analysis of East and West German Labor Markets: Before and After Unification. *Pages 405–446 of:* Freeman, Richard B., & Katz, Lawrence F. (eds), *Differences and Changes in Wage Structures*. University of Chicago Press, Chicago.

Krueger, Dirk, & Kumar, Krishna B. 2004a. Skill-Specific Rather than General Education: A Reason for US-Europe Growth Differences? *Journal of Economic Growth*, **9**(2), 167–207.

Krueger, Dirk, & Kumar, Krishna B. 2004b. US-Europe Differences in Technology-Driven Growth: Quantifying the Role of Education. *Journal of Monetary Economics*, **51**(1), 161–190.

Lazear, Edward P. 2009. Firm-Specific Human Capital: A Skill-Weights Approach. *Journal of Political Economy*, **117**(5), 914–940.

Lee, Lung-Fei. 1983. Generalized Econometric Models with Selectivity. *Econometrica*, **51**(2), 507–512.

Lee, Sokbae, & Salanié, Bernard. 2018. Identifying Effects of Multivalued Treatments. *Econometrica*, **86**(6), 1939–1963.

Malamud, Ofer, & Pop-Eleches, Cristian. 2010. General Education Versus Vocational Training: Evidence from an Economy in Transition. *The Review of Economics and Statistics*, **92**(1), 43–60.

Miller, Robert A. 1984. Job Matching and Occupational Choice. *Journal of Political Economy*, **92**(6), 1086–1120.

Montt, Guillermo. 2015. The Causes and Consequences of Field-of-Study Mismatch: An Analysis using PIAAC. *OECD Social, Employment and Migration Working Papers, No.167.*

Mountjoy, Jack. 2022. Community Colleges and Upward Mobility. *American Economic Review*, **112**(8), 2580–2630.

Neal, Derek. 1995. Industry-Specific Human Capital: Evidence from Displaced Workers. *Journal of Labor Economics*, **13**(4), 653–677.

Neuman, Shoshana, & Ziderman, Adrian. 1991. Vocational Schooling, Occupational Matching, and Labor Market Earnings in Israel. *The Journal of Human Resources*, **26**(2), 256–281.

Nicholson, Sean. 2002. Physician Specialty Choice under Uncertainty. *Journal of Labor Economics*, **20**(4), 816–847.

Nordin, Martin, Persson, Inga, & Rooth, Dan-Olof. 2010. Education-Occupation Mismatch: Is there an Income Penalty? *Economics of Education Review*, **29**(6), 1047–1059.

Oreopoulos, Philip, & Dunn, Ryan. 2013. Information and College Access: Evidence from a Randomized Field Experiment. *The Scandinavian Journal of Economics*, **115**(1), 3–26.

Parent, Daniel. 2000. Industry-Specific Capital and the Wage Profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics. *Journal of Labor Economics*, **18**(2), 306–323.

Peter, Frauke H., & Zambre, Vaishali. 2017. Intended College Enrollment and Educational Inequality: Do Students Lack Information? *Economics of Education Review*, **60**, 125–141.

Poletaev, Maxim, & Robinson, Chris. 2008. Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984-2000. *Journal of Labor Economics*, **26**(3), 387–420.

Quintini, Glenda, Martin, John P., & Martin, Sébastien. 2007. The Changing Nature of the School-to-Work Transition Process in OECD Countries. *IZA Discussion Paper, No. 2582.*

Ransom, Tyler. 2021. Selective Migration, Occupational Choice, and the Wage Returns to College Majors. *Annals of Economics and Statistics*, **142**, 45–110.

Robst, John. 2007. Education and Job Match: The Relatedness of College Major and Work. *Economics of Education Review*, **26**(4), 397–407.

Roy, Andrew D. 1951. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, **3**(2), 135–146.

Saniter, Nils, Schnitzlein, Daniel D., & Siedler, Thomas. 2019. Occupational Knowledge and Educational Mobility: Evidence from the Introduction of Job Information Centers. *Economics of Education Review*, **69**, 108–124.

Shaw, Kathryn L. 1987. Occupational Change, Employer Change, and the Transferability of Skills. *Southern Economic Journal*, **53**(3), 702–719.

Siow, Aloysius. 1984. Occupational Choice under Uncertainty. *Econometrica*, **52**(3), 631–645.

Spitz-Oener, Alexandra. 2006. Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, **24**(2), 235–270.

Stinebrickner, Ralph, & Stinebrickner, Todd R. 2014. A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *The Review of Economic Studies*, **81**(1), 426–472.

Sullivan, Paul. 2010. A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search. *International Economic Review*, **51**(1), 289–317.

Todd, Petra E., & Zhang, Weilong. 2020. A Dynamic Model of Personality, Schooling, and Occupational Choice. *Quantitative Economics*, **11**(1), 231–275.

Wiswall, Matthew, & Zafar, Basit. 2015. Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies*, **82**(2), 791–824.

Wolter, Stefan C., & Ryan, Paul. 2011. Apprenticeship. *Chap. 11, pages 521–576 of:* Hanushek, Eric A., Machin, Stephen, & Woessmann, Ludger (eds), *Handbook of the Economics of Education*, vol. 3. North Holland.

Yamaguchi, Shintaro. 2012. Tasks and Heterogeneous Human Capital. *Journal of Labor Economics*, **30**(1), 1–53.

Zafar, Basit. 2011. How Do College Students Form Expectations? *Journal of Labor Economics*, **29**(2), 301–348.

For Online Publication

Appendix to:

Training Specificity and Occupational Mobility:
Evidence from German Apprenticeships

Dita Eckardt

Department of Economics, University of Warwick

Appendix A. Tables

Table A.1: List of Occupations

KldB88 code	Occupation label	Sub-label	% in code
75-78	Office workers	Office workers	73.1
		Other	26.9
19-30, 32	Craft workers	Vehicle mechanics	14.4
		Machine fitters	10.7
		Plumbers	10.7
		Other	64.1
68-70	Sales, financial workers	Salespeople	34.3
		Banking experts	24.3
		Wholesalers, retail dealers	16.6
		Other	24.8
79-89	Health, social workers	Medical receptionists	25.9
		Nurses, midwives	23.0
		Nursery, childcare w.	10.2
		Other	40.9
44-51	Construction workers	Bricklayers, concrete w.	21.9
		Carpenters	21.2
		Decorators, painters	15.7
		Other	41.2
10-18, 52-54	Process, plant workers	Chemical, plastics proc. w.	26.4
		Unskilled laborers	19.2
		Other	54.5
71-74	Transport, stor. workers	Vehicle drivers	39.7
		Movers, warehousers	22.0
		Stock clerks	17.9
		Other	20.4

Table A.1 continued: List of Occupations

KldB88 code	Occupation label	Sub-label	% in code
60-63	Technical, lab. workers	Other technicians	22.6
		Technical drawers	17.0
		Electrical technicians	16.0
		Other	44.4
31	Electrical workers	Electricians	69.5
		Telephone technicians	17.2
		Electr. appliance fitters	13.3
		Other	0
90-93	Personal serv. workers	Hairdr., body care occ.	40.8
		Hospitality workers	28.4
		Other	30.8
39-43	Food prep. workers	Cooks, ready meal prod.	39.0
		Bakers, confectioners	28.5
		Butchers, fish processing w.	21.7
		Coopers, brewers, food prod.	10.8
		Other	0
01-09	Agricultural workers	Gardeners, florists, forest.	57.9
		Miners, oil production w.	22.9
		Farmers, zookeepers	19.2
		Other	0
33-37	Textile, garm. workers	Tailors, textile ind. w.	59.6
		Spinners, shoem.	40.4
		Other	0

Notes: The table lists all occupations contained in the baseline sample by fraction in the sample. Sub-labels are provided for all within-code shares greater than 10%.

