Essays in the Analysis of Matched Employer-Employee Data

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

Jasper Haller

Abteilungssprecher: Prof. Klaus Adam, PhD

Gutachter: Prof. Dr. Andrea Weber und Prof. Dr. Markus Frölich

Datum der mündlichen Prüfung: 18.10.2022

Hiermit erkläre ich, dass ich diese Arbeit selbstständig angefertigt und die benutzten Hilfsmittel vollständig und deutlich angegeben habe.

Contents

Chapter 1

Introduction

Traditionally, research in labor economics concerned itself with the study of workers or, less frequently, firms, but not both at the same time. While this research produced many classic results, it was necessarily silent on a host of important questions. Is there associative matching between high-wage workers and high-paying firms? To what extent are changes in worker outcomes attributable to sorting into different types of firms by unobservable characteristics? Conversely, to what extent are differences between firms, such as inter-industry wage gaps, attributable to sorting of different types of workers? How much, and through which channels, does the selection into high- or low-wage jobs contribute to wage growth? Do men obtain a higher share of firm-specific rents than women? These and many other important issues remained underexplored.

It was the advent of large administrative datasets containing matched employer-employee data that opened a new segment within labor economics to tackle these issues. The seminal paper of Abowd, Kramarz, and Margolis (1999) set out a statistical framework for the analysis of matched employer-employee data and, perhaps even more importantly, staked out the substantive economic questions on which the analysis of employer-employee data would prove fruitful. More than twenty years later, countless contributions from across labor economics have harnessed matched employer-employee data to progress important economic issues. This dissertation situates itself in this tradition, and while it contains three quite distinct chapters, each with a different focus, the common thread is asking what we can learn by exploiting the observed mobility of workers across workplaces.

Chapter 2 uses Austrian administrative matched employer-employee data to study the evolution of wage growth for young workers. Specifically, how much of wage growth is due to mobility towards higher-paying firms, and how much is due to improving idiosyncratic matches between worker and firm? The answer to this question has important implications, for example, for thinking about failing firms: if low-wage firms are not capable of contributing to wage growth, we should arguably care less when a low-wage firm fails than if all firms could contribute to wage growth by way of idiosyncratic matches with workers. I use fixed- and random-effects estimation to tackle this question, but the main results are the same regardless of the chosen method. First, mobility contributes to wage growth during the first few years on the labor market and fades out thereafter. Second, job moblity's contribution comes entirely in the form of improved firm effects, not improved idiosyncratic matches.

Chapter 3, written jointly with Anna Raute, turns to a different data source: administrative data from the human resource systems of the Evangelical Church in the Rhineland, Germany. Over a 26-year period, this unique dataset allows us to track the mobility of pastors across parishes, which we can link to a rich set of outcomes at the parish level, including church membership, service attendance, confirmations, and donations. In a contribution to the literature on the effect of increased female representation in firm management, we test for differences in outcomes across pastor gender during a period when the proportion of female pastors more than doubled. While we find no gender differences for most outcomes, there is one stark finding: female pastors attract significantly fewer donations than their male counterparts. This finding is echoed in some contributions from the literature of the sociology of religion.

Chapter 4, written jointly with Andrea Weber, returns to Austrian administrative matched employer-employee data and studies an important pathway into the labor market: firm-based apprenticeships. While there is a rich economic literature on apprenticeships, this literature generally treats all firms as homogeneous. This is a potentially serious omission in light of the now large body of literature that documents significant firm heterogeneity by productivity. We construct a stylized theoretical model to derive predictions about the effect of firm heterogeneity on the decision to offer training: more productive firms offer fewer apprenticeships but provide higher-quality training. We test some of our predictions empirically using various proxies for firm productivity, and consistently find support for our predictions.

This document has my name on it, but it has many more contributors. I would like to thank Andrea Weber for invaluable guidance; Markus Frölich for kindly agreeing to review this piece of work; Anna Raute, who served not just as a co-author but as a mentor and friend; and my friends and colleagues at the University of Mannheim, particularly Albrecht Bohne, Andreas Dzemski, Anna Hammerschmid, Florian Exler, Jan Nimczik, Johannes Dittrich, Johannes Schneider, Julia Schmieder, Katharina Momsen, Linh Nguyen and Maria Isabel Santana, for their advice and support. Most importantly, my wife Friederike provided the most tremendous support during the countless hours that I invested into this project, while my wonderful children Nathan, Tobias and Letizia never failed to remind me what matters most in life. It has been a very long road, and I could not have walked it without you.

Chapter 2

Firm Heterogeneity, Job Changes, and Wage Growth of Young Workers

2.1 Introduction

Two classic findings in labor economics characterize the careers of young workers as periods of rapid wage growth (Mincer, 1974) and high job mobility (Topel and Ward, 1992). More recently, labor economists have begun to look for a causal relationship between these two phenomena: does job mobility result in higher wages? The answer appears to be yes. For example, estimates by Altonji, A. Smith, and Vidangos (2013) attribute a fifth of all wage growth to moving across firms into higher-paying jobs.

While wage growth due to job mobility is now a well-documented phenomenon, its causes are not yet fully understood. One prominent possibility is that persistent differences between firms along dimensions such as productivity, bargaining power with workers, or product market competition lead some firms to pay higher wages than others to *all* of their workers. Indeed, Burdett and Mortensen (1998) show that in a wage-posting model with search frictions, even ex ante identical firms will pay different wages in equilibrium. In this case, the average worker experiences wage gains over time because he sorts himself into ever higher-paying firms. A second possibility, frequently considered in the matching literature (e.g., Mortensen and Pissarides, 1994), is idiosyncratic matches between worker and firm: some worker-firm pairings are just more productive than others. These productivity differences then translate into wage differences if workers capture a share of the generated surplus. As a result, wage growth occurs when workers move towards firms where matches are better, but these may be different firms for different workers.

This paper sets out to distinguish between these two rivaling explanations. To this end, I will use matched employer-employee data from Austrian social security records to estimate several extended versions of the models by Abowd, Kramarz, and Margolis (1999, hereafter AKM) and Woodcock (2008; 2015). I ask the following three questions: First, how important is job mobility for wage growth overall? Second, does wage growth happen because workers move towards firms which pay high wages generally, or does it arise because workers improve the idiosyncratic match between them and their employer? Third, does the answer to the previous question differ for workers at different stages of their career?

Distinguishing the different channels of wage differences between firms is important for at least two reasons. One, it may guide economic theory as to which assumptions about the wage process are attractive. Two, it can help us think about the value of holding a job and the costs of job displacement, due to, e.g., firm closures. If between-firm differences in pay arise mostly because of persistent differences between firms, then we should worry greatly about the costs of a firm closure by a "high-wage" firm, but less so if it is a "low-wage" firm. By contrast, if differences arise mostly because of match effects, then all firm closures should be similarly worrisome, because they all destroy worker-firm matches that may be hard to replace.

I find that job mobility accounts for 5 to 10 percent of total wage growth during the first 5 years on the labor market, and the effects fade out thereafter. Crucially, mobility-induced wage growth is almost entirely due to an improvement of firm effects, rather than match effects. Therefore, my findings strongly support the hypothesis that wage growth is driven by persistent differences across firms, rather than improvements in worker-firm matches.

Table 2.1: Creation of Baseline Sample

Sample Creation Step	Sample Size	% Lost
All employment records, split up into annual spells	52,664,572	
Restricted to 1 spell per year and firm-employee pair	46,659,104	11.40%
Restricted to spells that last at least 30 days in a year	45,276,104	2.96%
Restricted to spells with valid information on age and birthyear	45,275,916	0.00%
Restricted to spells between age 14 and 65	45,151,616	0.27%
Restricted to spells with valid wage information	38,062,728	15.70%
Restricted to spells after the beginning of a stable labor market career	36,074,232	5.22%
Restricted to one job per person and year	32,051,000	11.15%
Restricted to firms whose industry classification is known	32,047,163	0.01%

The last column shows the percentage by which the sample size was reduced during the respective step.

2.2 Data

The analysis uses data from the Austrian Social Security Database (ASSD), which includes detailed information on the universe of Austrian private-sector employment relationships derived from social security records. For every employment spell since 1972, the dataset contains the precise start and end date of the spell, an identifier for each employer[1] and employee, some characteristics of the worker, the firm, and the job spell, and the wages paid from the employer to the employee for each calendar year. Since social security contributions in Austria are only paid up to an annual earnings limit, annual earnings are recorded only up to that limit. I therefore merge the ASSD with the Austrian labor tax register to obtain uncensored earnings. As an extra benefit, the labor tax register contains an indicator for whether the job is part-time, which is not available in the ASSD.

I have constructed the baseline sample constructed as follows. Since the tax database only has had a valid part-time indicator since 2002, I have restricted the sample to the years from 2002 to 2012. For these years, I have merged the ASSD with the tax database to obtain uncensored

[1]From the data description, it is not entirely clear whether an employer identifier refers to a firm or a single establishment, of which a large firm might have several. Fink et al. (2010) study this issue in more detail by comparing the number of employer identifiers in the ASSD to the number of firms in other sources, and conclude that the vast majority of ASSD employer identifiers seem to refer to firms, not establishments. Therefore, I will use the terms employer and firm interchangeably.

earnings. There are a few workers with censored earnings for whom no corresponding uncensored record could be located in the tax data, as well as missing values for the part-time dummy in about 8 per cent of observations. In these cases, I have resorted to imputing the missing values, where the imputation is based on the observed uncensored distribution of earnings for other earners. I have also imputed missing information on the part-time dummy based upon the level of earnings as well as gender, whether the worker is white- or blue-collar, and whether she is observed working part-time in other years. All earnings have been adjusted for inflation using the consumer price index. For the later analyses, it will be necessary to define a main job for each person and year; for workers holding multiple jobs during a year, I have seclected the one with the highest total earnings for the year. The outcome variable is daily earnings, which is obtained by dividing annual earnings by the number of days worked. Finally, to limit the impact of outliers and recording errors, I have censored all earnings from below at 0.1 times the median wage for that year, and from above at 10 times the median wage. Table 2.2 contains some descriptive statistics of this sample.

The advantage of using the ASSD is the precise information on employment histories and earnings for an entire country over a long time frame. In particular, the dataset allows for the construction of a variable for labor market experience, which counts the number of years since 1972 during which a worker has held at least one job. There are a few drawbacks, however. First, the dataset contains no information on occupations, job titles, and educational attainment. I proxy educational attainment by using the age of first entry into the labor market. Second, being derived from administrative sources, the dataset contains no information on why a job spell ended. I therefore resort to a proxy for whether a transition is voluntary or not by considering whether there is a gap of at least 28 days of non-employment between employment spells.

Table 2.2: Descriptive Statistics, Baseline Sample

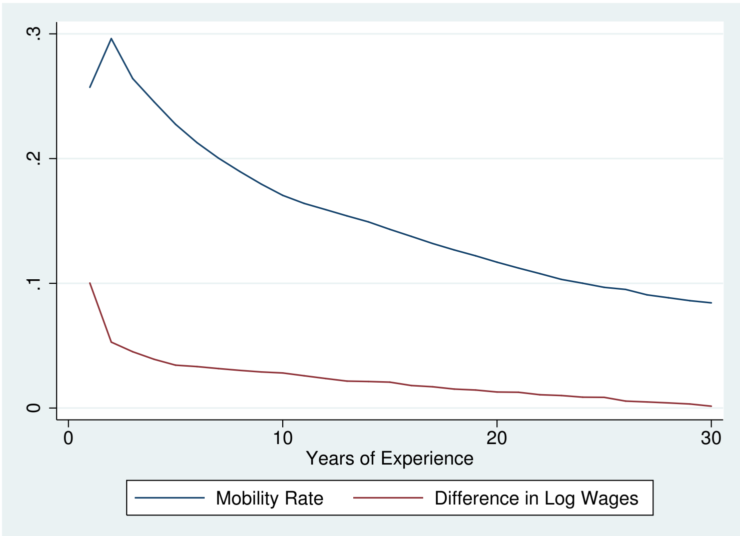
	Number of Observations	Mean	Standard Deviation
Worker-Years	32,047,163		
Daily Wage, in Euros		74.34	55.21
Blue-Collar Work		0.45	0.50
Part-Time Work		0.20	0.40
Labor Market Experience		14.75	10.08
Workers	4,423,311		
Women		0.47	0.50
Age at Labor Market Entry < 18		0.11	0.31
Age at Labor Market Entry 18–21		0.51	0.50
Age at Labor Market Entry > 21		0.38	0.49
Firms	503,230		
Firm Size		63.68	746.05
Firm Size < 5		0.38	0.49
Firm Size 5–20		0.31	0.46
Firm Size > 20		0.31	0.46

The sample covers the years 2002–2012. Each individual has only one entry per year, the main job, defined as the job with the highest total earnings during the year. If the individual was an apprentice with the firm for part of the year, the variable "Apprenticeship" measures the fraction of the days that the individual has spent as an apprentice of the total number of days he spent with the firm during the year.

2.3 Some Descriptive Evidence

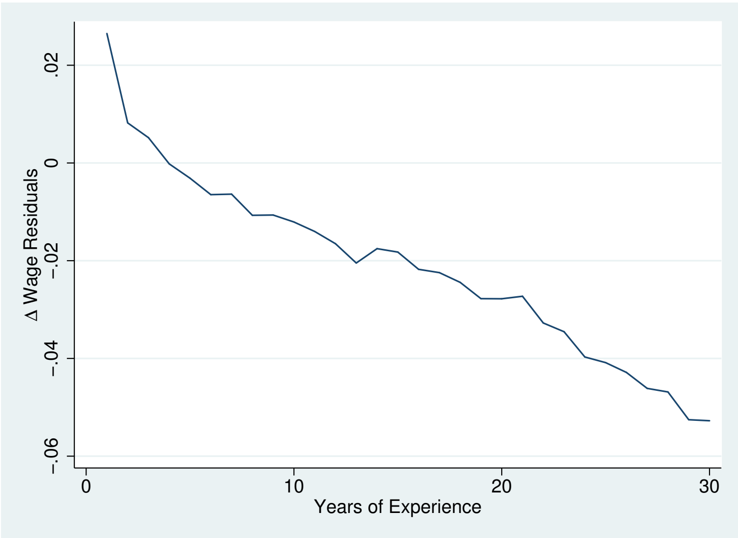
Previous research has extensively documented two basic facts about labor market careers over the life-cycle. First, it is well known, at least since the work of Mincer (1974), that wages grow over the course of a career, but at a decreasing rate. Second, as documented by Topel and Ward (1992) and others, job mobility is substantial over the first few years of workers' careers before tapering off. Both of these facts hold also for Austria, and are on display in Figure 2.1. Workers' nominal daily wages increase annually by 0.2–0.3 log points during the first 10 years on the labor market before growth tapers off steadily. Job mobility peaks after 3 years on the labor market, when almost 20% of workers switch employers. Three years later, the mobility rate has fallen well below 10%.

Figure 2.1: Annual Wage Growth and Job Mobility Rates, by Years of Experience



The fact that wage growth and job mobility are especially high at the beginning of workers' careers naturally leads to the question of whether the two phenomena are related. Using a variety of econometric techniques, Topel and Ward (1992), Altonji, A. Smith, and Vidangos (2013), and Bagger, Fontaine, et al. (2014) have all found that mobility increases wages; I begin here with a simple check by studying wage growth for workers who move from one job to the next, net of what they would have been expected to gain based on their increase in labor market experience alone. With the obvious caveat that the workers who move may be nonrandomly selected, Figure 2.2 shows residual wage changes of movers by labor market experience. The figure indicates some wage gains for movers during the first few years on the labor market, with zero or even negative returns to mobility thereafter. Taken together, Figures 2.1 and 2.2 therefore suggest a tight link between job mobility and wage growth, but only during the first few years on the labor market.

Figure 2.2: Residual Wage Growth at Job Changes, by Years of Experience



Wage residuals are derived by regressing log wages on a full set of experience dummies.