Table A.2: Spells as Percentage of Trainings/Occupations

		Occupation												
Training		01-09	10-54	19-32	31	33-37	39-43	44-51	60-63	68-70	71-74	75-78	79-89	90-93
01-09	Agricultural	51.8	6.3	4.9	0.8	0.2	0.4	4.8	3.3	5.3	9.6	6.5	4.7	1.4
		70.8	*2.5*	*0.8*	*0.5*	*0.9*	*0.4*	*1.5*	*1.2*	*0.9*	*3.9*	*0.7*	*1.2*	*1.5*
10-54	Process, plant	0.7	57.3	4.2	0.6	0.1	0.2	1.5	12.0	4.7	5.7	8.7	3.6	0.6
		0.8	*19.2*	*0.6*	*0.3*	*0.3*	*0.2*	*0.4*	*3.7*	*0.7*	*1.9*	*0.8*	*0.8*	*0.6*
19-32	Craft	0.9	9.5	55.3	1.6	0.2	0.4	2.5	8.9	3.9	9.2	4.8	2.4	0.6
		11.3	*36.9*	*84.3*	*8.9*	*8.8*	*3.6*	*7.5*	*32.3*	*6.2*	*36.2*	*5.0*	*5.7*	*6.7*
31	Electrical	0.6	5.3	8.7	47.0	0.1	0.2	1.2	17.1	3.8	4.7	8.1	2.8	0.5
		2.6	*6.4*	*4.1*	*84.1*	*0.9*	*0.7*	*1.1*	*19.2*	*1.9*	*5.7*	*2.6*	*2.1*	*1.6*
33-37	Textile, garment	0.4	9.7	8.0	0.4	35.6	1.5	3.6	7.0	8.3	5.6	12.7	4.7	2.4
		0.2	*1.5*	*0.5*	*0.1*	*70.1*	*0.6*	*0.4*	*1.0*	*0.5*	*0.9*	*0.5*	*0.5*	*1.1*
39-43	Food preparation	1.1	8.5	6.2	0.7	0.3	43.1	3.6	1.7	7.8	13.2	6.7	3.6	3.4
		2.6	*6.1*	*1.7*	*0.7*	*2.8*	*82.8*	*2.0*	*1.1*	*2.3*	*9.6*	*1.3*	*1.6*	*6.6*
44-51	Construction	1.1	7.5	5.7	0.5	0.3	0.4	60.2	4.3	3.1	9.4	3.5	2.9	0.9
		7.0	*13.5*	*4.0*	*1.2*	*6.3*	*2.1*	*85.1*	*7.2*	*2.3*	*17.0*	*1.7*	*3.2*	*4.3*
60-63	Technical, lab.	0.3	2.7	2.5	3.2	0.0	0.1	0.7	68.7	4.1	1.6	12.8	2.8	0.5
		0.4	*1.2*	*0.4*	*2.1*	*0.1*	*0.1*	*0.2*	*27.6*	*0.7*	*0.7*	*1.5*	*0.7*	*0.6*
68-70	Sales, financial	0.2	2.3	1.6	0.2	0.2	0.6	0.3	0.8	60.6	3.4	26.5	2.1	1.1
		1.7	*5.9*	*1.7*	*0.9*	*6.1*	*4.3*	*0.6*	*1.8*	*64.8*	*9.0*	*18.2*	*3.4*	*8.0*
71-74	Transport, storage	0.1	5.3	3.5	0.9	0.0	0.3	2.5	2.1	7.6	55.2	18.9	2.9	0.7
		0.0	*0.6*	*0.1*	*0.1*	*0.0*	*0.1*	*0.2*	*0.2*	*0.3*	*5.9*	*0.5*	*0.2*	*0.2*
75-78	Office	0.1	0.8	0.6	0.1	0.0	0.0	0.1	1.1	12.5	2.0	80.6	1.6	0.4
		1.0	*2.3*	*0.7*	*0.3*	*1.0*	*0.3*	*0.3*	*2.8*	*14.4*	*5.5*	*59.4*	*2.8*	*3.1*
79-89	Health, social	0.1	0.8	0.7	0.1	0.0	0.2	0.2	0.8	4.3	1.0	12.2	79.0	0.7
		0.8	*1.3*	*0.4*	*0.3*	*0.6*	*0.7*	*0.3*	*1.2*	*2.7*	*1.5*	*5.1*	*76.4*	*2.8*
90-93	Personal service	0.4	5.2	3.6	0.5	0.3	3.0	0.6	1.0	10.6	3.8	20.8	4.9	45.2
		0.7	*2.7*	*0.7*	*0.4*	*2.2*	*4.1*	*0.2*	*0.5*	*2.2*	*2.0*	*2.8*	*1.6*	*62.9*

Notes: The table reports the number of spells with a particular training-occupation combination as a percentage of all spells in the training (the occupation, second row in italics) for the baseline sample. Results are restricted to those with ten years of work experience.

Table A.3: Average On- versus Off-Diagonal Returns - no exp_k IV

	(1)	(2)	(3)	(4)
$D_{j=k} = 1$	-0.0144	0.1248	-0.0360	0.1220
	(0.0038)	(0.0274)	(0.0041)	(0.0257)
exp	0.0619	0.0617	0.0458	0.0457
	(0.0006)	(0.0006)	(0.0012)	(0.0012)
exp^2	-0.0010	-0.0010	-0.0004	-0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
exp_k			0.0175	0.0171
			(0.0009)	(0.0009)
exp_k^2			-0.0008	-0.0007
			(0.0000)	(0.0000)
Parametric cf	no	yes	no	yes
p-value cf		0.000		0.000
N	1,123,574	1,123,574	1,123,574	1,123,574

Notes: The table reports regression results for Equation (1) with $\tau_{jk} = \delta_k + \tau D_{j=k}$. All regressions include individual, occupation, region and time fixed effects. In contrast to Table 3, exp_k is *not* instrumented in the regressions. The control function accounts for this and uses exp_k instead of its instrument as a predictor for the choice probabilities. Bootstrap standard errors are reported in parentheses.

Table A.4: Average On- versus Off-Diagonal Returns - Robustness I

	(1) no movers	(2) app. length $2.5-3$ years	(3) no capped wages
$D_{j=k}=1$	0.1413	0.2079	0.0991
	(0.0329)	(0.0576)	(0.0339)
exp	0.0506	0.0666	0.0512
	(0.0025)	(0.0054)	(0.0021)
exp^2	0.0000	-0.0007	-0.0002
	(0.0001)	(0.0002)	(0.0001)
exp_k	0.0151	-0.0081	0.0115
	(0.0026)	(0.0051)	(0.0020)
exp_k^2	-0.0013	-0.0005	-0.0011
	(0.0001)	(0.0003)	(0.0001)
N	922,680	277,676	1,011,578

	(4) no app.-firm-switchers	(5) after re-unification	(6) firm-spec. exp. control
$D_{j=k}=1$	0.1538	0.1323	0.1386
	(0.0358)	(0.0317)	(0.0244)
exp	0.0549	0.0620	0.0500
	(0.0022)	(0.0017)	(0.0020)
exp^2	-0.0003	-0.0004	-0.0001
	(0.0001)	(0.0001)	(0.0001)
exp_k	0.0090	0.0043	0.0128
	(0.0022)	(0.0015)	(0.0020)
exp_k^2	-0.0010	-0.0008	-0.0012
	(0.0001)	(0.0001)	(0.0001)
N	1,008,167	1,007,346	1,123,574

Notes: The table reports regression results for Equation (1) with $\tau_{jk} = \delta_k + \tau D_{j=k}$. Columns (1)-(5) restrict the baseline sample as indicated in the column header. Column (6) uses the full sample and controls for full time firm-specific experience and its square. All regressions include individual, occupation, region and time fixed effects. As in Section 6, the variable exp_{ikt} is replaced by its first stage prediction for each sample. Bootstrap standard errors are reported in parentheses.

Table A.5: Average On- versus Off-Diagonal Returns - Robustness II

	(1) 10th-order polynomial	(2) train prob. cf rob.	(3) AR(1) vac. split	(4) quadr. vac. split
$D_{j=k}=1$	0.1401	0.1483	0.1450	0.1345
	(0.0246)	(0.0247)	(0.0276)	(0.0276)
exp	0.0553	0.0553	0.0554	0.0544
	(0.0020)	(0.0020)	(0.0021)	(0.0025)
exp^2	−0.0003	−0.0003	−0.0003	−0.0003
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
exp_k	0.0084	0.0084	0.0083	0.0094
	(0.0021)	(0.0021)	(0.0020)	(0.0025)
exp_k^2	−0.0010	−0.0010	−0.0010	−0.0010
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
N	1,123,574	1,123,574	1,125,764	1,122,069
	(5) 30 occ. categ.	(6) occ. x time FE	(7) ind. x time FE	(8) occ. x t x ind. FE
$D_{j=k}=1$	0.0921	0.1395	0.1391	0.1598
	(0.0228)	(0.0253)	(0.0330)	(0.0273)
exp	0.0542	0.0576	0.0496	0.0506
	(0.0018)	(0.0019)	(0.0020)	(0.0023)
exp^2	−0.0004	−0.0005	−0.0002	−0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
exp_k	0.0090	0.0044	0.0122	0.0085
	(0.0018)	(0.0019)	(0.0019)	(0.0022)
exp_k^2	−0.0009	−0.0008	−0.0011	−0.0009
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
N	1,123,574	1,123,571	1,122,163	1,121,883

Notes: The table reports regression results for Equation (1) with $\tau_{jk} = \delta_k + \tau D_{j=k}$: (1) controls for a tenth order polynomial in own vacancies; (2) estimates training probabilities using the region/time at the start of the apprenticeship as predictors; (3)/(4) uses an AR(1)/quadr. model to split vacancies; (5) uses 30 occ. categories; (6)-(8) include fixed effects as shown in the header. Industries are: 1) agricult., energy, mining; 2) prod. of rubber & plastic products, proc. of minerals, wood ind.; 3) chemicals; 4) metal prod. & proc., mech. engineering; 5) automotive, prod. of data proc. equipment, electrical & optical engineering; 6) consumer goods; 7) hospitality; 8) building ind.; 9) sale, maintenance & repair of motor vehicles & household goods; 10) transport & comm.; 11) credit & insurance interm., land & housing, rentals; 12) public & personal serv., hh serv.; 13) educ., soc. & healthcare; 14) public admin., soc. security. Bootstrap standard errors are reported in parentheses.