One might hypothesize that a large fraction of early-career job changes are voluntary, and therefore tend to be associated with wage gains, whereas later job changes tend to be the result of firings or firm closures, and are thus associated with wage losses. However, Figure 2.3 and Table 2.3 suggest that this is not the case. If anything, the share of transitions without an intervening unemployment spell, used as a proxy for the share of voluntary job changes, is *lower* for workers with little experience. As seen in Table 2.3, young workers gain from job-to-job transitions, on average, but lose if the transition is separated by an unemployment spell. By contrast, older workers lose from both kinds of job changes, although more so in the case of

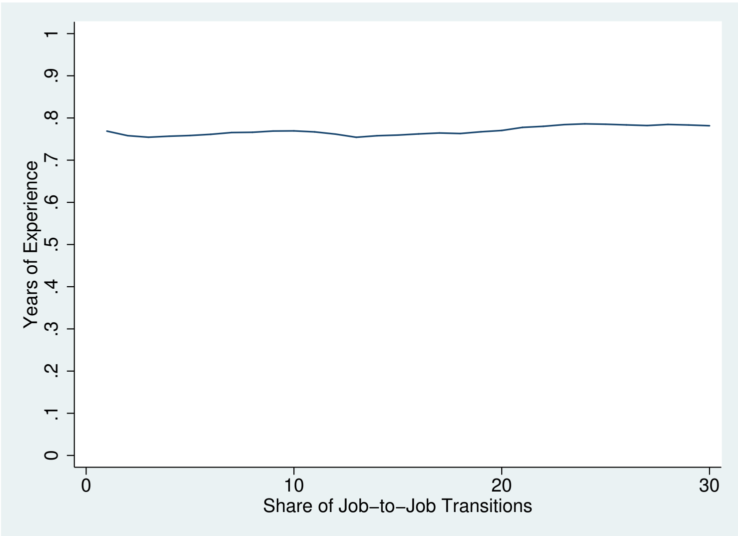
Table 2.3: Changes in Log Wages by Transition Type

Years of Experience	0–5	6–10	11–15	16–20	More than 20
Job-to-Job Transitions	0.020	0.001	-0.009	-0.018	-0.042
Job-Unemployment-Job Transitions	-0.028	-0.043	-0.043	-0.045	-0.058

Job-Unemployment-Job Transitions are transitions where either the old job is followed by an unemployment spell of at least 28 days, or the new job is preceded by an unemployment spell of at least 28 days. All other transitions are coded as Job-to-Job Transitions.

job-unemployment-job transitions.

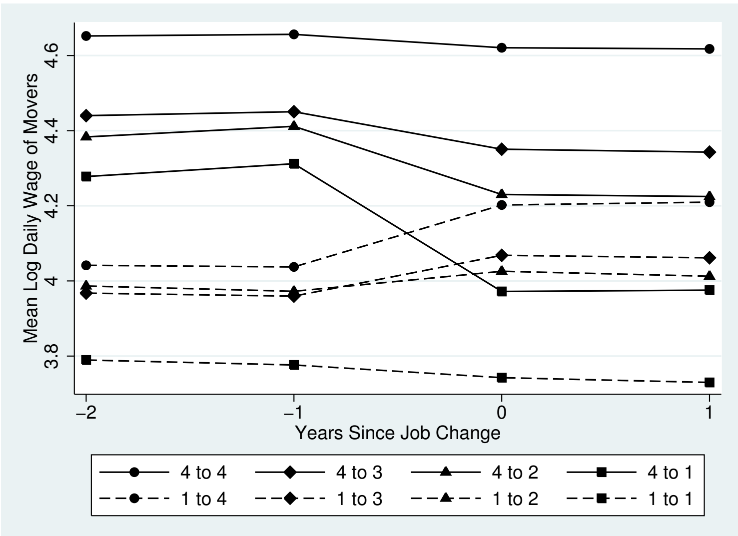
Figure 2.3: Job-to-job Transitions as a Fraction of All Job Transitions, by Years of Experience



A transition is defined as being job-to-job if the old employment spell is not followed by an unemployment spell of more than 28 days, *and* the new employment spell is not preceded by an unemployment spell of more than 28 days. Since the sample contains just one job per person and year, cases are possible where one condition is met but not the other.

The preceding observations indicate that job mobility and wage growth are important phenomena in the careers of young workers in particular, and it is worth setting up a more formal econometric model to study these in more detail. An important question is what kind of model would provide a good description of the data. Card, Heining, and Kline (2013) produce an event study of wage changes akin to Figure 2.4. The figure groups workers into quartiles based on the average wages received by their coworkers. Changing from quartile 1 to quartile 4 means that a worker leaves a firm in which her coworkers' average wage fall into the bottom quartile, and joins a firm in which her coworkers' average wage fall into the top quartile. As Card, Heining, and Kline (2013) point out, the gains associated with switching from a low- to a high-wage firm are approximately the same as the losses associated with switching from a high- to a low-wage firm. They therefore argue that a model containing additive worker and firm effects fits the data well. In particular, such a model need not contain a match effect that is idiosyncratic to the worker-firm pair and is allowed to be correlated with the assignment of workers to firms. If such match effects were important, the argument goes, workers would tend to improve them as they switch jobs, and they would experience wage growth regardless of whether they switched from firm A to B or from firm B to A. This would preclude the observed symmetry in wage gains and losses.

Figure 2.4: Mean Wages of Job Changes Grouped by Average Coworker Wages at Origin and Destination Firm



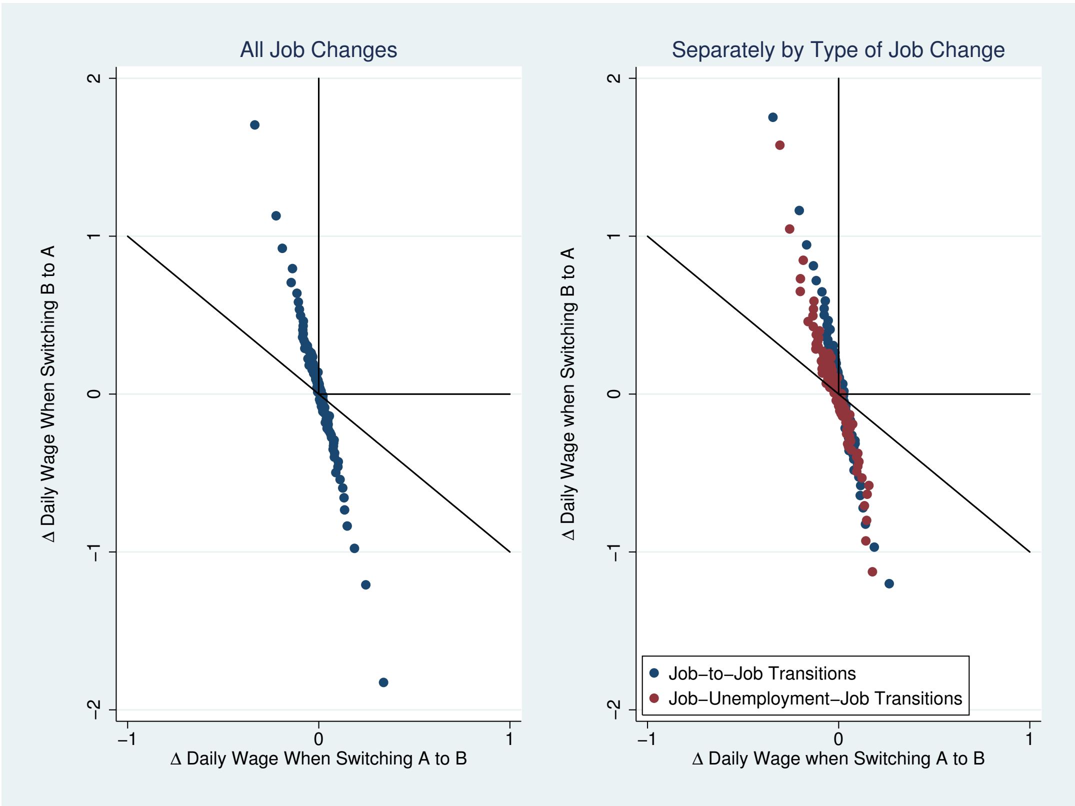
This figure shows the evolution of wages around job changes. Job changes are grouped according to the quartile of the origin and destination firms in the distribution of average wages paid to all other workers except the worker making the job change. Job changes are included in the creation of this figure only if the origin and destination job last at least two years.

Woodcock (2015) dissents and advocates a model containing an additive match effect in addition to worker and firm effects. He argues that what matters is the type of transition from one employment spell to the next: whereas workers who make a job-to-job transition tend to do so voluntarily and will improve their match on average, workers who make a job-unemployment-job transition tend to do so involuntarily and will see their match deteriorate on average. By this argument, the overall symmetry of wage gains and losses should not be taken as an indicator that match effects are unimportant, since match effects do not necessarily improve as workers change jobs. If, however, we focus only on those workers making job-to-job transitions, which should tend to improve the worker-firm match, wages should tend to improve regardless of whether workers switched from firm A to B or from firm B to A. For these transitions, the observed symmetry between wage gains and losses should disappear, or at least weaken substantially.

Figure 2.5 and Table 2.4 confront these two conflicting views with the data. I go one step further than Card, Heining, and Kline (2013) and identify firm pairs for which I observe at least one worker switching from firm A to firm B and one worker switching from B to A. In total, there are 315,901 such firm pairs in the data, formed by 1,137,918 job changes. By focusing on such firm pairs, I rule out that the firms towards which workers are moving are different along some time-invariant characteristic from the firms from which workers move away. In Figure 2.5, I have grouped these firm pairs by the percentile of wage changes as workers move from firm A to B. Ignoring random wage fluctuations over time, if Card, Heining, and Kline (2013) are correct, all points should line up on the diagonal line. That is, if a worker gains X Log-Euros moving from firm A to B, another worker should correspondingly lose X Log-Euros as he moves from B to A, and the correlation between gains and losses should be -1. By contrast, if workers switch jobs on the basis of match effects, we should see excess mass in the top right quadrant, where we find firm pairs where workers experience wage gains both by switching from A to B and from B to A.

Figure 2.5 provides some support for both stories, but more so for the former. Focusing first on the left panel, there are indeed some dots in the top right corner, but their number is small. As Table 2.4 shows, there are wage gains in both directions for 28.2% of firm pairs. This is not overwhelming, especially since there is an equal proportion of firm pairs for which we observe wage *losses* in both directions. Moreover, the correlation between A-to-B wage gains and B-to-A wage gains is -0.19. While this is significantly different from -1, the estimated correlation is likely to be heavily affected by the small number of movers: the median firm pair has just a single worker switching from firm A to B and one switching from B to A. This makes it difficult to measure the wage effect of switching from A to B or vice versa with precision, and will naturally depress the measured correlation towards zero.

Figure 2.5: Mean Wage Changes when Switching Firms



Each point is a group of firm pairs formed by the percentile in the distribution of wage changes when moving from firm A to B. The X-axis shows the mean change in daily wages, in Log-Euros, when switching from firm A to firm B for this group. The Y-axis shows the corresponding mean change in daily wages when switching from B to A for this group of firms.

The most remarkable aspect of Figure 2.5 is the clearly discernible slope. As we move towards firm pairs with a larger average wage gain for movers from A to B, we observe ever larger corresponding wage losses for movers from B to A. Most firm pairs show average wage gains for one direction of job movement and wage losses for the other. The right panel of

Table 2.4: Change in Log Wages when Transitioning between Firms A and B

Distribution of Wage Changes			
		Switches from B to A	
		Δ Log Wage < 0	Δ Log Wage > 0
Switches from A to B	Δ Log Wage > 0	28.23%	24.37%
	Δ Log Wage < 0	18.98%	28.43%
Correlation		-0.20	
Weighted Correlation		-0.19	

Weights are proportional to the number of transitions between firms A and B.

Figure 2.5, which differentiates by job-to-job and job-unemployment-job transitions, also fails to support the arguments of Woodcock (2015). If we follow his argument that improvements in the match only occur in job-to-job transitions, then we should see more blue than red points in the top right corner of the graph. Yet the patterns look remarkably similar for both types of transitions. Clearly, even for job-to-job transitions, firm pairs where workers gain wages by switching in either direction are the exception, not the rule.

In my view, Figure 2.5 supports the notion that a model which is additive in both worker and firm effects is a good starting point and can provide a good-but-not-great description of the data. Match effects may be important to some extent, but given that only a minority of firm pairs shows wage gains in both directions of job switches, the scope of match effects to explain the wage process appears limited.

2.4 Model

2.4.1 The Basic AKM Model

Suppose we observe N^* person-firm-year observations of N individuals working at J firms. The number of worker-firm pairings we observe in the data is M. The standard way of introducing worker and firm heterogeneity in wage setting is through the following simplified model of AKM:

$$Y = X\beta + U\gamma + D\alpha + F\psi + \epsilon \tag{2.1}$$

Here, Y denotes the stacked $N^* \times 1$ vector of log daily earnings that worker i earns at firm j in period t. X is an $N^* \times k$ matrix of k time-*varying* covariates, such as labor market experience, with associated $k \times 1$ parameter vector β, while U is the $N^* \times q$ vector of time-*invariant* covariates, such as gender, with associated $q \times 1$ parameter vector γ. Both X and U are worker-specific; the extension to firm-level covariates would be straightforward, but has been omitted to streamline the presentation. D is an $N^* \times N$ design matrix containing indicator variables for each worker, while F is the corresponding $N^* \times J$ design matrix of firms. The $N \times 1$ vector α and the $J \times 1$ vector ψ are known as the person and firm effects, respectively, and contain an additive time-invariant term for each worker and firm. α may reflect permanent differences such as general human capital, ability, or motivation, that are renumerated equally at every employer, while ψ reflects firm-specific differences in pay affecting all workers at all times equally, perhaps stemming from differences in productivity, workers' bargaining power, or product market competition. Finally, the $N^* \times 1$ vector ϵ reflects any left-out factors. It must be assumed to be serially uncorrelated, as well as uncorrelated with everything else on the right-hand side.

The basic AKM model can be estimated by OLS subject to some mild restrictions described below. However, it embeds a number of controversial assumptions. The most frequently discussed issue is the assumed uncorrelatedness between the error term ϵ and the matrix of firm identifiers D (Card, Heining, and Kline, 2013; Gruetter and Lalive, 2009). This assumption, which has sometimes been called "exogenous mobility" in the literature, rules out that any type of shock that impacts the level of wages also affects which worker works where. In particular, this assumption rules out that workers sort into firms on the basis of a worker-firm match effect in wages. To get some intuition about the type of bias introduced by match effects, suppose that that a firm screens applicants and will only hire workers if there are complementarities between the worker's skills and the firm's needs, which are reflected in the wage. Then, workers at such a firm will have high wages on average, not because the firm is a high-wage firm generally, but because the workers which select into the firm will earn high wages at this particular firm. The estimated firm effect will then be biased upwards.

A second issue is the restrictive functional form through which worker and firm heterogeneity are modeled. The fixed effects shift the intercept of the wage equation for each worker and firm, but by an amount that remains constant over time. Importantly, this assumption rules out different returns to experience for different workers. If the source of worker heterogeneity is unobserved differences in levels of human capital, then this functional form requires the returns to human capital to be constant over time. This is a questionable assumption, especially because the returns to *observed* levels of human capital (schooling) vary substantially and non-linearly over the life-cycle, as documented by Bhuller, Mogstad, and Salvanes (2017) and others. Likewise, the model does not allow for the firm effect to vary over time, thus failing to capture temporary wage shocks on the firm level.

The previous literature has done little to address these shortcomings. Even the simple AKM model requires $N + J$ degrees of freedom for estimation of the heterogeneity terms; modeling firm or worker heterogeneity in a more general way would require even more and might lead to severe precisions issues. Moreover, more complex forms of heterogeneity could give rise to new forms of bias. For example, the only more general form of modeling heterogeneity I am aware of can be found in the original AKM paper, where the authors introduce firm-specific terms for linear returns to tenure. For OLS to consistently estimate these terms, it is required that the length of tenure cannot be related to innovations in wages, which is at odds with basic theoretical models of job duration such as Jovanovic (1979). AKM do not discuss this issue in detail, but merely state that the firm-specific tenure terms are identified under comparable assumptions as in Topel (1991).[2] In any case, the subsequent literature has not pursued firm-specific returns to tenure any further. Overall, while the functional form with which worker and firm heterogeneity are modeled may not seem fully satisfactory, it is also difficult to see how this could be relaxed in a feasible and sensible way.

[2] I have not been able to verify this claim, and while it is debatable what "comparable" means, I am, in fact, fairly sure it is wrong. The central part of Topel's (1991) two-step identification strategy is a regression of residuals from a first regression on initial experience when starting a job. Initial experience is entirely absent from the list of covariates in AKM.

The basic AKM model as stated above cannot directly be estimated by OLS, for three reasons. First, even if X contains no intercept, there is a problem of perfect multicollinearity: one could add a constant to every firm effect, subtract the same constant from every person effect, and still end up with the same model. For this reason, I do include an intercept and normalize both person and firm effects to be mean-zero. This normalization makes effects easily interpretable: a person effect of 0.1, for example, means that this person earns 0.1 log-Euros per day more than the average person-year in the sample. Moreover, the presence of the U matrix with time-invariant covariates introduces further multicollinearity, so I normalize the person effects to be mean-zero for every group of persons formed by combinations of the elements of U. Formally,

$$E[\alpha|U] = E[\psi|U] = 0. \tag{2.2}$$

Second, Abowd, Creecy, and Kramarz (2002) point out that, even after normalizations, OLS can yield estimates of person and firm effects only for a connected set of firms in the graph-theoretic sense, where the connections are formed by workers moving from one firm to the next. In practice, this is not restrictive, as most datasets, including the data used for the present study, contain one large connected set that is composed of nearly all firms.

Third, there is a practical problem in trying to apply OLS that stems from the high dimensionality of the matrix Z, defined as

$$Z = \begin{pmatrix} X'X & X'U & X'D & X'F \\ U'X & U'U & U'D & U'F \\ D'X & D'U & D'D & D'F \\ F'X & F'U & F'D & F'F \end{pmatrix}. \tag{2.3}$$

The OLS estimator is defined as $\hat{\zeta} = (Z'Z)^{-1}Z'Y$, where $\zeta = [\beta, \gamma, \alpha, \psi]'$. Even with modern computing power, inverting the high-dimensional, non-sparse matrix $Z'Z$ is computationally infeasible. Abowd, Creecy, and Kramarz (2002) suggest using an iterative solver to minimize the expression $(Z'Z)\tilde{\zeta} - Z'Y$. With a suitable $\tilde{\zeta}$, this expression is nearly zero, and $\tilde{\zeta}$ should provide a good approximation to the OLS estimator $\hat{\zeta}$. Below, I will use a conjugate gradient algorithm to obtain an estimate of ζ.

2.4.2 Orthogonal Match Effects

The basic AKM model outlined above is not suitable to analyze whether workers gain by switching to higher-paying firms or finding better matches, since it contains no measure of matches. Match effects can be straightforwardly included by decomposing the error term as

$$\epsilon = W\phi + \eta, \tag{2.4}$$

where W is an $N^* \times M$ design matrix identifying each worker-firm match, ϕ is the associated $M \times 1$ vector of match effects, and η is an error term. Here, calculating ϕ is simply a matter of estimating the basic AKM model by OLS and then taking the within-match mean of the residuals. Plugging (2.4) into (2.1) and taking first differences, we then obtain

$$\Delta Y = \Delta X\beta + \Delta F\psi + \Delta W\phi + \Delta\eta. \tag{2.5}$$

The individual-specific terms $U\gamma$ and $D\alpha$ have dropped out, and equation (2.5) says that any change in the wage must come from the change in covariates, the change in the firm-fixed effects, the change in the match-specific effect, and the change in the residual. The change in the residual will be zero in expectation, and if X includes labor-market experience, $\Delta X\beta$ will indicate by how much the wage would have been expected to rise in the absence of job mobility. Therefore, the changes in ψ and ϕ indicate the extent to which a wage gain when changing employers can be attributed to finding a higher-paying firm and finding a better match, respectively.

There are three issues with this approach. First, as match effects are calculated from residuals, the cost of this procedure is that it assumes the match effect to be orthogonal to the covariates and the firm and person effects. It therefore does not allow workers to sort into firms on the basis of potential match effects. As noted above, Card, Heining, and Kline (2013) do not believe this to be a serious concern; they therefore resort to a model of orthogonal match effects.

A second problem is that this procedure restricts the duration-weigted average of the match effect for each person and each firm in the sample to be zero. Therefore, the model imposes that every worker and every firm are equally good at finding productive matches. This would be an attractive assumption if every individual held many jobs and every firm employed many workers – after all, the ability to consistently achieve good matches should reasonably be interpreted as being "of a high-wage type". But in practice, many individuals hold few jobs over the 12 years I observe them for, and as Table 2.2 showed, many Austrian firms are small. Consider a firm with just two employees, and paying high wages to both. Is this a high-wage firm, or just a firm which got lucky and obtained a good match twice in a row? In the above model, the mean match effect for every firm must be zero, so the high wage for both employees will be entirely apportioned to the firm-fixed effect. This is the intuition why this model understates the role of match effects and overstates the role of firm- and person-fixed effects in the case of few matches per worker or firm. Woodcock (2008) and Jackson (2013) discuss this bias in more detail.

Third, the model requires the residual η to be serially uncorrelated. If serial correlation is present in the data, then consecutive observations with higher-than-expected wages at a worker-firm pair, which has come about as the result of serially correlated shocks, will be interpreted as a large positive match effect by the model. As a result, the importance of the match effect will be overestimated. This is potentially serious, as the earnings dynamics literature (e.g. Meghir and Pistaferri (2004)) finds evidence of complex patterns of serial correlation in earnings data. That said, it is also difficult to address this problem convincingly: Card, Heining, and Kline (2013) try parametric corrections for serial correlation in η, and report the results to be unsatisfactory.

2.4.3 The Random Effects Model

In light of the limitations of the orthogonal match effects model, Woodcock (2008; 2015) has proposed a random effects alternative. It, too, takes as a starting point the log wage equation with linear person, firm, and match effects:

$$Y = X\beta + U\gamma + D\alpha + F\psi + W\phi + \eta \tag{2.6}$$

This model treats person, firm and match effects as random. In addition to spherical errors η, it makes use of the following assumptions:

$$E[\alpha|U] = E[\psi|U] = E[\phi|U] = 0 \tag{2.7}$$

$$Cov\left(\begin{array}{c|c} \alpha & \\ \psi & U \\ \phi & \end{array}\right) = \begin{pmatrix} \sigma_\alpha^2 I_N & 0 & 0 \\ 0 & \sigma_\psi^2 I_F & 0 \\ 0 & 0 & \sigma_\phi^2 I_M \end{pmatrix}. \tag{2.8}$$

The first assumption is simply an extension of (2.2). With any AKM-type model, we will be unable to see whether the higher wages paid to, say, more educated workers are due to a causal effect of education or because more educated workers are higher-wage types, so an assumption of this kind is unavoidable. More problematic is the second assumption, which states that worker, firm and match effects are all uncorrelated with one another. This is more restrictive than the orthogonal match effects model, which at least allowed for assortative matching of high-wage workers with either high- or low-wage firms. However, one assumption that the random effects model does *not* have to make is to assume uncorrelatedness between the random effects and the time-varying covariates X. This is less restrictive than the orthogonal match effects model, which had to assume that the match effect ϕ is uncorrelated with the time-varying covariates X.

Aside from this, the main advantage of Woodcock's random effects model for my purposes is the correction of the downward bias on the variance of the match effects outlined above for the orthogonal match effects model. To follow up on the above example, if a firm is seen paying two workers a high wage relative to what they earn elsewhere, it is not automatically assumed that the firm must have a high firm effect and the match effects are both small. Rather, the random effects estimation would assign some portion of the high wages to the firm effect, and the remainder to its having found good matches with the two workers. The relative importance of match versus firm effects will be chosen to minimize overall root mean squared error.

Like the basic AKM model, the random effects model as stated above does not incorporate serial correlation in the error term η. It is again possible to try parametric corrections for serial correlations. Woodcock (2015) reports the result of a variety of such specifications, some of which, as expected, do reduce the variance of the match effect somewhat. Again, it is questionable whether such parametric specifications are sufficiently rich to accomodate the complex serial correlation patterns encountered in earnings data.

Since it is difficult to say definitively whether the orthogonal match effects model or the random effects model should be preferred, I estimate both. Estimating the random effects model is a two-step procedure. First, I estimate the coefficient vector on the time-varying covariates β by a simple regression of wages on the covariates X and the design matrix for matches W. This assures that only within-match variation is used to identify β and leaves me with the problem of estimating

$$Y - X\hat{\beta} \equiv R = U\gamma + D\alpha + F\psi + W\phi + \eta. \tag{2.9}$$

Therefore, in the second step, I take the residuals R and estimate the parameters on the time-invariant parameters γ, the random effects α, ψ and ϕ as well as the variances σ^2_α, σ^2_ψ, σ^2_ϕ and σ^2_η using Restricted Maximum Likelihood (REML).[3] For this step, I use the algorithm

[3] REML is unfamiliar to many economists; Searle, Casella, and McCulloch (1992), Chapter 6, provide an introduction. To grasp the main idea, suppose that the random effects and the error term are all normally distributed in addition to being mutually uncorrelated. Then, we could first regress our outcome, R, on the covariates U and obtain residuals from this regression, and then apply maximum likelihood to these residuals

proposed by Searle, Casella, and McCulloch (1992, pp. 275–286) and implemented by Witkofsky (2012). Intuitively, this is an iterative algorithm that tries different values for the variances of the random effect components and the parameters on the covariates U and, as part of the algorithm, solves the following system of equations (Henderson et al., 1959):

$$\begin{pmatrix} U'U & U'D & U'F & U'W \\ D'U & D'D + (\hat{\sigma}^2_\eta/\hat{\sigma}^2_\alpha) & D'F & D'W \\ F'U & F'D & F'F + (\hat{\sigma}^2_\eta/\hat{\sigma}^2_\psi) & F'W \\ W'U & W'D & W'F & W'W + (\hat{\sigma}^2_\eta/\hat{\sigma}^2_\phi) \end{pmatrix} \begin{pmatrix} \hat{\eta} \\ \hat{\alpha} \\ \hat{\psi} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} U' \\ D' \\ F' \\ W' \end{pmatrix} R \tag{2.10}$$

Once convergence has been reached, solving the above system of equation yields estimates $(\hat{\eta}, \hat{\alpha}, \hat{\psi}, \hat{\phi})$ that are known as Best Linear Unbiased Predictors (BLUPs) in the statistical literature. They are best in the sense that, among all linear predictors of the random effects, these predictors assign person, firm and match effects so as to minimize overall mean squared error.

2.5 Previous Literature

As noted above, the approach of estimating wage equations containing an employee- and an employer-specific component was pioneered by AKM. Since then, the main methodological innovation has been the development of Woodcock's (2008; 2015) random effects model. More recently, Abowd, McKinney, and Schmutte (2019) have estimated a latent class model that incorporates match effects that are allowed to be arbitrarily correlated with person and firm effects, and does not need restrictive assumptions on worker mobility. However, their Bayesian estima-

in order to obtain estimates of the variances of the various random effects. The likelihood is derived not by considering the joint distribution of R, but rather the joint distribution of $K'R$, where the matrix K' contains row vectors k' that are orthogonal to U, i.e. $k'U = 0$. The solution to this maximization problem is invariant to the actual choice of K. Of course, normality of the random effects is a strong assumption, but Jiang (1996) shows consistency and asymptotic normality of the REML estimates even if normality is violated. This is similar in spirit to the familiar result from linear regression that the maximum likelihood estimator derived under normality is consistent and asymptotically normal even if the normality assumption does not hold.

tion method is computationally so intensive that it must be carried out on a small subsample of the data.

In terms of economics, much of the literature has been concerned with one of two questions, or both. The first question concerns the sign, size, and interpretation of the correlation between estimated person- and firm-fixed effects. AKM, Goux and Maurin (1999), Woodcock (2008), Gruetter and Lalive (2009), and Abowd, McKinney, and Schmutte (2019) report estimates of the correlation; Andrews et al. (2008) and Andrews et al. (2012) discuss statistical reasons why such a correlation could be downward-biased; Bagger, K. Sørensen, and Vejlin (2013) and Card, Heining, and Kline (2013) study trends in the correlation over time; and Lentz and Mortensen (2010), Eeckhout and Kircher (2011), and Lopes de Melo (2018) show that the correlation of wage effects need not correspond to a correlation in productivity. The second main strand of the literature has used AKM-type models in order to decompose wage differentials, notably the inter-industry wage differential (AKM, Goux and Maurin, 1999; Gruetter and Lalive, 2009; T. Sørensen and Vejlin, 2013) and the gender wage gap (Card, Cardoso, and Kline, 2016). I have seen only two estimates of the relative importance of firm and match effects for wage growth. T. Sørensen and Vejlin (2013) find that firm and match effects are of equal importance when estimated by fixed effects, but find a larger role for firm than for match effects when using random effects. Woodcock (2015) finds that firm effects are far more important regardless of whether estimation is carried out by fixed or random effects. Neither article asks whether there is heterogeneity by years of experience.

2.6 Results

2.6.1 Fixed Effects Estimation

I begin by presenting the results from estimating equation (2.1) by fixed effects. Having estimated the baseline model, I noticed an anomaly in the distribution of the estimated person and firm effects. As Figure 2.7 makes clear, there is a heap of observations with both very large person and very small firm effects, and vice versa. Table 2.5 demonstrates that these observations come disproportionally from small firms. Inspection of the data revealed that most of these anomalous observations come about in the following way: consider a firm with just a single employee, who held one other job before joining this one-person firm. Say that his wage at the one-person firm is about average, but for some reason, perhaps a recording error, his wage at the previous job is very high. Then, the model will assign this worker a high person effect to account for the high wage at the first job. However, in order to justify the average wage at the one-person firm despite the high person effect, the model must assign a very low firm effect to the firm. This explains the incidence of extreme observations with both a person effect on one end of the distribution, and a firm effect on the other. The problem is most acute for small firms (where a single employee has a large effect on the estimated firm effect) that have few connections to other firms through movers (such that only few movers identify the firm effect).[4]

Therefore, in addition to the baseline model, I have estimated the model using various samples and specifications; these are compared in Table 2.6. To overcome the problem of extreme estimated person and firm effect, I have tried two approaches. In column (2), I have restricted the sample to firms include only well-connected firms by deleting all firms with a core number smaller than 3.[5] Column (3) restricts the sample to firms with at least 10 employees. Column (4) retains firms with fewer than 10 employees in the sample but treats groups of such firms formed by a combination of 2-digit industry level and Austrian federal state as one single firm, with a common fixed effect, in the estimation. The remaining columns contain robustness checks. Column (5) estimates a model on the baseline sample but removes the age at labor market entry as a proxy for education. Column (6) restricts the sample to those individuals for whom we have non-missing uncensored earnings and part-time information. Column (7)

[4]Interestingly, Figure VIII of Card, Heining, and Kline (2013), who estimate a comparable AKM model for Germany, shows no such phenomenon. Perhaps this is because their figure is not as fine-grained as mine, or because of a greater number of very small firms in Austria relative to Germany.