Table A.6: Full Matrix of Returns - No Selection Control

	Occupation												
Training	01-09	10-54	19-32	31	33-37	39-43	44-51	60-63	68-70	71-74	75-78	79-89	90-93
01-09: Agric.	0	0.01	0.06	−0.04	0	0.08	−0.02	0.11	0.03	0.01	0.02	0.00	−0.12
		(0.04)	(0.03)	(0.07)		(0.13)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.10)
10-54: Process	−0.12	0	−0.11	−0.04	0.09	−0.13	−0.09	0.06	−0.08	−0.13	0.02	−0.12	−0.79
	(0.07)		(0.04)	(0.07)	(0.12)	(0.08)	(0.05)	(0.02)	(0.05)	(0.04)	(0.03)	(0.05)	(0.24)
19-32: Craft	−0.10	0.01	0	−0.01	−0.13	−0.03	0.02	0.13	0.04	−0.03	0.06	−0.09	−0.25
	(0.04)	(0.01)		(0.02)	(0.06)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.06)
31: Electrical	−0.00	0.06	0.08	0	−0.11	0.11	−0.02	0.16	0.11	−0.02	0.19	0.04	−0.08
	(0.03)	(0.02)	(0.01)		(0.15)	(0.07)	(0.04)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)	(0.07)
33-37: Textile	0.06	0.06	0.16	0.21	0	−0.30	0.13	0.14	−0.02	−0.01	−0.01	0.06	−0.18
	(0.24)	(0.04)	(0.09)	(0.20)		(0.14)	(0.06)	(0.06)	(0.07)	(0.07)	(0.05)	(0.09)	(0.08)
39-43: Food	0.01	0.00	0.13	−0.07	0.28	0	0.09	0.20	0.15	0.07	0.14	0.04	−0.04
	(0.05)	(0.01)	(0.02)	(0.06)	(0.11)		(0.02)	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)
44-51: Constr.	−0.10	−0.10	−0.00	−0.03	−0.52	−0.08	0	0.07	−0.04	−0.09	−0.02	−0.15	−0.24
	(0.03)	(0.01)	(0.02)	(0.04)	(0.49)	(0.05)		(0.01)	(0.03)	(0.01)	(0.02)	(0.03)	(0.06)
60-63: Techn.	−0.03	−0.12	0.03	−0.09	−0.16	0.03	−0.10	0	0.05	−0.11	−0.02	−0.19	−0.05
	(0.07)	(0.04)	(0.05)	(0.03)	(0.01)	(0.11)	(0.05)		(0.04)	(0.08)	(0.04)	(0.10)	(0.21)
68-70: Sales	−0.39	−0.02	0.05	0.02	0.07	−0.09	−0.05	0.17	0	−0.01	0.01	−0.11	−0.29
	(0.14)	(0.02)	(0.03)	(0.07)	(0.06)	(0.06)	(0.04)	(0.03)		(0.02)	(0.01)	(0.03)	(0.05)
71-74: Transp.	−0.15	−0.08	0.02	−0.10	0	−0.54	0.08	0.05	0.01	0	0.02	−0.05	−0.13
	(0.07)	(0.05)	(0.07)	(0.13)		(0.31)	(0.11)	(0.09)	(0.07)		(0.03)	(0.08)	(0.14)
75-78: Office	−0.43	−0.08	0.01	−0.10	0.08	−0.23	−0.08	0.08	0.07	−0.06	0	−0.07	−0.47
	(0.15)	(0.02)	(0.03)	(0.10)	(0.03)	(0.14)	(0.05)	(0.03)	(0.01)	(0.02)		(0.02)	(0.06)
79-89: Health	−0.23	−0.10	−0.09	0.01	0.16	−0.41	−0.13	0.08	−0.02	−0.12	−0.07	0	−0.60
	(0.07)	(0.04)	(0.07)	(0.09)	(0.86)	(0.14)	(0.09)	(0.04)	(0.02)	(0.05)	(0.02)		(0.08)
90-93: Person.	−0.06	0.22	0.34	0.29	0.28	0.05	0.21	0.22	0.17	0.15	0.17	0.16	0
	(0.10)	(0.04)	(0.04)	(0.06)	(0.05)	(0.04)	(0.04)	(0.06)	(0.03)	(0.04)	(0.03)	(0.04)	

Notes: The table shows coefficient estimates $\hat{\tau}_{jk}$ from Equation (1), estimated without selection control. Bootstrap standard errors are reported in parentheses.

Table A.7: Full Matrix of Returns - Parametric Selection Control

	Occupation												
Training	01-09	10-54	19-32	31	33-37	39-43	44-51	60-63	68-70	71-74	75-78	79-89	90-93
01-09: Agric.	0	−0.40	−0.03	−0.47	0	0.13	−0.16	0.04	0.08	−0.49	−0.44	−0.16	−0.41
		(0.13)	(0.15)	(0.22)		(0.27)	(0.19)	(0.15)	(0.15)	(0.13)	(0.12)	(0.17)	(0.30)
10-54: Process	0.26	0	0.32	0.02	0.86	0.43	0.29	0.51	0.50	−0.14	0.05	0.24	−0.56
	(0.29)		(0.21)	(0.23)	(0.90)	(0.27)	(0.21)	(0.19)	(0.20)	(0.15)	(0.19)	(0.18)	(0.50)
19-32: Craft	0.17	0.06	0	0.03	0.43	0.38	0.27	0.43	0.44	−0.04	0.08	0.16	−0.09
	(0.17)	(0.04)		(0.13)	(0.64)	(0.17)	(0.08)	(0.05)	(0.06)	(0.04)	(0.04)	(0.09)	(0.22)
31: Electrical	0.49	0.29	0.60	0	0.77	0.76	0.46	0.70	0.76	0.13	0.37	0.51	0.28
	(0.20)	(0.06)	(0.08)		(0.77)	(0.24)	(0.12)	(0.08)	(0.08)	(0.09)	(0.06)	(0.10)	(0.29)
33-37: Textile	0.03	−0.27	0.20	−0.16	0	−0.14	0.10	0.19	0.15	−0.44	−0.39	0.00	−0.37
	(0.46)	(0.31)	(0.32)	(0.43)		(0.45)	(0.34)	(0.31)	(0.30)	(0.33)	(0.31)	(0.33)	(0.51)
39-43: Food	0.32	0.04	0.49	−0.06	0.94	0	0.39	0.57	0.65	0.03	0.13	0.33	0.13
	(0.20)	(0.08)	(0.11)	(0.18)	(0.78)		(0.13)	(0.10)	(0.11)	(0.09)	(0.10)	(0.12)	(0.29)
44-51: Constr.	−0.06	−0.29	0.07	−0.25	−0.14	0.13	0	0.17	0.17	−0.36	−0.26	−0.13	−0.32
	(0.19)	(0.05)	(0.08)	(0.15)	(0.76)	(0.19)		(0.06)	(0.09)	(0.06)	(0.06)	(0.11)	(0.26)
60-63: Techn.	−0.62	−1.00	−0.49	−1.00	−0.37	−0.36	−0.68	0	−0.34	−1.08	−0.95	−0.79	−0.78
	(0.29)	(0.17)	(0.17)	(0.21)	(0.88)	(0.38)	(0.21)		(0.16)	(0.17)	(0.15)	(0.19)	(0.41)
68-70: Sales	−0.27	−0.15	0.20	−0.13	0.44	0.18	0.06	0.34	0	−0.20	−0.13	−0.03	−0.30
	(0.21)	(0.06)	(0.10)	(0.16)	(0.63)	(0.23)	(0.10)	(0.08)		(0.06)	(0.07)	(0.10)	(0.24)
71-74: Transp.	0.06	−0.19	0.29	−0.24	0	−0.13	0.28	0.35	0.42	0	−0.15	0.12	−0.09
	(0.41)	(0.25)	(0.27)	(0.33)		(0.52)	(0.29)	(0.28)	(0.20)		(0.21)	(0.31)	(0.41)
75-78: Office	−0.62	−0.50	−0.13	−0.55	0.20	−0.24	−0.26	−0.05	0.00	−0.56	0	−0.27	−0.75
	(0.26)	(0.07)	(0.10)	(0.19)	(0.66)	(0.24)	(0.09)	(0.06)	(0.06)	(0.07)		(0.10)	(0.22)
79-89: Health	−0.69	−0.80	−0.51	−0.74	−0.05	−0.70	−0.64	−0.34	−0.35	−0.92	−0.80	0	−1.19
	(0.26)	(0.07)	(0.13)	(0.17)	(1.15)	(0.23)	(0.14)	(0.10)	(0.08)	(0.09)	(0.07)		(0.22)
90-93: Person.	−0.21	−0.20	0.24	−0.17	0.45	0.06	0.05	0.13	0.17	−0.36	−0.29	−0.03	0
	(0.25)	(0.12)	(0.13)	(0.20)	(0.78)	(0.25)	(0.17)	(0.14)	(0.13)	(0.10)	(0.10)	(0.16)	

Notes: The table shows coefficient estimates $\hat{\tau}_{jk}$ from Equation (1), estimated using the parametric control function estimator. Bootstrap standard errors are reported in parentheses.