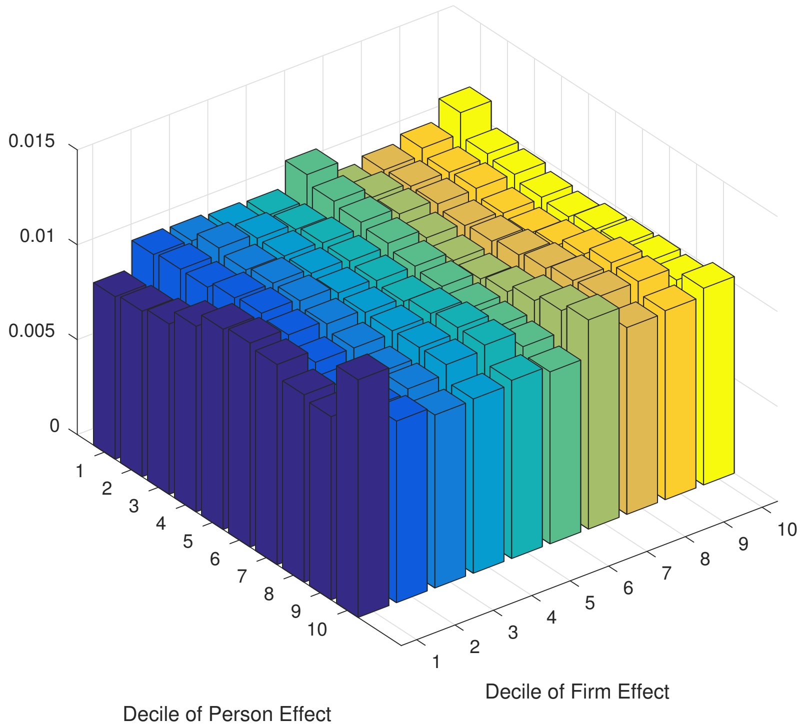
[5]In graph theory, the core number is a measure of the connectedness of a node. The degree is defined as the number of nodes a node is connected to via edges. A node has a core number of 3 or greater if, after removing all nodes that have a degree less than 3, the node still has a degree greater than 0. That is, after removing all firms that are poorly connected from the network, the remaining firms are still connected to others in the network.

Table 2.5: Outliers in Estimated Person and Firm Effects

	Percentage of Observations (Expected)	Average Firm Size
All observations	100% (100%)	929.04
Top 10% PE, bottom 10% FE	0.78% (1%)	419.14
Top 10% FE, bottom 10% PE	1.51% (1%)	526.66
Top 5% PE, bottom 5% FE	0.29% (0.25%)	318.30
Top 5% PE, bottom 5% FE	0.54% (0.25%)	291.59
Top 1% PE, bottom 1% FE	0.06% (0.01%)	8.50
Top 1% PE, bottom 1% FE	0.07% (0.01%)	9.06

"Expected" refers to the percentage we would expect to find if person and firm effects were independently distributed. Average firm size is calculated as the average over of all person-year-firm observations, rather than the average of all firms. This is why the numbers shown here differ from those in Table 2.2.

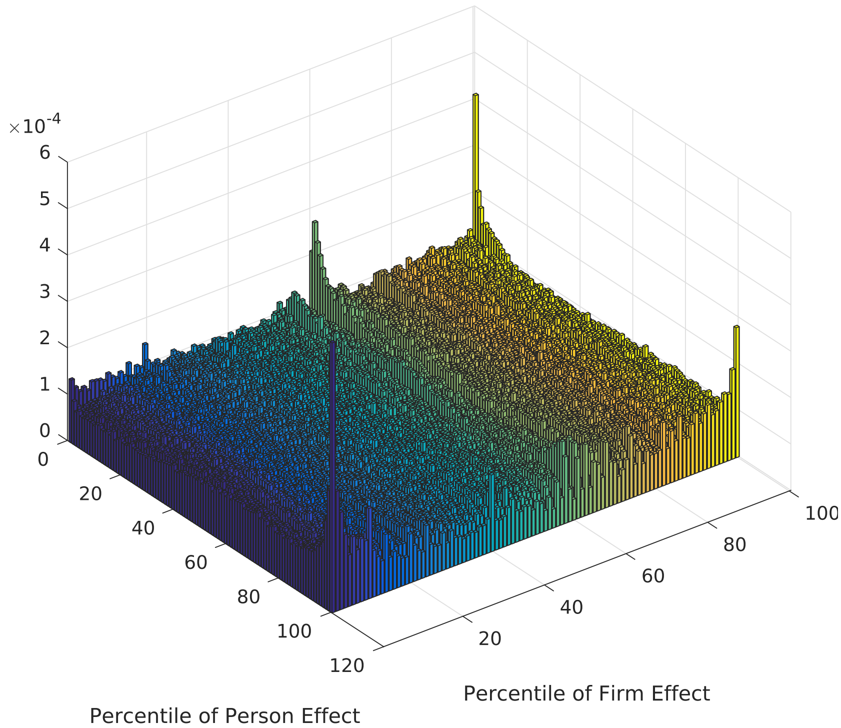
Figure 2.6: Joint Distribution of Person- and Firm Fixed Effects



Person and firm effects are estimated by fixed effects using the baseline specification.

contains the baseline sample, but earnings have not been trimmed from above and below. The sample used in column (8) does not use the uncensored earnings data from the tax data, but instead imputed censored wages using a series of tobit regressions, akin to what has been done in the previous literature when access to uncensored earnings data was not available. Finally, column (9) restricts the data to men, and also extends the observation window back to 1994, the first year for which we have uncensored earnings data.

Figure 2.7: Joint Distribution of Person- and Firm Fixed Effects



Person and firm effects are estimated by fixed effects using the baseline specification.

Table 2.6: Comparison of Different AKM Models and Samples

	(1)	(2)	(3)	(4)	(5)	(6)
Blue-Collar	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09
Part-Time	-0.27	-0.27	-0.27	-0.28	-0.27	-0.30
Female	-0.33	-0.32	-0.28	-0.37	-0.30	-0.36
18-21 at Labor-Market Entry	0.42	0.41	0.35	0.46		0.43
$>$ 21 at Labor-Market Entry	0.63	0.61	0.54	0.69		0.66
Female*18-21 at Labor-Market Entry	0.05	0.05	0.02	0.06		0.05
Female * $>$ 21 at Labor-Market Entry	0.08	0.08	0.04	0.09		0.09
Only Firms with Core Number >3		Yes				
Only Firms with ≥ 10 Employees			Yes			
Small Firms Aggregated by Industry-State				Yes		
Earnings and Part-Time Imputation						No
Trimming of Earnings						
Tobit Imputation						
Men Only						
Corr. with (1), Person Effects	1	1.00	0.98	0.98	0.94	0.98
Corr. with (1), Firm Effects	1	1.00	0.98	0.98	1.00	0.97
N^*	31,588,721	29,795,648	25,910,278	32,051,197	31,588,721	29,352,645
N	4,346,026	4,191,745	3,921,679	4,423,748	4,346,026	4,240,669
J	435,396	205,004	84,532	85,782	435,396	421,242
$RMSE$	0.22	0.22	0.21	0.23	0.22	0.21
Adjusted R^2	0.86	0.86	0.86	0.85	0.86	0.88
Perc. of Outliers, 5% Criterion	0.82	0.54	0.68	0.41	0.79	0.88
Perc. of Outliers, 1% Criterion	0.13	0.03	0.05	0.03	0.12	0.17
Match Effect Test Statistic	244,773	223,728	8,626	306,760	268,755	227,691
Productive Workforce Test Statistic	1,372	982	1,262	1,450	1,107	1,447
Years	2002–2012	2002–2012	2002–2012	2002–2012	2002–2012	2002–2012

Covariates also include a full set of experience dummies interacted with gender and the age at which workers enter the workforce, where applicable, and also interacted with a dummy indicating that the worker was born before 1957. Furthermore, the model contains a full set of cohort dummies indicating the first year in which we observe a worker in the ASSD. An observation is an outlier according to the 5% criterion if it is assigned a person effect in the top 5% of the distribution and a firm effect in the bottom 5% of the distribution, or vice versa. Likewise, an observation is an outlier according to the 1% criterion if it is assigned a person effect in the top 1% of the distribution and a firm effect in the bottom 1% of the distribution, or vice versa. If person and firm effects were distributed independently, we would expect the share of outliers by the 5% criterion to be 0.5% ($2 \times 5\% \times 5\%$) and by the 1% criterion to be 0.02% ($2 \times 1\% \times 1\%$). Under the null hypothesis of exogenous mobility, the match effect test statistic would be χ^2-distributed with 8,100 degrees of freedom, while the productive workforce test statistic would be χ^2-distributed with 810 degrees of freedom.

Table 2.6: Comparison of Different AKM Models and Samples (cont.)

	(1)	(7)	(8)	(9)
Blue-Collar	-0.09	-0.09	-0.09	-0.11
Part-Time	-0.27	-0.27	-0.30	
Female	-0.33	-0.33	-0.36	
18-21 at Labor-Market Entry	0.42	0.42	0.43	0.41
$>$ 21 at Labor-Market Entry	0.63	0.63	0.66	0.58
Female*18-21 at Labor-Market Entry	0.05	0.05	0.05	
Female * $>$ 21 at Labor-Market Entry	0.08	0.08	0.09	
Only Firms with Core Number $>$ 3				
Only Firms with $\geq$ 10 Employees				
Small Firms Aggregated by Industry-State				
Earnings and Part-Time Imputation			No	
Trimming of Earnings		No		
Tobit Imputation			Yes	
Men Only				Yes
Corr. with (1), Person Effects	1	1.00	0.98	0.91
Corr. with (1), Firm Effects	1	1.00	0.97	0.88
N^*	31,588,721	31,588,721	29,352,645	28,599,013
N	4,346,026	4,346,026	4,240,669	2,748,983
J	435,396	435,396	421,242	449,350
$RMSE$	0.22	0.23	0.22	0.23
Adjusted R^2	0.86	0.85	0.87	0.82
Perc. of Outliers, 5% Criterion	0.82	0.86	0.92	0.70
Perc. of Outliers, 1% Criterion	0.13	0.14	0.18	0.11
Match Effect Test Statistic	244,773	263,037	244,464	334,085
Productive Workforce Test Statistic	1,372	1,330	1,203	1,459
Years	2002–2012	2002–2012	2002–2012	1994–2012

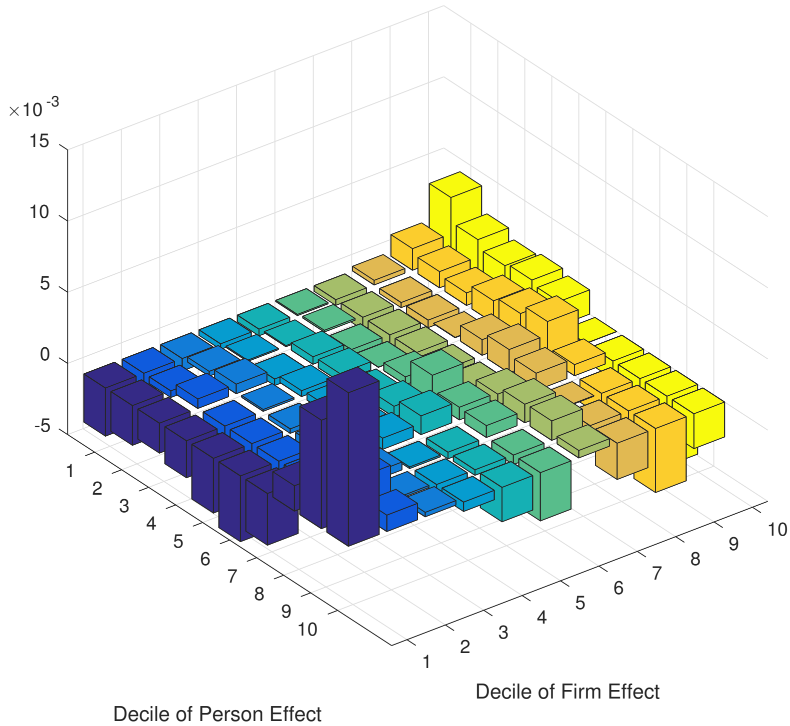
As the bottom panel of Table 2.6 shows, restricting the sample to larger or better-connected firms reduces the share of outliers somewhat, but they do not disappear entirely. However, grouping small firms by industry and state reduces the share of outliers to the levels expected under independent worker and firm effects. Otherwise, the top and middle panel of the table show a remarkable robustness of the estimated coefficients across specifications. In particular, the estimated firm and person fixed effects are virtually the same across specifications, as they are all nearly perfectly correlated with the estimated effects from the baseline model. The estimates in Table 2.6 do not contain standard errors, but I know of no computationally feasible way to estimate these.[6] But given the huge sample sizes involved, I would expect standard errors to be very small. Therefore, I conclude that the various specifications yield practically identical results, and I will report results only for the baseline model below.

One somewhat troubling result of Table 2.6 is the result of two specification tests devised by Abowd, McKinney, and Schmutte (2019). Intuitively, the first test asks whether the current match effect predicts the firm effect at a worker's future firm; the second test asks whether the average size of the person effects of a firm's workers is related to the average residual of its workers in the past. Both tests roundly reject the exogenous mobility assumption. While this does indicate that we should proceed with some caution, the utility of these tests is also quite limited: with the enormous sample sizes involved, *any* model will have difficulties passing a specification test. Ultimately, the tests tell uns nothing about whether the discrepancies between the model and the data are not just statistically but also economically significant. Indeed, as we saw earlier in section 2.3, there is good reason to believe that the additive fixed effects model fits the data reasonably well.

As further specification checks, Figures 2.8 and 2.9 reproduce Figures VI and VII from

[6]Even if one assumed homoskedasticity, a standard error estimate would consist of the diagonal elements of $(Z'Z)^{-1} \times s^2$, where s^2 is an estimate of the error term. Though I would only need its diagonal elements, I would first need to calculate the whole matrix $(Z'Z)^{-1} \times s^2$. This is a non-sparse matrix with over 4.7 million rows and columns, which is impossible for a computer to hold in memory. Bootstrapping is also not an option, because anytime one samples with replacement from the dataset, there are some observations that will not be sampled, and therefore the largest connected set on which the estimation can be carried out will be smaller than in the baseline sample. As a result, some of the person and firm fixed effects will not be identified.

Figure 2.8: Distribution of Residuals by Decile of Person- and Firm Fixed Effects

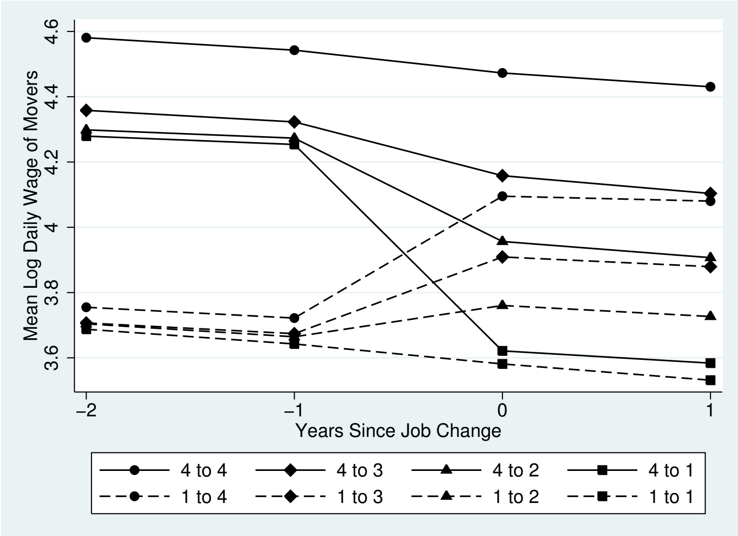


Residuals, person and firm effects are estimated by fixed effects using the baseline specification.

Card, Heining, and Kline (2013). Figure 2.8 graphs the mean residual for groups of observations formed by deciles of worker and firm effects. We might expect a failure of the assumption that wages are additively separable into a worker and a firm effect to result in anomalously large residuals for certain groups of worker-firm pairs. As in Figure VI in Card, Heining, and Kline (2013), there does seem to be a pattern of large resiudals (in absolute value) for certain cells involving small person- and firm-fixed effects, but the magnitude of these resiudals is relatively small. Second, Figure 2.9 draws an event study similar to Figure 2.4, but with firms grouped by the quartile of estimated firm effects, rather than coworker wages. Again, sorting on match effects could result in an asymmetry in wage gains and losses when switching from high-wage to low-wage firms and vice versa. The figure provides no evidence of such an asymmetry.

Moving now to the main purpose of the estimation, the extent to which workers improve firm effects and match effects over the course of their career, Figure 2.10 shows mean firm and

Figure 2.9: Mean Wages of Job Changes Grouped by Quartile of Firm Effects at Origin and Destination Firm

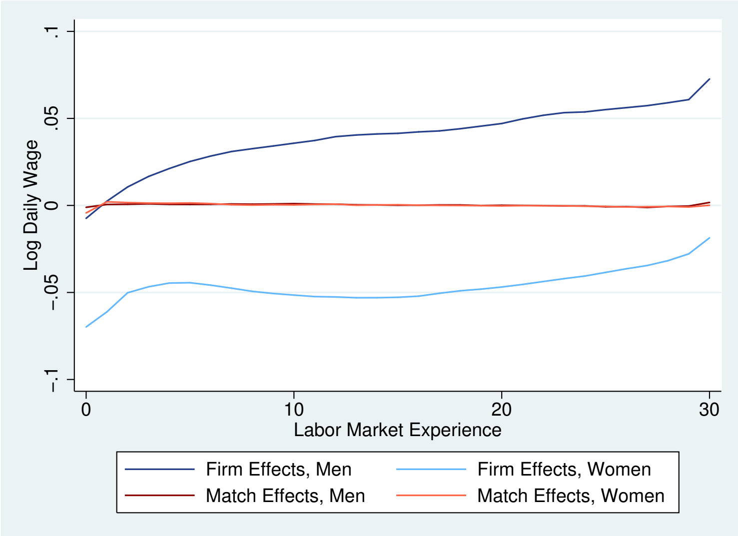


This figure shows the evolution of wages around job changes. Job changes are grouped according to the quartile of the origin and destination firm effects as estimated by fixed effects using the baseline sample. Job changes are included in the creation of this figure only if the origin and destination job last at least two years.

match effects by labor market experience, seperately by gender. Three features stand out. First, workers' average firm fixed effects increase markedly with labor market experience. A worker who has been on the labor market for 30 years will be found in a firm that pays between 5 and 10% more, on average, than a worker who is just starting out. This is true for both men and women, but slightly more so for men. Second, the figure shows a remarkable segregation of men and women by workplace. Women start out in establishments that pay an average of about 5% less than men, and never catch up. Third, the mean match effect is near zero for every level of experience. Taken together, Figure 2.10 suggests that it is firm effects, not match effects, that account for improvements in wages over time.

Of course, the changing firm effects with experience could come about as a result of genuine improvements of firm effects by job switchers, or because of composition effects as some workers drop out of the labor force over time. To account for this, Figure 2.11 and Table 2.7 decompose the year-to-year changes in the wage over time. Improvements in the firm effect do contribute to wage growth somewhat during the first five years on the labor market, whereas the contribution

Figure 2.10: Mean Firm and Match Effects by Experience



Firm and match effects are estimated by fixed effects using the baseline specification.

from improvements in match effects stays negligible throughout. The pattern looks similar for men and women; even though Figure 2.10 had demonstrated that they start out at different kinds of firms, the importance of firm and match effect *improvements* over the life-cycle is similar for both genders.[7] Therefore, in the following, I no longer provide separate analyses by gender. Table 2.7 demonstrates that improving firm fixed effects contribute between 5 and 10% to overall wage growth during the first five years on the labor market, but this contribution fades out thereafter.

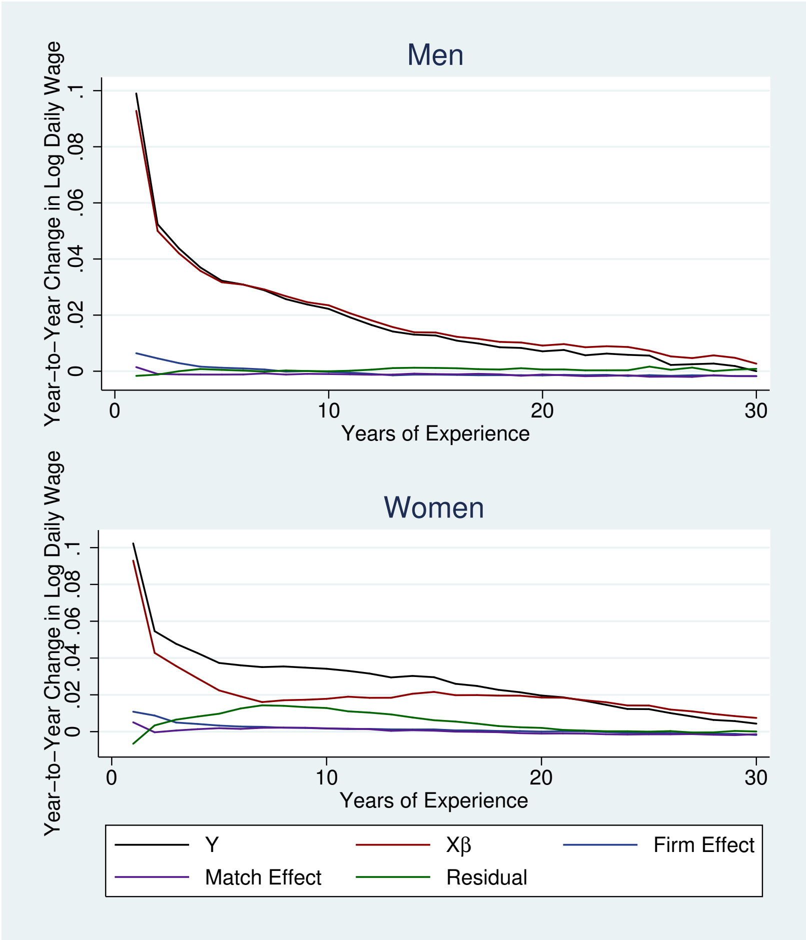
The above results are consistent with an explanation where young workers who are stuck in a low-wage firm make a conscious effort to improve their lot by switching jobs into a higher-paying firm. We can bring two more pieces of evidence to bear on this story. First, Figure 2.12 displays probabilities of moving away from a firm conditional on the current firm or match effect. As expected, workers working in a firm with a low firm fixed effect are more likely to move to another

[7] Table 2.7 indicates that the change in *residuals* also contributes to wage growth, which is unexpected. As Table 2.9 shows below, this pattern repeats itself for the random effects estimation. I have not been able to explain this phenomenon. It is worth noting that, as seen in Figure 2.11, the finding is entirely driven by the residuals for women. By contrast, the change in average residuals for men is close to zero, as one would have expected.

Table 2.7: Decomposition of Wage Growth, Fixed Effects Estimation

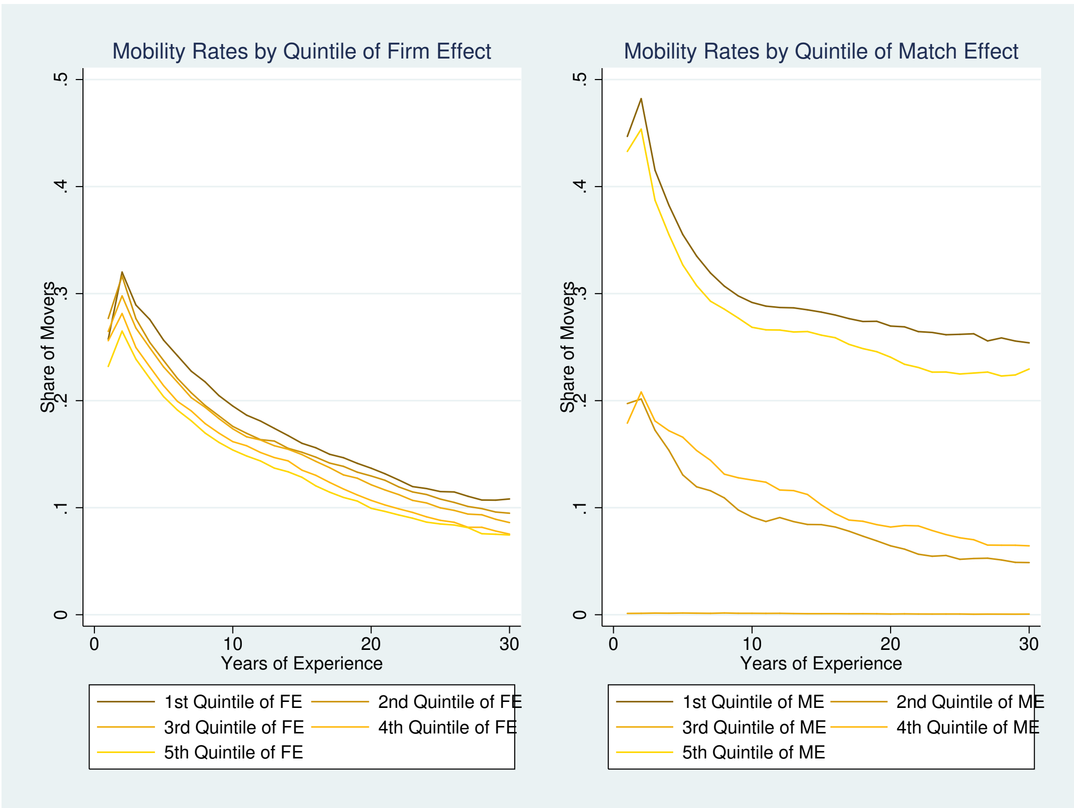
Experience	Total Wage Growth, in Logs	Contributions to Wage Growth, in %			
		X	Firms	Matches	Residual
1	0.10	92.44	8.32	3.03	-3.80
2	0.05	87.73	12.02	-1.30	1.55
3	0.05	86.01	8.35	-0.76	6.40
4	0.04	82.68	7.01	-0.12	10.44
5	0.03	79.55	6.28	0.58	13.58
6	0.03	76.56	5.44	0.20	17.80
7	0.03	72.59	4.87	1.77	20.77
8	0.03	73.13	3.13	1.35	22.39
9	0.03	72.50	3.44	1.87	22.20
10	0.03	73.80	2.51	1.30	22.40
11	0.03	76.20	1.66	0.87	21.28
12	0.02	76.35	1.04	0.14	22.46
13	0.02	78.89	-0.74	-1.85	23.71
14	0.02	80.06	-0.21	-0.26	20.41
15	0.02	84.00	-0.11	-1.21	17.32
16	0.02	87.28	-1.85	-3.05	17.62
17	0.02	90.79	-2.23	-3.07	14.51
18	0.02	96.53	-3.50	-4.61	11.57
19	0.01	100.75	-3.97	-8.44	11.66
20	0.01	104.31	-5.81	-8.49	9.99
Total	0.62	83.61	2.28	-1.10	15.21

Figure 2.11: Decomposition of Wage Changes



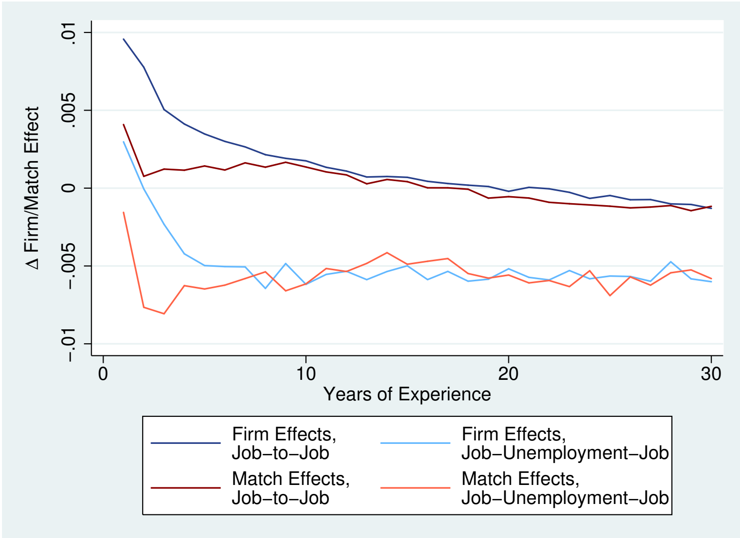
firm than workers working at a high-firm-effect firm. The right-hand panel, which shows mobility as a function of the match effect, is distorted by the fact that almost 40% of workers never move, and therefore their match effect is normalized to 0. This lowers the observed mobility rates for the three middle quintiles of the match effect distribution. The important feature of the right panel in Figure 2.12 is that, at least during the first ten years of worker experience, mobility rates are only slightly higher for workers in the top than in the bottom quintile of the match effects distribution. This underscores that moving to better firm effects, not match effects, is the main driver of worker mobility. Second, to the extent that job-to-job transitions can serve as a proxy for voluntary transitions, we would expect workers making such transitions to improve their firm effect by more than workers who experience an unemployment spell. Figure 2.13 confirms this intuition; whereas job-to-job transitions result in improvements of average firm effects at least during the first ten years or so on the labor market, job-unemployment-job transitions result in worse average firm effects for all but the least experienced workers. As for match effects, we do not see any visible improvements resulting from job-to-job transitions. However, there are deteriorations in the average match effect for job-unemployment-job transitions. Thus, Figure 2.13 provides support for half of Woodcock's (2008) argument: while voluntary job transitions do not generally improve match effects, matches may suffer from involuntary job separations.

Figure 2.12: Mobility Rates by Quintile of Firm and Match Effect



In conclusion, the results from the fixed effects estimation overwhelmingly support firm effects, not match effects, as a driver of mobility and wage growth. However, as noted earlier, fixed effects estimation may have understated the role of match effects. I therefore turn to

Figure 2.13: Change in Firm and Match Effects, by Years of Experience and Type of Transition



estimation by random effects to check the robustness of my findings.

2.6.2 Random Effects Estimation

Estimation of the variances of person, firm and match effects turned out to be computationally very demanding. I therefore had to split the sample into two periods, 2002–2006 and 2007–2012, and estimate the model separately for both periods. Table 2.8 compares the estimation results with those from the fixed effect estimations for the baseline sample and specification. Reassuringly, the parameter estimates on the covariates are quite similar across estimation techniques, as well as across time periods. However, the estimated variances of person, firm and match effects are quite different. As expected, I find evidence that the fixed effects estimator understates the variance of match effects, as the random effects estimator finds an estimate that is more than twice as high. What is unexpected, however, is that the variance of the firm effect is also twice as high when estimated by random rather than fixed effects. By contrast, the estimated variance of the person effect is found to be slightly lower by random effects estimation.

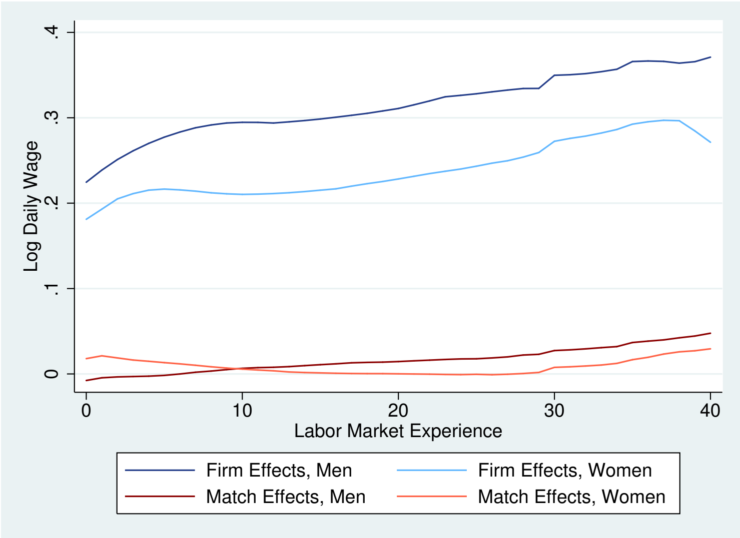
Table 2.8: Comparison of Estimation Results by Random and Fixed Effects

	Fixed Effects			Random Effects (2002–2006)			Random Effects (2007–2012)		
Coefficient Estimates									
Blue-Collar		-0.09			-0.07			-0.07	
Part-Time		-0.33			-0.24			-0.24	
Female		-0.27			-0.34			-0.35	
18-21 at Labor-Market Entry		0.42			0.35			0.35	
$>$ 21 at Labor-Market Entry		0.63			0.53			0.51	
Female*18-21 at Labor-Market Entry		0.05			0.06			0.07	
Female * $>$ 21 at Labor-Market Entry		0.08			0.08			0.12	
Observation Numbers									
N^*		31,588,721			13,960,397			18,093,960	
N		4,346,026			3,496,915			3,910,905	
J		435,396			359,150			376,258	
Variance and Covariance Estimates	PE	FE	ME	PE	FE	ME	PE	FE	ME
Person Effects	0.1384	0.0021	-0.0000	0.0818	0.0000	0.0000	0.0760	0.0000	0.0000
Firm Effects		0.0522	-0.0000		0.1349	0.0000		0.1366	0.0000
Match Effects			0.0168			0.0464			0.0524

Covariates also include a full set of experience dummies interacted with gender and the age at which workers enter the workforce.