Table A.8: List of Tasks and Fraction Performing

Task	01-09	10-54	19-32	31	33-37	39-43	44-51	60-63	68-70	71-74	75-78	79-89	90-93
1: Cultivate	80	3	1	1	0	2	4	1	0	2	0	2	1
2: Repair	33	23	60	76	27	6	54	19	4	15	4	8	5
3: Equip	38	60	53	48	45	39	30	29	8	24	9	12	12
4: Manufact.	25	36	42	33	62	66	46	15	4	3	3	5	8
5: Serve	4	1	0	1	0	37	0	1	4	1	2	10	30
6: Clean	25	18	13	10	25	45	19	4	12	15	3	17	66
7: Teach	24	13	20	26	19	22	21	41	45	14	39	65	21
8: Sell	35	5	11	15	14	28	16	25	79	14	34	20	29
9: Pack	38	37	22	19	16	28	32	11	32	78	17	14	14
10: Research	25	36	42	51	31	32	30	63	30	16	35	46	14
11: Design	20	10	16	19	14	19	20	49	20	9	24	30	12
12: Secure	10	10	9	10	5	8	9	11	5	11	6	19	5
13: Ex. laws	5	2	3	6	1	5	2	20	8	6	25	20	2
14: Employ	21	11	14	20	8	21	16	49	33	14	38	36	14
15: Nurse	16	4	6	5	7	15	3	6	17	9	9	43	26
16: Publish	1	0	0	1	0	1	0	6	4	1	5	17	2
17: Program	9	18	14	23	8	6	7	47	35	14	54	29	6
18: Calculate	16	4	6	9	4	15	10	31	34	5	41	9	9
19: Correct	20	11	9	16	7	10	10	37	50	21	74	44	12

Notes: The table shows the average percentage of individuals indicating they perform the given task. To construct averages, observations in each wave are weighted using survey weights and subsequently combined giving equal weight to each wave. Task 1: cultivate; task 2: repair, renovate, reconstruct; task 3: equip or operate machines; task 4: manufacture, install or construct; task 5: serve or accommodate; task 6: clean; task 7: teach or train others; task 8: sell, buy or advertise; task 9: pack, ship or transport; task 10: research, evaluate or measure; task 11: design, plan, sketch; task 12: secure; task 13: execute laws or interpret laws; task 14: employ, manage personnel, organize, coordinate; task 15: nurse or treat others; task 16: publish, present or entertain others; task 17: program; task 18: calculate or do bookkeeping; task 19: correct texts or data.

Table A.9: Training-Occupation Distances - Selected Categories

Statistics	Training j	Occupation k	$Dist_{jk}$
Overall mean			0.3461
Standard dev.			0.1500
Weight. mean			0.2805
	Craft workers	Electrical w.	0.0209
	Craft workers	Construction w.	0.0421
	Construction w.	Electrical w.	0.0632
	Craft workers	Process, plant w.	0.0834
	⋮	⋮	⋮
	Office workers	Craft workers	0.5428
	Craft workers	Personal serv. w.	0.5435
	Electrical w.	Personal serv. w.	0.5585
	Office workers	Textile, garment w.	0.5887

Notes: The table reports summary statistics on the distance measure $Dist_{jk}$, and distances for the five most similar and the five most distant training-occupation pairs.

Table A.10: Training-Occupation Distances - Five Largest Occupations

		Occupation				
		Office workers	Craft workers	Sales, fin. workers	Health workers	Constr. workers
Training	Office workers	0				
	Craft workers	0.54	0			
	Sales, fin. w.	0.11	0.53	0		
	Health, soc. w.	0.16	0.44	0.20	0	
	Construction w.	0.51	0.04	0.46	0.42	0

Notes: The table reports the distance measure $Dist_{jk}$ for the five largest occupations.

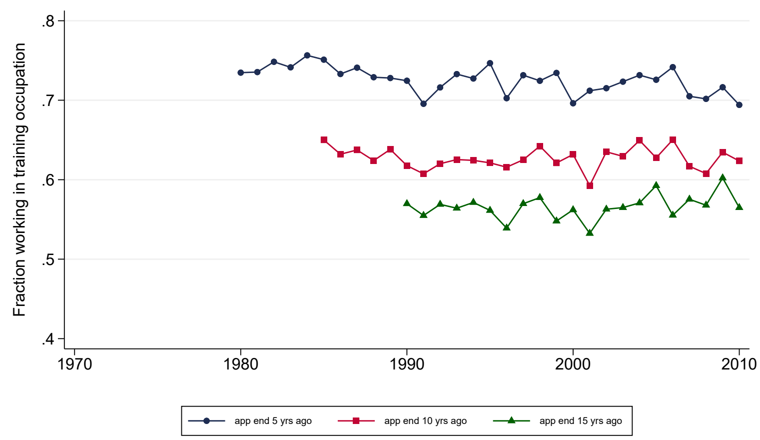
Table A.11: Match Returns and Task Distance - Robustness

τ_{jk} estimated	with parametric control fcn.		without selection control	
	(1)	(2)	(3)	(4)
A. No on-diag. obs.				
$Dist_{jk}$	-0.0814	-0.0471	-0.0164	0.0112
	(0.0357)	(0.0330)	(0.0157)	(0.0122)
Occ. FE	no	yes	no	yes
Mean of $\hat{\tau}_{jk}$	-0.0563	-0.0563	-0.0259	-0.0259
R-squared	0.0379	0.3105	0.0095	0.4748
N	156	156	156	156
B. Largest trainings				
$Dist_{jk}$	-0.0683	-0.0333	-0.0286	0.0042
	(0.0339)	(0.0340)	(0.0151)	(0.0119)
Occ. FE		yes		yes
Mean of $\hat{\tau}_{jk}$	-0.1596	-0.1596	-0.0791	-0.0791
R-squared	0.0452	0.3419	0.0434	0.6046
N	65	65	65	65
C. Quadratic spec.				
$Dist_{jk}$	-0.0579	-0.0265	-0.0124	0.0084
	(0.0378)	(0.0337)	(0.0169)	(0.0132)
$Dist^2_{jk}$	0.0240	0.0294	0.0133	0.0072
	(0.0379)	(0.0408)	(0.0154)	(0.0135)
Occ. FE		yes		yes
Mean of $\hat{\tau}_{jk}$	-0.0520	-0.0520	-0.0239	-0.0239
R-squared	0.0395	0.3009	0.0217	0.4535
N	169	169	169	169

Notes: The table reports regression results from Equation (18). $Dist_{jk}$ is scaled by its standard deviation. Observations are weighted by the sample fraction in the relevant training. Robust standard errors are reported in parentheses. Panel A excludes on-diagonal observations where $Dist_{jk} = 0$. Panel B only includes cells for the five largest trainings. Panel C uses a quadratic specification in task distance.