Figure 2.14 is the random effects counterpart to Figure 2.10.[8] Both figures tell a similar story: men tend to have greater firm effects than women, but both improve theirs rapidly during the first few years on the labor market. After about five years for women, and ten for men, the speed of improvement slows down, though the mean firm effect continues to increase steadily. By contrast, average match effects remain near zero throughout, for both genders.

Figure 2.14: Mean Firm and Match Effects by Experience



Firm and match effects are estimated by random effects using the baseline specification.

Table 2.9 decomposes wage gains at job changes analogously to Table 2.7 previously. As expected from Figure 2.14, firm effects do contribute to wage growth, particularly during the first few years on the labor market, but their impact gradually fades out. Compared to the results from fixed effects estimation, the impact of firm effects is somewhat larger, and the effect does not fade out to zero nearly as fast as had been the case for the fixed effects estimation. There is some contribution to wage growth from match effects after five years on the labor market, but the importance of match effects remains small compared to firm effects throughout. In sum, the findings from random effects estimation underscore the importance of firm effects,

[8] The mean firm effect shown in this figure is about 0.2, which would seem to contradict assumption (2.7), where firm effects were normalized to zero. The reason is that the mean *firm* has a zero firm effect, but Figure 2.14 shows mean *person-firm years*. Larger firms, which receive more weight in the construction of Figure 2.14, have a higher firm effect on average.

Table 2.9: Decomposition of Wage Growth, Random Effects Estimation

Experience	Total Wage Growth, in Logs	Contributions to Wage Growth, in %			
		X	Firms	Matches	Residual
1	0.10	95.91	10.79	1.61	-4.59
2	0.05	97.33	17.91	-3.41	-5.51
3	0.05	92.39	16.20	-2.28	1.79
4	0.04	88.25	15.61	-0.60	6.40
5	0.03	83.61	15.56	0.69	10.51
6	0.03	79.20	15.09	1.70	13.79
7	0.03	75.68	15.05	3.68	16.65
8	0.03	75.50	12.61	3.30	19.02
9	0.03	75.81	12.44	3.93	18.05
10	0.03	76.44	11.70	4.29	18.52
11	0.03	79.05	10.93	3.75	17.42
12	0.02	81.64	10.98	3.15	16.54
13	0.02	84.15	10.05	1.74	17.99
14	0.02	86.59	10.21	3.06	13.72
15	0.02	89.16	11.40	2.80	12.41
16	0.02	92.69	9.69	1.64	14.88
17	0.02	96.09	11.64	1.70	10.37
18	0.02	103.13	8.07	-0.30	5.00
19	0.01	107.96	8.84	-4.06	1.86
20	0.01	110.18	7.80	-3.42	1.38

not match effects, as a driver of wage growth.

2.7 Conclusion

I estimate models with additive worker, firm and match effects to study the contribution of improvements in these effects to overall wage growth. Improvements in firm effects contribute between 5 and 15 percent of overall wage growth during the first years on the labor market, but the effect fades out over time. By contrast, workers seem unable to improve idiosyncratic matches to their firms over the course of their careers.

For researchers in labor economics, the results underscore the findings of Card, Heining, and Kline (2013) that a model with additive worker and firm effects serves as a good first-order approximation to the distribution of wages. Moreover, they emphasize the importance of including firm heterogeneity in pay into theoretical models of the labor market. Finally, my findings clarify the channel through which job mobility helps young workers improve their labor market prospects.

Chapter 3

Female Leadership in Protestant Churches, Religiosity and Market Outcomes[1]

3.1 Introduction

A large body of literature in labor economics has documented a persistent glass ceiling for women in management positions that is only slowly eroding (see Blau and Kahn (2017) for an overview). Where female represenation in leadership positions does increase, it is natural to ask whether women will display a different leadership style that will affect firm performance. Such effects could be expected because there is ample evidence of meaningful gender differences in terms of preferences, such as risk aversion (Croson and Gneezy, 2009), forward-lookingness (Silverman, 2003) or competitiveness (Buser, Niederle, and Oosterbeek, 2014). Furthermore, the preferences of managers manifest themselves in distinct management styles that yield meaningfully different outcomes at the firm level (Bertrand and Schoar, 2003).

[1]This chapter is joint work with Anna Raute, School of Economics and Finance, Queen Mary University, Mile End Road, London E1 4NS, UK. a.raute@qmul.ac.uk.

Empirical evidence of the effect of female managers on firm productivity has been difficult to come by, however, because performance measures for individual managers are elusive. The previous literature has largely focused on top executives or board members and proxied their performance with firm-wide outcomes. Results have been mixed: whereas Adams and Ferreira (2009), Matsa and Miller (2013) and Weber and Zulehner (2010) find evidence of a distinctly female style of leadership, Wolfers (2006), Gagliarducci and Paserman (2015) and N. Smith, V. Smith, and Verner (2006) find no evidence of meaningful gender effects. As noted by Wolfers (2006), given the many determinants of firm performance and the fact that management decisions may need time to take effect, tests based on firm-wide performance are hampered by a lack of statistical power. Moreover, this body of research is naturally confined to the study of the very top of the corporate hierarchy or very small firms.

We contribute to the literature by studying the productivity of a particular type of *mid-level* managers in a large organization: men and women working as pastors in the Evangelical Church in the Rhineland, Germany. In a novel dataset, we observe several measures of output on the parish level, including church membership, mass attendance, and donations, as well as the number of masses, volunteers, communions, baptisms and confirmations performed, all over a 26-year time period. Since each parish is staffed with at most a few pastors (often just a single one), this dataset offers a unique opportunity to study the performance of workers below the top management tier.

Our identification strategy exploits pastors moving across parishes to isolate pastor gender effects from parish effects. A natural concern with this strategy is that mobility may be endogenous, for example, if women were allocated to struggling parishes more frequently than men. However, an event study that shows no systematic differential trends outcomes for parishes before a new male or female pastor moves in. Moreover, Cox proportional hazard estimations indicate that, while women move across parishes more frequently than men, this correlation is entirely explained by the type of service, as women are more likely to work part-time or still be in training, which have shorter spell durations. These results give us comfort that our identifying assumption of exogenous mobility is valid.

After accounting for composition effects, we find no significant gender effects across a wide range of outcomes, with one notable exception: female pastors raise approximately 7.5% fewer donations than their male counterparts. This effect is highly statistically significant and appears both for donations for the pastor's own parish and for other causes. There is some suggestive evidence that the effects are strongest for the youngest and oldest cohorts of pastors. Gender effects appear strongest at the top of the distribution of donations, i.e. the most successful male fundraisers outperform the most successful female fundraisers by a wider margin than at the median. We think there are two – not mutually exclusive – possible explanations for our findings: differences in skills or preferences between male and female pastors, or discrimination against female pastors by parish members.

3.2 Institutional Setting and Data

We use administrative data from the Evangelical[2] Church in the Rhineland ("EKIR"), the second-largest regional protestant church in Germany with 2.6 million members as of 2015. Our analysis combines several different datasets, all of which cover the time period from 1990 until 2015. Our principal dataset records all changes to a pastor job assignment on a daily basis. The main possible assignment changes are entry into or exit from the clergy, a change of position, a change in hours from full-time to part-time or vice versa, beginning or end of temporary leave, or death. In all cases, the dataset contains an identifier for the pastor, identifiers for the outgoing and incoming parish, and information on the positions that the pastor held before and after the change. We also observe the pastor's gender and year of birth.

We have supplemented our dataset of pastor transitions with a second dataset which records each pastor's parish on an annual basis. This second dataset allows us to identify those pastors

[2] Unlike the United States, where a multitude of protestant denominations exist side-by-side, German protestantism is dominated by regional churches which comprise the vast majority of practicing protestants in their respective territories. Theologically, these churches are mostly liberal; the term "evangelical" does not have the same connotation of ideological rigidity in German as in English.

who never experience any change in status during the entire 26-year period from 1990 until 2015. Additionally, we have used this supplementary dataset to correct apparent errors in the our primary dataset, which result in contradictions between consecutive assignment changes.[3] From these sources, we have constructed a monthly panel dataset that records each pastor's assignment over time.

Table 3.1: Descriptive statistics of the sample size.

	Male	Female	Total
Number of pastors	3,042	1,374	4,417
Number of pastors who ever serve in a parish	2,474	1,085	3,559
Number of pastor-years in parishes	30,520	9,845	40,365
Number of parishes			939
Number of parish-years			20,841

Notes: Pastor-years and parish-years counted as of January each year.

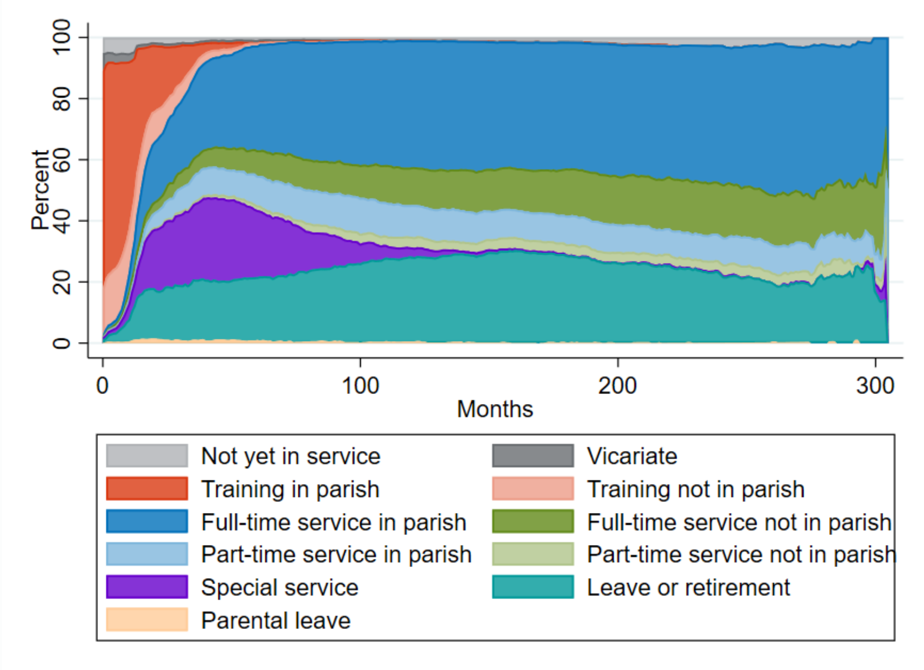
Our dataset contains information on more than 4,400 pastors, although roughly one-fifth of them never serve in a parish. The remaining 3,559 pastors account for 41,365 pastor-years during which they serve in a parish, for an average of nearly 12 years per pastor. Women make up just over 30% of pastors, but fewer than one quarter of pastor-years. The sample contains 939 unique parishes[4] that account for 20,841 parish-years. The number of parish-years is only half of the number of pastor-years, because it is not uncommon to staff parishes with two or even more pastors.

Figures 3.1 and 3.2 depict the evolution of pastoral careers, separately for men and women. After graduation from university, a pastoral career begins with the vicariate, a two-year period of practical training. Towards the end of the vicariate, the candidates receive ordination, at which point they are authorized to perform all functions of a pastor. Following the vicariate, pastors enter another period of service while training ("Probedienst"), during which they work

[3]For example, if we observe a pastor going on leave in year T, but then observe the same pastor leaving a parish from active service in year T+3, it must be the case that the pastor returned from leave to active service in years T+1 or T+2, and the return failed to be recorded in the transition dataset. In these cases, we use the supplementary dataset to infer whether the pastor returned from leave to active service in year T+1 or year T+2.

[4]For the case where two parishes merge to form a third, this figure counts all three parishes separately. If, instead, all three parishes connected by a merge were counted as one single parish, the figure would come out to 756 unique parishes.

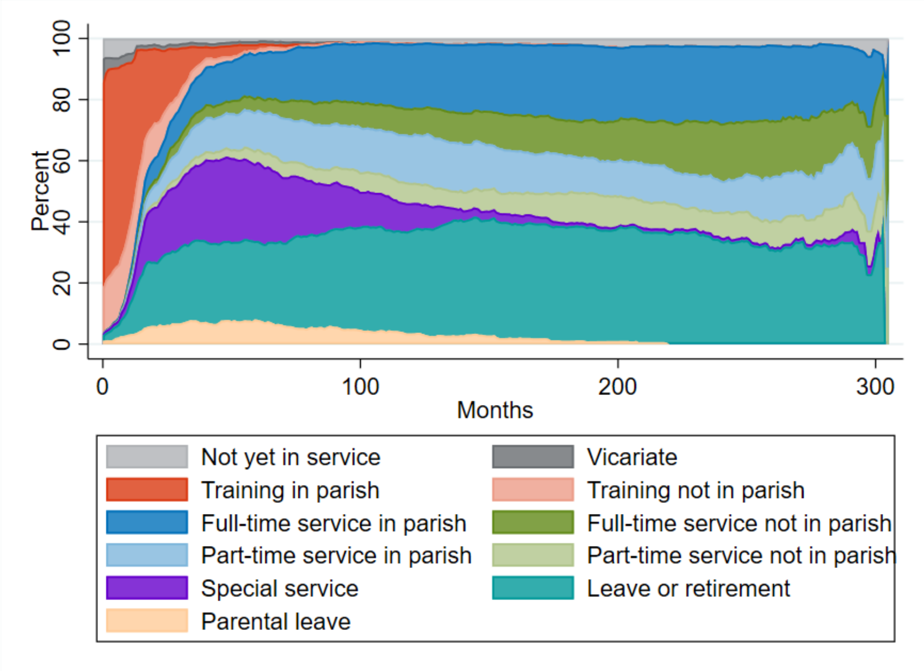
Figure 3.1: Shares of male pastors by job category and month since ordination.



as a regular pastor, but receive continued supervision and mentoring. At the conclusion of their training service, pastors may join regular church service. Both regular and training services mostly take place in parishes, but there are exceptions, such as services in hospitals, prisons, schools, or the church administration.

The assignment process of pastors to jobs is as follows. For the training service, pastors are assigned placements by the central church administration. Upon completion of the training service, pastors cannot generally stay on in the same parish, and will need to obtain another position. There are two assignment mechanisms. Two-thirds of open positions are posted publicly, and any ordained pastor from the EKIR may apply. The remaining third of positions is assigned centrally by the church, taking into account the requirements of the particular job and the qualifications and personal circumstances of the candidates. Once a pastor occupies her first position after completing training service, she will be promoted into the rank of church official, meaning, among other things, that she can no longer be fired except under extraordinary

Figure 3.2: Shares of female pastors by job category and month since ordination.



circumstances. In principle, the pastor may keep her position for as long as she wishes, although in recent years, the church has encouraged pastors to change jobs every 10 years or so.

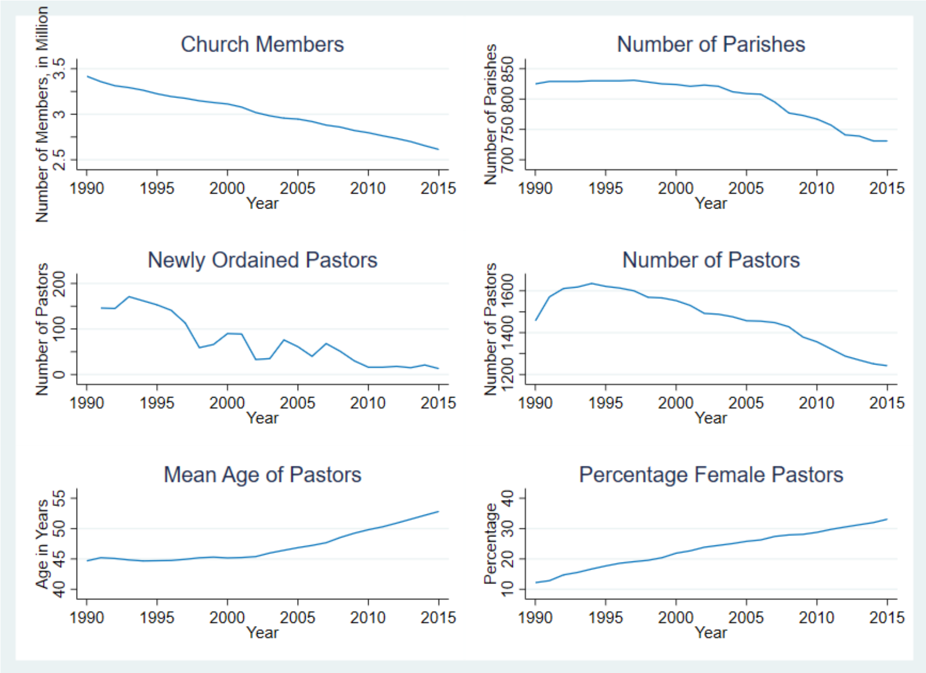
A notable feature of figures 3.1 and 3.2 is the significant number of pastors who do not find an open position after conclusion of their training service. We understand from conversations with EKIR staff that there was a boom in the number of students who pursued divinity studies in the 1980s, leading to a glut of newly trained pastors with no open positions available to them in the early 1990s. In response, the church created term-limited “special service” assignments outside of parishes, in order to place at least some of the newly trained pastors until they found a permanent position. However, even this option was not available to all pastors, and many pastors would be placed on leave indefinitely and eventually exit the clergy.

It is also remarkable that a non-negligible fraction of male pastors works part-time, whereas part-time work is predominantly performed by women in the German general population. The likely reason is that EKIR allows full-time positions to be split into two part-time positions, which is attractive to couples where both partners are pastors and wish to work in the same parish. By contrast, very few male pastors ever go on parental leave.

Figure 3.3 displays some trends in the data. Church membership has declined steadily from approximately 3.4 million members in 1990 to 2.6 million in 2015. As a consequence of declining membership, EKIR has begun to merge parishes, a process that has accelerated in recent years; the number of parishes has declined from 825 in 1990 to 731 in 2015. As the boom in divinity studies from the 1980s has subsided and the demand for new pastors has tapered off, the number of ordinations has declined precipitously, from a maximum of 171 ordinations in 1993 to just 13 in 2015. The result has been a decline in the total number of pastors from over 1,600 in the early 1990s to fewer than 1,300 in 2015. Correspondingly, the average age of pastors has increased from 45 to 53 years between 1990 and 2015. Finally, and of particular interest for this paper, the percentage of women among pastors was just 17% in 1990, as the church allowed the first ordinations of women only in the 1970s. While the share of women has increased steadily over time, women have remained the minority: in 2015, fewer than 40% of pastors were women.

In a final step, we have merged the data on individual pastors with a rich dataset of outcomes at the parish level. Table 3.2 presents summary statistics for the most relevant outcomes contained in the data. We observe the stock and flows of members in and out of the church via baptisms, entries, exits, or deaths. As measures of activity in the parish, we observe the number of confirmations, marriages, funerals, masses held, youth club meetings, and the number of volunteers from the parish. Four times per year, the parishes are required to record church attendance at the gates: at Christmas, Good Friday, the first sunday in Advent, and the first sunday of Lent. The church has chosen the latter date because it is an ordinary Sunday, and the church views attendance on this day as broadly representative of average attendance. We also have measures of the amounts of donations raised, both for the own parish and for other causes. Finally, we observe some outcomes specific to female members of the parish: the share of women among all entries, exits, volunteers, and presbyters, an elected committee of parish members that leads the parish. All data are available annually, with the exception of the fundraising

Figure 3.3: Trends in church membership, parishes, and pastors, 1990–2015.



data, which are gathered biannually. The data are self-reported by the parishes, but there are no rewards or penalties attached to performance, and in conversations, church officials have expressed confidence that parishes would have no reason to strategically tamper with the data.

3.3 Pastor Mobility Across Parishes

Since our identification strategies will rely on mobility of pastors in and out of parishes to distinguish pastor from parish effects, we begin with a descriptive analysis of the way pastors move across parishes in our data. Table 3.3 displays some descriptive statistics of spells. While we observe multiple spells for most pastors in our sample – the average number of spells per pastor is 3.8 – fewer than half of these spells actually reflect service in a parish. Instead, a large portion of spells represents leaves or service outside of parishes, e.g. in schools, prisons, hospitals, or administrative positions. On average, a pastor in our dataset will hold just a single regular full-time service position in a parish and serve just 1.09 parishes over the period under

Table 3.2: Descriptive statistics for the parishes in our sample.

	Mean	SD	Min	Max
Number of parish members	4,164.83	3,176.93	142	32,604
Parish member exits	29.74	34.58	0	519
Adult parish member entries	9.05	9.46	0	139
Adult baptisms	2.83	3.84	0	121
Child baptisms	32.93	26.18	0	266
Confirmation candidates	36.89	29.48	0	1,115
Marriages	9.31	8.75	0	84
Funerals	49.47	40.03	0	351
Masses	166.51	102.07	32	1,055
Children's masses	37.62	37.65	0	1,112
Youth club meetings	6.30	6.27	0	77
Volunteers	138.79	115.80	0	1,449
Attendance last Sunday in Lent	109.19	97.03	0	3,886
Children's service attendance last Sunday in Lent	23.59	30.40	0	900
Attendance Good Friday	156.70	127.90	0	2,000
Attendance first Sunday in Advent	161.40	137.43	0	1,800
Attendance Christmas Eve	1,063.90	847.71	0	8,978
Funds raised for the parish, in 2010 Euros	5,295.69	5,342.54	0	99,712
Funds raised for other causes, in 2010 Euros	15,509.20	12,555.44	0	165,980
Female share of exits	0.44	0.17	0	1
Female share of entries	0.58	0.51	0	49
Share female volunteers	0.71	0.11	0	2
Share female presbyters	0.50	0.15	0	1

Notes: All monetary amounts in 2010 Euros. All observations recorded annually from 1990 until 2015, except eucharists, which were not recorded in 1992, and donations, which were only recorded in even years.

study. In other words, the case of a pastor switching from one full-time position to another in a different parish is relatively atypical.

Table 3.3 also reveals some interesting splits by gender. Women are much more likely to shuttle between full- and part-time work and to go on parental leave (where they stay over twice as long) compared to men. Perhaps surprisingly, women are also much more likely to go on other types of leave and to serve outside of parishes. As a result, they have many more spells – an average of 5.25 per pastor, as opposed to 3.14 for men. However, the number of parishes in which men and women serve is nearly identical, at 1.08 for men and 1.12 for women. The average spell duration for a full-time regular service in a parish is also much longer for men (almost 3,000 days, or more than eight years) than women (1,767 days, or under five years).

These splits might reflect inherent gender differences in mobility, but to an extent they

Table 3.3: Descriptive statistics for service spells.

		Average number of spells			Average spell duration in days		
		All	Men	Women	All	Men	Women
Regular service in parish	Full-time	1.00	1.06	0.84	2,655	2,968	1,767
	Part-time	0.33	0.24	0.54	1,483	1,446	1,519
Training service in parish	Full-time	0.50	0.37	0.80	570	600	538
	Part-time	0.01	0.01	0.03	368	328	391
Other service outside of parish		1.28	0.98	1.96	1,252	1,494	980
Parental leave		0.13	0.06	0.28	544	311	647
Other leave		0.54	0.43	0.79	2,655	2,778	2,506
All spells		3.79	3.14	5.25	1,724	2,037	1,304
All spells in parishes		1.84	1.68	2.21	1,861	2,223	1,244
Total number of parishes served		1.09	1.08	1.12			

might also be simply an artifact of sample selection. As noted above, the percentage of female pastors increases over time, so women are more highly represented among younger cohorts. Since mobility is more common early in a pastor's career, when they are likely to move at the conclusion of their training service or because they may have accepted temporary special assignments, a simple comparison by gender will necessarily be skewed towards showing higher mobility for women. Similarly, women are better represented in the later years of our sample; as noted earlier, the church has recently begun to encourage pastors to move across parishes at the time. Finally, the left- and right-censoring of our data distorts simple average comparisons of spell durations.

We have investigated this issue further by fitting some simple Cox proportional hazard models to the data. Table 3.4 displays the estimated hazard ratios. Column (1) confirms the greater mobility of women in the sample: across all types of spells, female pastors have a hazard rate that is approximately 33% higher than that of men. For column (2), we have added control variables for the pastor's age decile at the start of the spell. While there are sizable effects of age on mobility – the estimated relationship is U-shaped, with mobility first declining and then increasing with age – the estimated hazard ratio for pastor gender is practically unchanged. The same goes for column (3), where we control for cohort effects by adding a fixed effect for the five-year window in which the spell started. Even though mobility appears to increase gradually over time, controlling for cohort effects does not materially affect the estimated hazard ratio for female versus male pastors. By contrast, when we control for spell type in column (4), the hazard ratio for female pastors falls to near one. Part-time spells, parental and other leaves have much higher hazard ratios than other types of spells, and as we saw in table 3.3, these are also the kind of spells in which women are overrepresented.

Columns (5) through (8) of table 3.4 restrict the sample to only those spells where a pastor serves in a parish, which are the main interest of this paper. The main findings are the same as in columns (1) through (4), but the contrast between the unconditional hazard ratio for gender, which now stands at 1.45, and the hazard ratio of only 1.10 after controlling for spell type, is stronger still. It is also worth noting that the hazard ratio is no longer statistically significantly different from 1 once we control for spell type. Therefore, the seemingly greater mobility of female pastors is almost entirely explained by the fact that they select into certain positions, such as part-time positions, that are characterized by shorter spell durations generally. After controlling for position, the gender differences in the frequency of mobility largely disappear.

Table 3.4: Results of Cox proportional hazard estimations for service spell duration.

	All spells				Services in parishes only			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	1.33	1.29	1.31	1.11	1.45	1.45	1.39	1.10
	[1.26,1.41]	[1.21,1.36]	[1.24,1.39]	[1.05,1.18]	[1.31,1.61]	[1.31,1.61]	[1.26,1.53]	[0.99,1.22]
Age at start 30-39		1.15				1.33		
		[0.95,1.40]				[0.98,1.81]		
Age at start 40-49		0.76				1.08		
		[0.62,0.93]				[0.79,1.48]		
Age at start 50-59		0.69				1.30		
		[0.56,0.86]				[0.93,1.81]		
Age at start 60+		1.12				1.35		
		[0.76,1.63]				[0.65,2.83]		
Spell starts 1996-2000			1.19				1.28	
			[1.09,1.29]				[1.12,1.47]	
Spell starts 2001-2005			1.14				1.33	
			[1.05,1.24]				[1.16,1.53]	
Spell starts 2006-2010			1.29				2.15	
			[1.81,1.42]				[1.83,2.52]	
Spell starts 2011-2016			1.31				1.62	
			[1.17,1.47]				[1.35,1.94]	
Regular service in parish, part-time				8.15				11.00
				[7.28,9.13]				[9.31,12.98]
Training service in parish, full-time				1.02				1.94
				[0.93,1.13]				[1.74,2.15]
Training service in parish, part-time				6.50				12.83
				[5.69,7.42]				[9.57,17.19]
Other service outside of parish				1.81				
				[1.65,1.98]				
Parental leave				9.21				
				[7.27,11.65]				
Other leave				2.43				
				[2.25,2.63]				
Number of observations	8,413	8,413	8,413	8,413	2,927	2,927	2,927	2,927
Number of failures	5,729	5,729	5,729	5,729	1,922	1,922	1,922	1,922

Notes: Table displays hazard ratios. 95% confidence intervals in brackets. Standard errors clustered by pastor.

We have also explored whether there are gender differences not only in the *frequency* of mobility, but also in the *kinds of parishes* that pastors select into. A first piece of evidence is table 3.5, which shows the number of pastors per parish. The table shows, first, that parishes are frequently staffed with multiple pastors: even among regular-service, full-time positions, fewer than one-fourth of pastors work in a parish by themselves. Second, women tend to select into larger parishes with more colleagues, a finding that holds true across types of positions. For example, for pastors in regular-service, full-time positions, women have 2.7 colleagues on average, compared to just 2.4 colleagues for men. Interestingly, this pattern already begins during training service, where female pastors are somewhat more likely to select into parishes where they have three or more colleagues, and less likely to have zero or one colleague.

Table 3.5: Number of colleagues per pastor

	Regular service, full-time		Regular service, part-time		Training service	
Number of colleagues	Men	Women	Men	Women	Men	Women
0	23.5%	15.1%	10.3%	4.7%	3.3%	2.7%
1	25.1%	23.9%	32.4%	32.8%	31.9%	26.4%
2	20.0%	23.8%	21.1%	20.4%	23.2%	23.7%
3	12.7%	14.7%	13.3%	14.9%	15.9%	17.7%
4	7.3%	8.1%	7.9%	9.0%	8.9%	11.9%
5+	11.4%	14.3%	15.0%	18.1%	16.7%	17.6%
Average	2.1	2.4	2.4	2.7	2.8	2.9

Notes: Unit of observation is a pastor-year. Number of colleagues as of January in a given year.

As we show in table 3.6, the finding that women tend to select into larger parishes extends to some other dimensions, but not others. A female pastor works in a parish that has, on average, 400 more members, which is a sizable difference. However, turning to outcomes, the picture is far more muddled. Female pastors perform more confirmations, funerals and services, as well as bringing in more donations. However, there is no significant difference in baptisms, and the number of marriages and attendance is actually lower for parishes served by female pastors compared to male pastors, despite the larger size of those parishes. We will explore the question to which extent these discrepancies reflect differences in productivity or selection more fully in the following section.