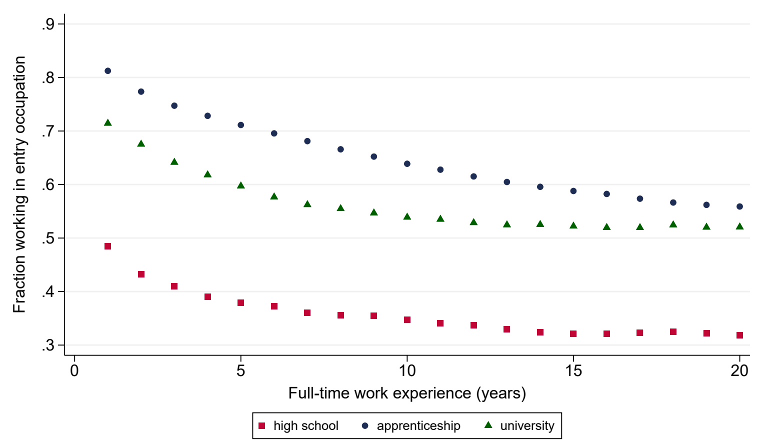
Appendix B. Figures

Figure B.1: Fraction On Diagonal over Time



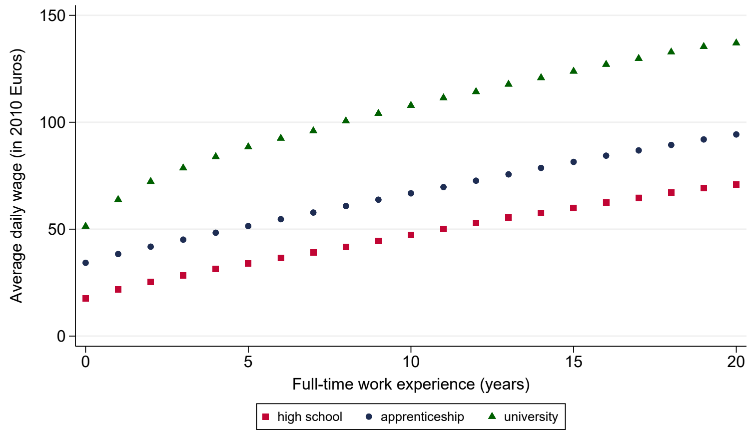
Notes: The figure plots the fraction of individuals working in an occupation equal to their training occupation over time for the baseline sample. The three lines plot this fraction for individuals who finished their apprenticeship 5, 10, or 15 years prior to the date shown on the x-axis.

Figure B.2: Fraction in Entry Occupation by Education Groups



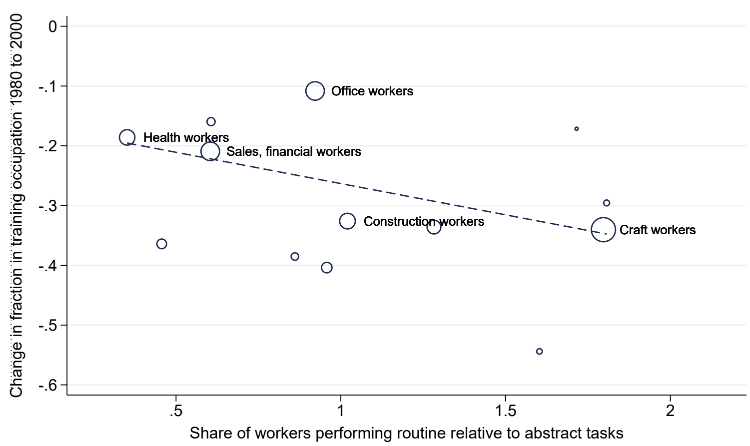
Notes: The figure plots the fraction of full-time workers working in their entry occupation for high school, apprenticeship and university graduates. Entry age is restricted to 18, 23, 27 for high school, apprenticeship and university graduates, respectively (based on average graduation ages).

Figure B.3: Wage-Experience Profiles by Education Groups



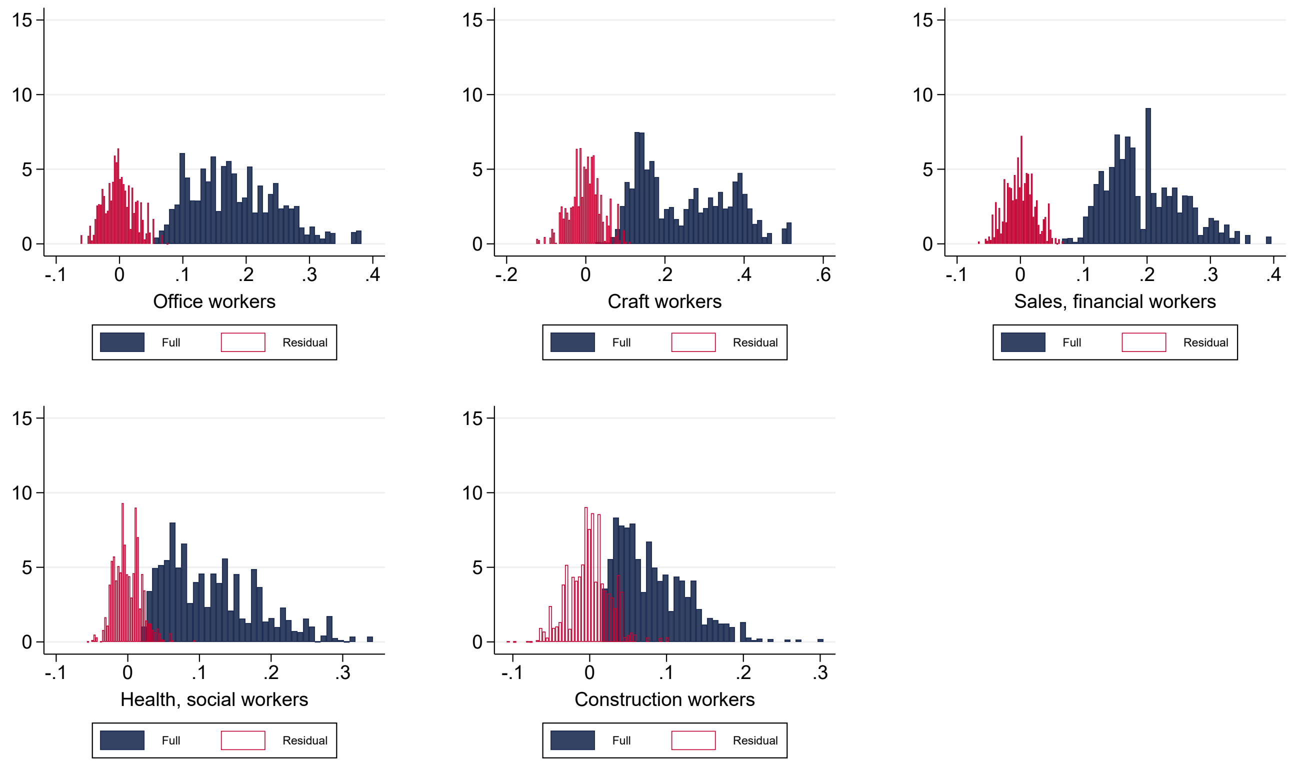
Notes: The figure plots average full-time wages (in 2010 Euros) by full-time experience for high school, apprenticeship and university graduates. Entry age is restricted to 18, 23 and 27 for high school, apprenticeship and university graduates, respectively (based on average graduation ages).

Figure B.4: 1980 to 2000 Change in Fraction on Diagonal and Tasks



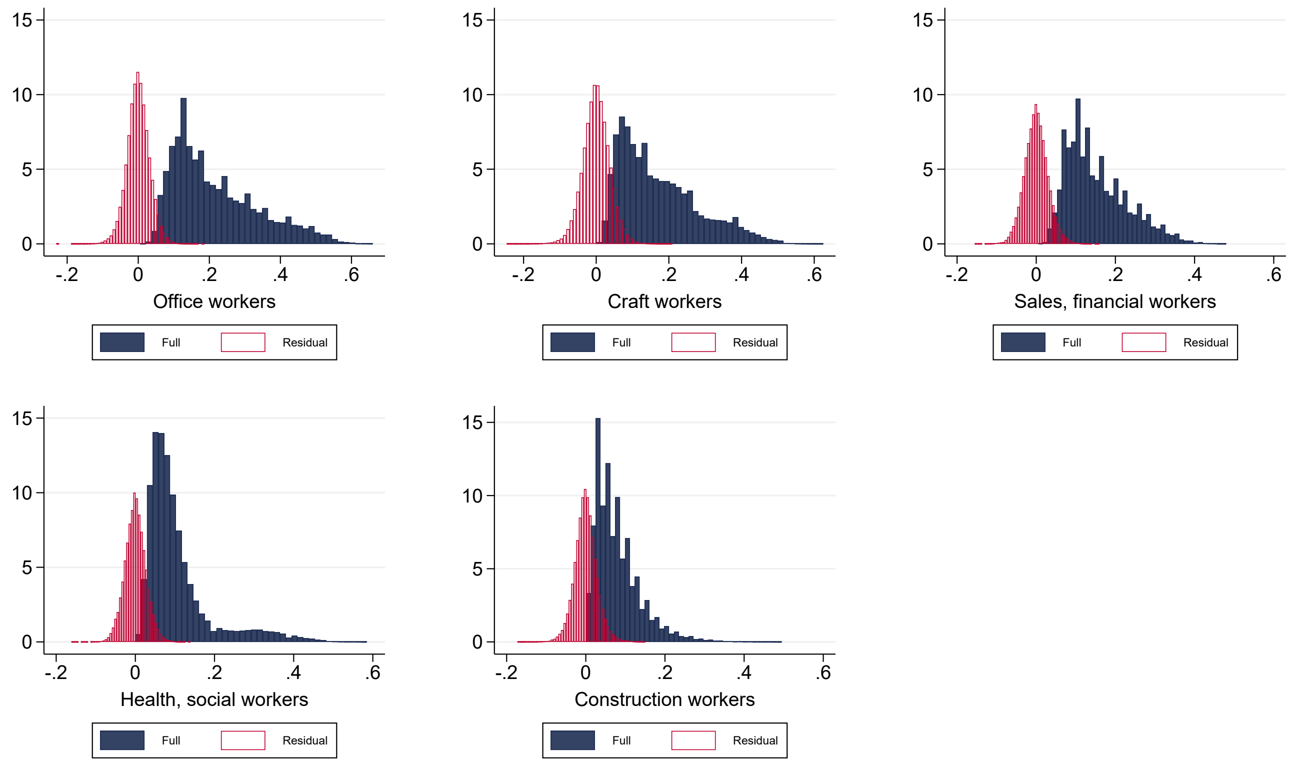
Notes: This figure plots the fraction of workers performing "routine" tasks divided by the fraction performing "abstract" tasks on the x-axis (measured following Dustmann et al. (2009)), and the *change* from 1980 to 2000 in the fraction of workers trained between 1975 and 1980 working in their training occupation on the y-axis. Marker size is proportional to fraction of spells in sample.

Figure B.5: First Stage Variation in Selection Probabilities - Training



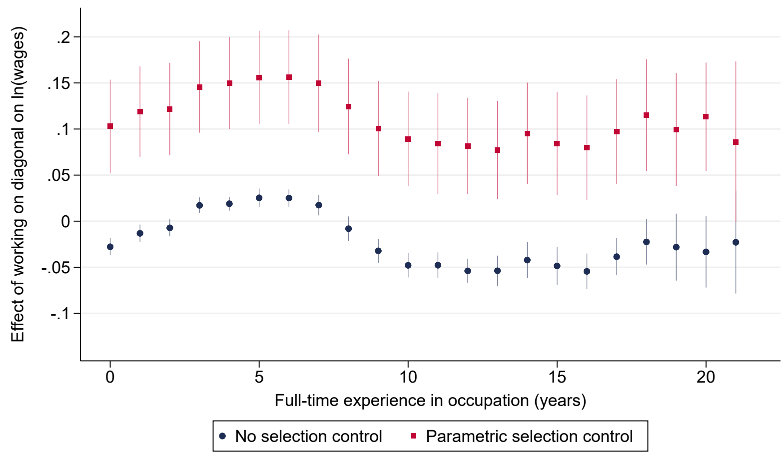
Notes: The figure shows a set of histograms of the estimated selection probabilities for the five largest trainings. Histograms in blue show the full variability in estimated selection probabilities. Histograms in red are restricted to male workers, and residualized using location and time of training fixed effects.

Figure B.6: First Stage Variation in Selection Probabilities - Occupation



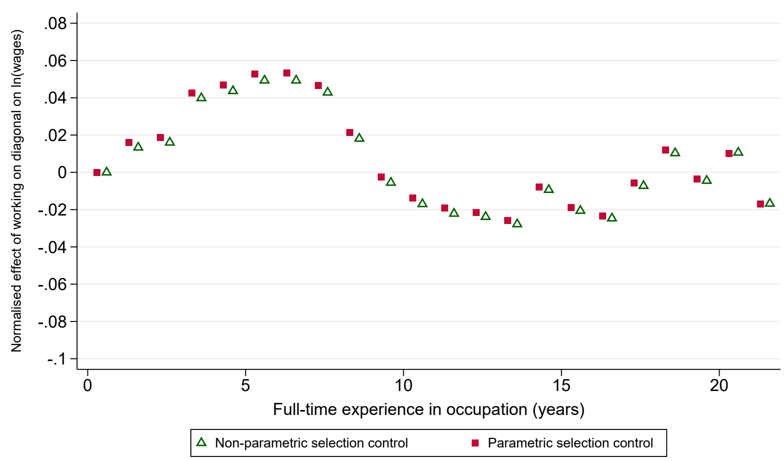
Notes: The figure shows a set of histograms of the estimated selection probabilities for the five largest occupations. Histograms in blue show the full variability in estimated selection probabilities. Histograms in red are restricted to on-diagonal workers and residualized using individual fixed effects, full-time experience, vacancies in the given occupation, and region and time fixed effects.

Figure B.7: On- versus Off-Diagonal Returns by Occ.-Specific Experience



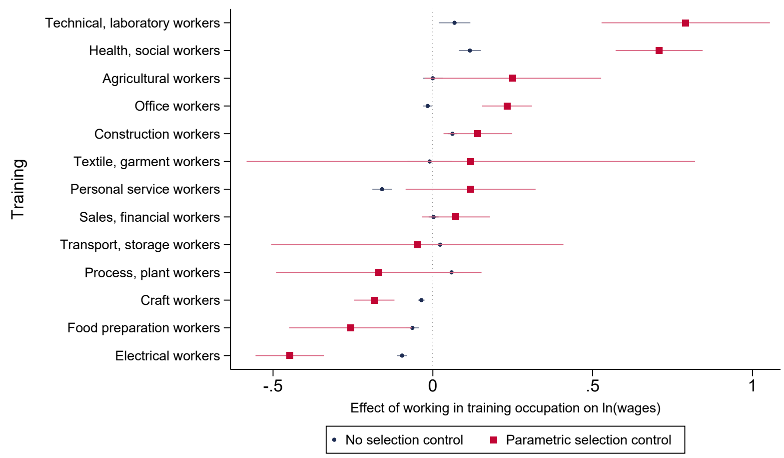
Notes: The figure plots regression coefficient estimates for τ^{exp_k} in a version of Equation (1) with $\tau_{jk} = \delta_k + \tau^{exp_k} D_{j=k}$. 95% bootstrap confidence intervals are shown.

Figure B.8: Normalized On- versus Off-Diagonal Returns by Occ.-Specific Experience



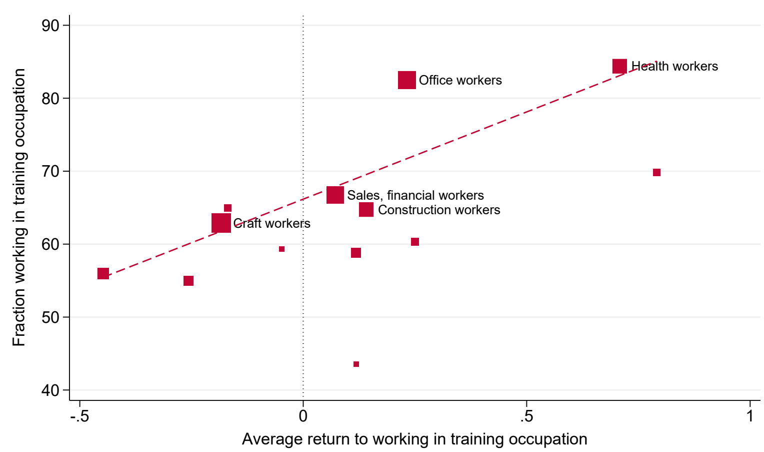
Notes: The figure plots regression coefficient estimates for τ^{exp_k} in a version of Equation (1) with $\tau_{jk} = \delta_k + \tau^{exp_k} D_{j=k}$. Coefficient estimates are normalized to zero at zero years of work experience.

Figure B.9: Average On- versus Off-Diagonal Returns by Training



Notes: The figure shows regression coefficient estimates for τ_j in a version of Equation (1) with $\tau_{jk} = \tau_j D_{j=k}$. 95% bootstrap confidence intervals are shown.

Figure B.10: Average Return and Fraction Working in Training Occupation



Notes: The figure plots average on- versus off-diagonal returns for each training against the fraction working on the diagonal. The fitted line corresponds to a weighted regression using the sample fraction in each training as weights. Marker size is proportional to the weights.

Appendix C. Model Details

Wages. The formulation of wages can be motivated by an aggregate technology in which within-occupation human capital units are perfect substitutes (e.g., Heckman & Sedlacek (1985), Keane & Wolpin (1997)). Individual i is endowed with h_{ikt} units of occupation-k-specific human capital in t. Output in occupation k in region r at time t, Y_{krt}, depends on the sum of occupation-specific human capital units H_{krt} across workers and non-labor inputs K_{krt}. The aggregate production function is given by

$$Y_{krt} = F^{(krt)}(H_{krt}, K_{krt}),$$

where $F^{(krt)}$ is a twice-continuously differentiable function that is strictly concave in its arguments. The marginal product of a unit of human capital is given by

$$MPH_{krt} = \frac{\partial F^{(krt)}(H_{krt}, K_{krt})}{\partial H_{krt}}.$$

With perfectly competitive markets, aggregate demand for human capital is pinned down by $\bar{W}_{krt} = MPH_{krt}$ where $\bar{W}_{krt}$ is the market skill price per unit of human capital. For any level of $\bar{W}_{krt}$, wages of individual i in occupation k are given by

$$w_{ikrt} = \bar{W}_{krt} h_{ikt} e^{\upsilon_{ikrt}},$$

where υ_{ikrt} captures random influences. Following standard human capital formulations (e.g., Griliches (1977), Keane & Wolpin (1997)), I assume that the human capital production function is time-invariant and takes an exponential form:

$$h_{ikt} = e^{\delta_i} e^{\tau_{j(i)k}} e^{\beta' X_{ikt}} e^{\varphi_{ik}},$$

where δ_i denotes i's time-invariant general human capital endowment, $\tau_{j(i)k}$ captures the contribution of i's training in j to human capital in occupation k, X_{ikt} includes general and occupation-specific work experience and their squares, and φ_{ik} captures individual unobserved human capital components such as occupation-specific ability. Allowing for a flexible effect of different trainings on occupation-specific human capital has the advantage of avoiding assumptions on the relationship between training skills and tasks. Parameters $\tau_{j(i)k}$ are also directly policy-relevant as they link to workers' discrete choice sets. Using the above formulation, equilibrium log wages are given by

$$ln(w_{ikrt}) = \bar{w}^*_{krt} + \delta_i + \tau_{j(i)k} + \beta' X_{ikt} + \epsilon_{ikrt},$$

where $\bar{w}^*_{krt}$ denotes the equilibrium log skill price per unit of human capital and $\epsilon_{ikrt} = \varphi_{ik} + \upsilon_{ikrt}$ captures non-random and random influences on wages.

Suppose the market has many firms that post vacancies to hire workers in each occupation, and that the vacancy posting cost is constant. Assuming a spot market model of the labor market where workers are hired for one period, aggregate posting vac_{krt} in each period is pinned down by the labor demand function $\bar{W}_{krt} = MPH_{krt}$.

In equilibrium, log skill prices then depend on aggregate posting and supply factors as follows: $\bar{w}^*_{krt} = \delta_r + \delta_t + f(ln(vac_{krt}))$, where δ_r and δ_t denote region and time fixed effects. Note that this specification abstracts from skill-price effects of occupation-specific supply within time and regions which are assumed to be small in the short run. In the empirical analysis, I show robustness to several skill price parameterizations. Using the baseline formulation, log wages are given by Equation (1) in Section 4 .

Information Sets. Workers make choices according to Equations (3) and (7) observing all current-period variables, including the error terms e_{ikrt}, $e_{ijr_0t_0}$ that are unobservable to the researcher. The utility error terms e_{ikrt} when choosing *occupations* contain the wage error term ϵ_{ikrt} and a signal $s_{tr}(\phi_{ik})$ a worker receives about their time-invariant occupation-specific preferences ϕ_{ik}: $e_{ikrt} = \epsilon_{ikrt} + s_{tr}(\phi_{ik})$.

Similarly, the utility error terms $e_{ijr_0t_0}$ when choosing a *training* contain a preference signal $s_{t_0r_0}(\phi_{ij})$. In contrast to the occupation choice stage, workers also face uncertainty regarding their occupation-specific abilities, and observe a signal about their ability in the occupation that corresponds to each training j, $s_{t_0r_0}(\varphi_{ij})$. The error term is then given by $e_{ijr_0t_0} = s_{t_0r_0}(\phi_{ij}) + s_{t_0r_0}(\varphi_{ij})$.

The above decompositions capture the notion that workers may learn about their preferences and abilities over time, in line with the idea that imperfect information at the time of training choice causes occupational mismatch (see Section 3.2). Specifically, the above model allows for updating about preferences in every period and for updating about occupation-specific abilities relative to the time of training choice. In addition, occupational mismatch may occur because future observations of vac_{krt} are not observed. I assume that workers cannot affect the information they obtain through choices, but impose no further assumptions on learning.

Appendix D. Estimation Details

Lee/Dahl Approach. To ensure tractability, the control function approach in Section 5.2 approximates $E[\epsilon_{ikrt}|M_{ijkrt} = 1]$ by $E[\epsilon_{ikrt}|\zeta_{ijkrt} \leq \Phi^{-1}(p_{ijkrt})]$. The terms are exactly equivalent in simultaneous choice problems where selection is driven by a single maximum order statistic (see Dahl (2002)). In the given sequential setting, the conditioning set on M_{ijkrt} is a subset of that on ζ_{ijkrt} since the function defining ζ_{ijkrt} is increasing in its two arguments, so that $\zeta_{ijkrt} \leq \Phi^{-1}(p_{ijkrt})$ holds for those selecting into jk but it may also hold for some who only select into either j or k.

How closely does the conditioning set on ζ_{ijkrt} approximate that on M_{ijkrt}? To understand this, note that, at the minimum, both sets exclude combinations of selection errors such that neither choice j nor k is optimal. Depending on the distribution of maximum order statistics, the set on ζ_{ijkrt} may further exclude many error combinations that imply that only j or k is optimal (thus implying a close approximation), but this is difficult to gauge empirically. That possibility aside, excluding selection errors such that neither j nor k is optimal resolves most of the selection problem if i) the excluded subset is large and if ii) outcome errors in the subset are the dominant cause of bias. In the given setting, the former holds since the set of jk cells is large, so that combinations of errors such that neither j nor k is optimal are most likely.[41] The latter holds if individuals with occupation-k abilities that are most different from those of individuals selecting into jk choose neither j nor k. This is plausible for on-diagonal cells since both the training and occupation choice likely reveal information about abilities. By the same logic, it is also plausible for off-diagonal cells where j and k are similar in terms of their tasks, and these are the ones that dominate empirically. Taken together, these points suggest that $E[\epsilon_{ikrt}|\zeta_{ijkrt} \leq \Phi^{-1}(p_{ijkrt})]$ is likely to approximate $E[\epsilon_{ikrt}|M_{ijkrt} = 1]$ closely for the majority of the data.

To understand how Dahl (2002)'s approach differs from Lee (1983)'s, define the control function $\lambda_{jk}(\cdot) = E[\epsilon_{ikrt}|M_{ijkrt} = 1]$, which depends on the conditional joint distribution of the error and the two maximum order statistics from Equation (10), where the conditioning is on all sub-utility differences. Given a mapping between these differences and selection probabilities, the distribution can instead be conditioned on

[41] The weighted average of the probability of selecting into neither j nor k across all jk cells, where the weight is the sample size in each cell, exceeds 0.8.

the probabilities and the control function may be written as

$$\lambda_{jk}(\cdot) = \lambda_{jk}(p_{i1r_0t_0}, ..., p_{iJr_0t_0}, p_{i(1|j)rt}, ..., p_{i(K|j)rt}). \quad \text{(D.1)}$$

With sequential choices, $\lambda_{jk}(\cdot)$ depends only on occupation probabilities that condition on the observed training, but estimating $\lambda_{jk}(\cdot)$ flexibly is infeasible. Lee's transformation and parametric assumption make estimation feasible, but A6 is equivalent to assuming that the distribution of outcome errors and maximum order statistics does not depend on the sub-utility differences (see Dahl (2002) for details). For parameters on variables that vary within selection cells, Dahl shows that this may be relaxed by allowing for dependence on a small set of selection probabilities, a so-called index sufficiency assumption. Implementation may proceed by estimating the probabilities that affect the control function, and approximating $\lambda_{jk}(\cdot)$ using non-parametric techniques. In the given setting, parameters that vary within jk cells include the change in on-diagonal returns across levels of occupation-specific experience. For these parameters, I employ Dahl's non-parametric method, where I approximate $\lambda_{jk}(\cdot)$ using a polynomial function, interacted with selected jk-cells (see Section 6.2.1).

Splitting Vacancies into Expectation and Shock. To obtain the instruments defined in Equations (14) and (15), vacancies need to be split into expectations and shocks. To do so, I estimate linear time trend models for each cell r_0t_0, where log vacancies in each occupation are explained using five years of past data. Note that this eliminates the first five years in the sample. I estimate the following models:

$$ln(vac_{krt}) = \kappa_k^{r_0t_0} + \pi_k^{r_0t_0} \times t + \varepsilon_{krt}. \quad \text{(D.2)}$$

Note that I allow intercepts and slopes to be occupation-specific. In a robustness check, I also include a quadratic time variable. For each r_0t_0, 30-year-ahead predictions for vacancies are computed as conditional expectations using Equation (D.2):

$$E[ln(vac_{k(t_0+\tau)})|\Omega_{r_0t_0}] = \hat{\kappa}_k^{r_0t_0} + \hat{\pi}_k^{r_0t_0} \times (t_0 + \tau), \quad \forall \tau = 0, ..., 30, \forall k. \quad \text{(D.3)}$$

Note that the training instruments defined in Equation (14) use a j instead of k to denote the training in an occupation. I set vacancy predictions smaller than one to one, but this only affects 1.7% of the predictions, even at the 30-year horizon. For any $t = t_0 + \tau$, individual-specific shocks to vacancies are then defined as residuals relative to the expectation formed at the time of training choice t_0 in region r_0:

$$ln(vac_{krt}) - E[ln(vac_{k(t_0+\tau)})|\Omega_{r_0 t_0}], \quad \forall \tau = 0, ..., 30, \forall k. \tag{D.4}$$

The conditional expectations based on Equation (D.3) serve as training instruments I^j, and the residuals from Equation (D.4) serve as occupation instruments I^k.

I conduct a further robustness exercise where I use an AR(1) model for each time period t_0 to predict vacancies using five years of past data. I estimate this using an Arellano-Bond estimator. 30-year ahead predictions and shocks are derived as before.

Estimating the Selection Probabilities. The variables used to predict occupation choices conditional on training, $occ_{i(k|j)rt}$, include 13 instruments for all occupations, $13 \times 31 = 403$ training instruments, full-time experience, the instrument for occupation-specific experience (see below), vacancies in all occupations, the training choice, and region and time fixed effects. Variables are adjusted according to the regression specification, e.g., the instrument for occupation-specific experience is not used for specifications that do not control for exp_{ikt}. Since random forests are trained and applied to separate samples of individuals, individual fixed effects cannot be included. To partly account for preferences, I include a gender variable.

The explanatory variables used for the prediction of training choices, $train_{ij}$, include $13 \times 31 = 403$ instruments for all trainings defined by Equation (14). Vacancies, region and time fixed effects, and the occupation instruments are not included as they are determined after the training choice, and cannot affect this choice. However, the latter may be correlated with a training choice through the region and time in which the training was started r_0, t_0. I thus provide a robustness check where I also use r_0, t_0 as explanatory variables in the prediction of training probabilities (see column (2) in Table A.5). Note that there is no natural way of including the omitted variables themselves as training choices do not vary by occupation, region and time. As before, individual fixed effects cannot be included in the prediction, and I instead include a gender variable to capture the effect of individual characteristics on training choice.

The explanatory variables are used as inputs into random forests which are amongst the most accurate classifiers (Breiman (2001), see Hastie *et al.* (2009) for details on the algorithm). I use 50% of individuals as training dataset to grow random forests for training and occupation choices. Both forests are based on 500 trees, where 1000 randomly selected observations from the training dataset are used to grow each tree.

In a second step, the forests are applied to the remaining 50% of the sample. Probability predictions for each training or occupation option in this dataset are computed as the proportion of counts for that option across all trees in the final nodes.

Instrument for Occupation-Specific Experience. Occupation-specific experience exp_{ikt} reflects past selection into occupations. To address this endogeneity concern, I follow Altonji & Shakotko (1987), and use the deviation of occupation-specific experience from its individual-specific mean in the sample as an instrument. Consider the following decomposition of the error term in Equation (1): $\epsilon_{ikrt} = \varphi_{ik} + \upsilon_{ikrt}$. Since the instrument sums to zero over the sample years in which i works in occupation k, it is orthogonal to φ_{ik} by construction. Occupation-specific experience and its instrument will also not be related to the transitory component if υ_{ikrt} is serially uncorrelated, which appears to be a reasonable assumption in the given context.

To implement this strategy, I run a first stage model to predict exp_{ikt} using the instrument and its square, and the exogenous variables from the outcome equation (experience and its square, vacancies, the occupation instruments, individual, occupation, region and time fixed effects). Note that the training instruments are absorbed by individual fixed effects. I then use the prediction as control in the regressions. In Section 6.2.1, I discretize the predicted value of exp_{ikt} into yearly categories.

Table D.1: First Stage for Occupation-Specific Experience

Dependent variable	exp_k	exp_k^2
exp	0.4257	−7.1160
	(0.0041)	(0.0900)
exp^2	−0.0102	0.5026
	(0.0002)	(0.0061)
IV	0.8231	15.3436
	(0.0040)	(0.0741)
IV^2	0.0260	0.7913
	(0.0002)	(0.0073)

Notes: $N = 1,123,574$. The table reports regression results with occ.-specific experience and its square as dependent variables. Regressions control for individual, occupation, region and time fixed effects, a fourth order polynomial in vacancies, and the occupation instruments. The instrument IV is defined in this section. Bootstrap standard errors are reported in parentheses.

Appendix E. Welfare and Policy

Locked-in Workers. In addition to off-diagonal workers, a second group of workers affected by the lack of information at the time of training choice are workers who are locked into their training. These workers would choose a different occupation in the absence of off-diagonal penalties, but currently work on the diagonal as their payoff elsewhere is insufficient to compensate for the lack of training. As in Section 8.1, consider the on- versus off-diagonal model where $\tau_{jk} = \delta_k + \tau D_{j=k}$. Using Equations (1) and (2) in Section 4, the welfare loss relative to the optimal allocation in the training-occupation matrix for locked-in workers is bounded from above by the on-diagonal return τ. This is because, by revealed preference, locked-in workers are worse off working off the diagonal. In particular, for an on-diagonal worker currently working in k, we know that

$$\begin{aligned} \tilde{U}_{i(k|j)rt} + e_{ikrt} &\geq \tilde{U}_{i(k'|j)rt} + e_{ik'rt}, \ \forall k' \neq k \\ \tau &\geq (\tilde{U}_{i(k'|j)rt} + e_{ik'rt}) - (\tilde{U}_{i(k|j)rt} + e_{ikrt}) + \tau, \ \forall k' \neq k, \end{aligned} \tag{E.1}$$

where the term on the right hand side is the gain from switching to an occupation k' while keeping the on-diagonal return τ. Using the estimate from column (4) in Table 3, the welfare loss for on-diagonal workers is thus bounded from above by 14%.

Since locked-in workers work on the diagonal, they are not directly observed in the sample, and estimating the share of these workers requires further assumptions. Consider the fraction of on-diagonal workers over the experience schedule from Figure 3 in Section 3.1. The downward sloping pattern partly arises due to changes in the on- versus off-diagonal return considered in Section 6.2.1. In addition, other factors including new information about own abilities or the labor market contribute to the declining share on the diagonal. If one is willing to make assumptions about the latter, it is possible to calculate the fall in on-diagonal work induced by a 1*pp* reduction in on-diagonal returns. To proceed with this exercise, I assume that information updates arrive at a constant rate between 5 and 15 years of work experience. Next, it can be seen from Figures B.7 and 3 that between 5 and 10 years of experience, the on-diagonal return falls by about 7*pp*, and the fraction of on-diagonal individuals falls by about 5*pp*. Then, between 10 and 15 years of experience, the on-diagonal return is stable, and the fraction of on-diagonal individuals falls by about a further 3*pp*. Since there is no change in on-diagonal returns during this time, these 3*pp* can be attributed to information updates rather than changes in returns. With constant information updates, this implies that, between 5 to 10 years of experience, $2pp$ of the $5pp$ fall in individuals on the diagonal may be attributed to the $7pp$ fall in the on-diagonal return. With a total on-diagonal return of 14%, this in turn means that $4pp$ fewer individuals would be on the diagonal in the absence of any return. These individuals are, by definition, locked into their training.

This implies that the total welfare loss from locked-in workers is at most 10% of that of off-diagonal workers. Given the required assumptions on worker learning, this result should be considered as suggestive.

Retraining Calculations. *Total costs in Euros:* Average annual costs per apprentice are around 5,400 Euros for firms and 6,800 Euros for government bodies (figures available for 2012/13. Source: *Finanzierung der beruflichen Ausbildung in Deutschland, BWP 2/2016, BiBB*). In terms of private cost, the average yearly difference in earnings between an apprentice and a trained worker with less than 15 years of work experience was about 20,200 Euros in 2010 (Source: *BiBB, press release 01/2011* and author's own calculations using the administrative data).

Net benefits in Euros: My estimates suggest that the annual average gain of retraining τ corresponds to 14% of wages per worker. The cost of a year of foregone work experience is about 6%. Assuming that the effective foregone work experience of two years of retraining is one year (apprentices work in firms for two thirds of their time), the net gain of retraining is therefore equal to 8%. Based on average annual earnings of 32,600 Euros in 2010, this amounts to 2,608 Euros in 2010.

Cost-benefit calculations: Assuming a discount factor of 0.98, retraining costs would be recouped after 35 years of subsequent work in the new occupation:

$$32,400 + \beta \times 32,400 = \beta^2 \times 2,608 \times \frac{1-\beta^{t+1}}{1-\beta}$$
$$t \approx 34.5. \tag{E.2}$$

Based on an average training completion age of 23, and a retirement age of 67, off-diagonal workers would thus need to switch out of their training with at most seven years of work experience for retraining to be profitable ($67 - 23 - 2 - 35 = 7$ years).