Table 3.6: Average characteristics of parishes by pastor gender

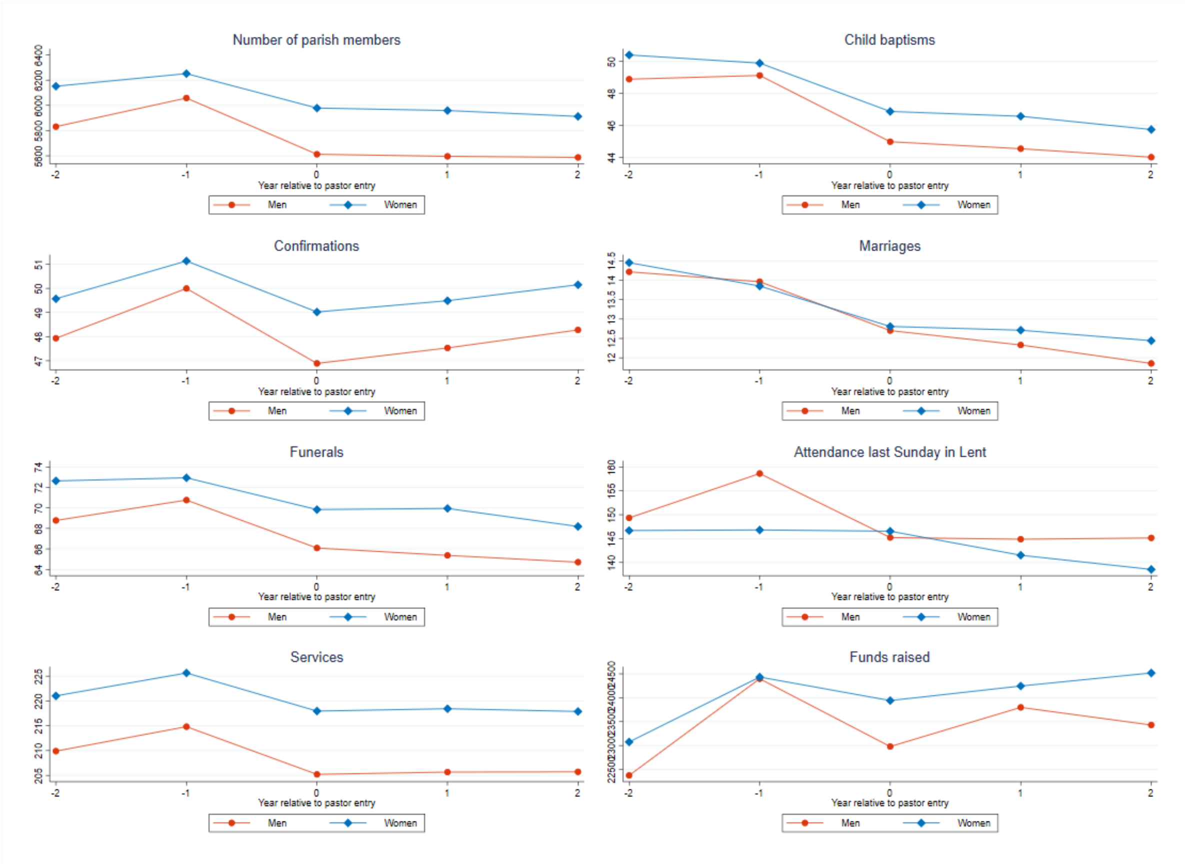
	Women	Men	All
Parish members	6,325.1	5,989.1	6,070.9*
	(4,547.7)	(4,533.6)	(4,539.2)
Child baptisms	47.0	46.7	46.8
	(35.6)	(36.5)	36.3)
Confirmations	52.2	49.5	50.2*
	(38.2)	(36.4)	(36.8)
Marriages	12.3	13.1	12.9*
	(10.6)	(11.7)	(11.4)
Funerals	74.2	71.0	71.8*
	(51.8)	(53.4)	(53.0)
Attendance last Sunday in lent	141.3	147.2	145.8*
	(119.1)	(130.7)	(128.0)
Services	222.3	216.3	217.7*
	(135.6)	(137.8)	(137.3)
Total donations, in Euro	25,002.7	24,408.2	24,551.9
	(18,773.9)	(18,950.5)	(18,909.3)
Number of observations	9,845	30,520	40,365

Notes: Unit of observation is a pastor-year. Standard deviations in parentheses. * indicates that the difference between genders is significant at the 95% level.

One concern with any identification strategy that relies on parish fixed effects is that men and women might select into parishes with different pre-existing trends. To investigate this possibility, we have conducted a simple event study, whose results we show in figure 3.4. Each marker shows, separately for male and female pastors, average outcomes from two years before until two years after the pastor has entered the parish. If male and female pastors systematically selected into parishes with different pre-existing trends, the slopes of the red and blue lines would be different, at least until year zero. However, this is not what we observe – the two lines generally run in parallel. One notable exception, for which we do not have a satisfactory explanation, is an unusual spike in church attendance during the year before a male pastor moves in. Investigation of the data confirmed that this is a general phenomenon in most years and not driven by a few outliers. However, since relative attendance two years before a move is quite similar for men and women, we find it implausible that this one-time event should constitute a general differential trend by gender.

In conclusion, there is some evidence that men and women systematically sort into differ-

Figure 3.4: Event study of average outcomes before and after pastor entry.



ent kinds of parishes, which indicates that it will be important to control for selection biases. However, there is no strong evidence that the parishes men and women select into have different pre-existing time trends, which supports the use of linear models with parish fixed effects.

3.4 Estimations of gender differentials in productivity

For our baseline estimations, we restrict the dataset to the set of parishes served by just a single pastor at a given time. We rely on the following specification:

$$y_{it} = \beta_0 + \beta_1 \Gamma_{it} + \beta_2 X_{it} + \alpha_i + \delta_t + \epsilon_{it}. \tag{3.1}$$

Here, y_{it} is the outcome of interest at parish i and time t. β_1 is the parameter of interest and measures the effect of $\Gamma_i t$, a dummy variable indicating wthether the pastor in parish i at time t is female. As control variables X_{it}, we include pastor age and the number of parish members along with its square. The model includes a full set of parish-fixed effects α_i and year-fixed effects δ_t.

The top panel of table 3.7 displays the results of estimating versions of equation (3.1) for various outcome variables. Each cell in the table reports the estimated regression coefficient on the dummy variable for a female pastor. Column (1) contains specifications without any control variables, and is meant mostly to illustrate whether male and female pastors select into different kinds of parishes. The results are somewhat uneven and do not provide strong evidence of differential selection by female pastors. In particular, while we had seen earlier in table 3.5 that women tend to select into larger parishes staffed with more pastors, there is no consistent pattern of women serving in larger or more active parishes once we restrict our sample to parishes served by just a single pastor, as we do in table 3.7.

The most striking result in table 3.7 is the negative effect of female pastor gender on donations, both for the own parish and for other causes. The fact that this effect appears for two different measures of donations is notable, as is the remarkable size and statistical significance of the effect.[5] If we recall from table 3.6 that the average annual size of donations is on the order of €5,300 for one's own parish and €15,500 for other causes, the coefficient estimates of 389 and 1,167, respectively, imply that women raise approximately 7.5% fewer donations than men, on average.

Given the large number of outcomes available in our data, we were concerned that studying each of them individually might obscure some patterns in the data. Therefore, in the bottom panel of table 3.7, we have tested several further hypotheses. First, we have pooled all observations of our attendance variables, as one might reasonably expect that pastors who fail to attract attendance at their services would do so every Sunday of the year. Again, however, we find no notable average difference between male and female pastors. Second, we have summed up all donations in a given year, and the results only confirm the effects we observed when we studied both measures of donations individually. In the final row, we have aggregated all outcomes from table 3.7, excluding membership and the female shares of exits, entries, volunteers and presbyters, which are not, strictly speaking, measures of performance. To make all variables comparable, we have converted them into z-scores, so that a value of 100 means that the value is 1 standard deviation above the mean.[6] While results are now statistically significant, they are also very small: on average across all outcomes, female pastors perform worse by a mere 3% of a standard deviation, which, for practical purposes, is essentially identical to zero.

We have repeated our baseline estimation for the larger sample of all parishes, including those with more than one pastor. To form the variables for pastor gender and age, we have taken averages across all pastors serving in a given parish, with pastors in part-time positions receiving half of the weight of their full-time counterparts. While one might expect that the resulting larger sample size will yield more powerful tests, a drawback of including larger parishes is that it is not clear that effects are linear. For example, if two pastors share duties in a parish,

[5] We do not wish to overstate the finding that our results are statistically significant, because our confidence intervals are not adjusted for the multiple testing problem (Romano, Shaikh, and Wolf, 2010). We have not yet found a suitable correction for this problem, because the clustered structure of our standard errors, coupled with the fact that one would ideally wish to test for joint significance of both specifications with donations as outcome variables, make ours a nonstandard problem that is not easily addressed with existing software packages.

[6] Since fewer exits indicate better performance, we first multiplied exits by -1.

one of whom likes working with children but the other does not, we might expect them to split responsibilities so that all children's services are held by the pastor who enjoys doing them. In this case, the number of children's services would be determined only by one pastor, rather than some average of what each pastor would offer if they were in a parish by themselves. This effect is why we prefer the baseline specification with just a single pastor per parish, which offers a sharper test of our hypothesis of gender-specific effects on outcomes.

Nevertheless, we regard the results of the estimations on the full sample, which we present in table 3.8, as quite interesting. Column (1), which again contains the results for the specification without any control variables, shows significant differences in selection by gender now that the larger parishes are taken into account. Women select into much larger parishes than men – on average, a parish staffed only with female pastors will have about 820 more members than one staffed entirely with male pastors. Unsurprisingly, these larger parishes also have both more entries and more exits, and more activities along many dimensions, including baptisms, confirmation candidates, funerals (but not weddings), the number of volunteers, and church attendance, at least on Christmas Eve. However, only the amount of funds raised for the own parish is, on average, larger for parishes staffed with more female pastors, while the opposite is true for funds for other causes. Neither estimate is statistically significant. Finally, we observe that parishes served by a larger share of female pastors also have more female presbyters. While this finding is interesting on its face, the estimated relationship is quite weak, with a parish served exclusively by female pastors having an average of 4% more female presbyters than one served exclusively by male pastors. Further, the relationship need not be causal, and even if it is, it is not clear in which direction the causality runs: female pastors may serve as a role model and enable women to successfully run for a position as presbyter, or female presbyters may be more likely to select female pastors for job openings.

Column (2) displays the results for the full specification including all control variables. Not a single regression coefficient comes out as significant on the 5% level – not even the z-score, which pools observations across outcomes and relies on an estimation with over 300,000 observations.

Perhaps most importantly, the coefficients on both donation variables are negative, but neither is statistically significant and, at -268 and -530, the coefficients come in at 31% and 55% smaller than the respective coefficients from the baseline specification. We regard this difference as evidence that the full sample just does not offer such a sharp test as the sample of parishes served by just a single pastor.

In light of the stark contrast between the estimations with and without control variables in the full sample, we have explored in a bit more detail which control variables are responsible for the disappearing effects on the gender variable. Therefore, in table 3.9, we present the results from the Gelbach (2016) decomposition, which apportions the change in regression coefficients on the gender dummy to the various control variables added between columns (1) and (2) of table 3.7. For example, the last row of table 3.9 shows that, in the regressions with the z-score of all outcomes as the dependent variable, the baseline coefficient without control variables is 7.05 and the coefficient on the full specification with control variables is -1.26. The difference of 8.31 is more than entirely accounted for by the addition of parish size controls, and somewhat diminished by the addition of the full set of parish fixed effects. In other words, the fact that female pastors select into particularly large parishes fully explains why, in the regression without control variables, they appear to have better performances than male pastors; in fact, the parishes they select into have actually somewhat worse outcomes conditional on their size. The finding that the difference between the unconditional and the conditional gender effect is due to differences in parish size, and is somewhat mitigated by the full set of parish fixed effects repeats itself multiple times throughout table 3.9. However, it is not universal: for example, in the case of the number of volunteers, the addition of year and parish fixed effects is about twice as important as the parish size effects to account for the reduction of the regression coefficient on the gender effect.

Table 3.7: Baseline regression results of the female pastor dummy, single-pastor parishes

Dependent variable	(1)	(2)	Observations
Number of parish members, in '000	0.13	0.03	7,339
	(0.10)	(0.03)	
Parish member exits	1.21	0.53	7,337
	(0.85)	(0.52)	
Adult parish member entries	0.36	0.00	7,337
	(0.41)	(0.68)	
Adult baptisms	0.24	0.44	7,337
	(0.27)	(0.62)	
Child baptisms	-0.17	-0.72	7,337
	(0.84)	(0.61)	
Confirmation candidates	0.99	-0.59	7,337
	(1.00)	(0.53)	
Marriages	-0.44	-0.62**	7,337
	(0.35)	(0.30)	
Funerals	1.31	0.16	7,337
	(1.24)	(0.49)	
Masses	1.05	-0.26	7,337
	(4.75)	(3.04)	
Children's masses	-5.34***	0.62	7,337
	(1.88)	(2.32)	
Youth club meetings	-0.16	0.44*	7,337
	(0.31)	(0.27)	
Volunteers	5.44	0.86	7,337
	(4.93)	(3.22)	
Log attendance first Sunday in lent	-0.05	0.02	7,260
	(0.04)	(0.04)	
Log attendance first Sunday in lent (children)	-0.08	-0.05	6,264
	(0.06)	(0.08)	
Log attendance Good Friday	-0.09*	-0.07*	7,331
	(0.05)	(0.04)	
Log attendance Advent	-0.10**	-0.04	7,230
	(0.05)	(0.04)	
Log attendance Christmas Eve	0.02	-0.03	7,322
	(0.05)	(0.03)	
Funds raised for the parish, in 2010 Euros	-210.39	-388.97**	3,648
	(262.45)	(194.07)	
Funds raised for other causes, in 2010 Euros	-1,131.65**	-1,167.48**	3,648
	(528.06)	(534.24)	
Female share of exits	0.02*	-0.00	6,986
	(0.01)	(0.02)	
Female share of entries	-0.01	-0.05**	6,302
	(0.01)	(0.02)	
Share female volunteers	0.01	-0.00	7,330
	(0.01)	(0.01)	
Share female presbyters	0.02	0.00	7,316
	(0.01)	(0.02)	
Log attendance combined	-0.05	-0.04	29,421
	(0.04)	(0.02)	
Funds raised, in 2010 Euros	-1,342**	-1,556.44*	3,648
	(664.11)	(536.64)	
All outcome variables	-3.96**	-3.21**	123,794
	(1.69)	(1.59)	
Parish size controls		Yes	
Pastor age controls		Yes	
Month-by-year fixed effects		Yes	
Parish fixed effects		Yes	

Notes: [1] Each cell reports the result of a dummy on pastor gender. [2] No parish size controls in the regressions where number of parish members is the dependent variable. [3] Parish size controls are the number of parish members and number of parish members squared. [4] Standard errors clustered by parish or, in the case of regressions with Z-scores as the dependent variable, by parish and year.

Table 3.8: Baseline regression results of the female pastor dummy, all parishes

Dependent variable	(1)	(2)	Observations
Number of parish members, in '000	0.82***	0.00	18,127
	(0.26)	(0.05)	
Parish member exits	7.01***	-0.10	18,123
	(2.41)	(0.60)	
Adult parish member entries	2.28***	-0.02	18,123
	(0.68)	(0.34)	
Adult baptisms	0.58**	0.06	18,123
	(0.25)	(0.26)	
Child baptisms	2.45	-0.18	18,123
	(1.88)	(0.58)	
Confirmation candidates	5.48**	-0.48	18,121
	(2.18)	(0.50)	
Marriages	-0.64	-0.41	18,123
	(0.52)	(0.28)	
Funerals	9.48***	0.36	18,123
	(3.06)	(0.61)	
Masses	17.70**	0.43	18,122
	(7.66)	(2.91)	
Children's masses	-6.82***	1.13	18,123
	(2.16)	(1.75)	
Youth club meetings	0.77*	0.32	18,123
	(0.43)	(0.23)	
Volunteers	35.01***	0.58	18,123
	(9.02)	(4.13)	
Log attendance first Sunday in lent	-0.01	-0.01	18,025
	(0.05)	(0.03)	
Log attendance first Sunday in lent (children)	0.03	0.01	16,144
	(0.05)	(0.05)	
Log attendance Good Friday	-0.05	-0.03	18,113
	(0.05)	(0.02)	
Log attendance Advent	0.02	-0.02	17,973
	(0.06)	(0.03)	
Log attendance Christmas Eve	0.14**	-0.02	18,099
	(0.06)	(0.02)	
Funds raised for the parish, in 2010 Euros	520.08	-267.66	9,054
	(377.67)	(213.12)	
Funds raised for other causes, in 2010 Euros	-387.59	-530.33	9,054
	(913.72)	(360.84)	
Female share of exits	0.03***	0.00	17,712
	(0.01)	(0.01)	
Female share of entries	-0.01	-0.03	16,888
	(0.01)	(0.02)	
Share female volunteers	0.02***	0.01	18,110
	(0.01)	(0.01)	
Share female presbyters	0.04***	0.02*	18,077
	(0.01)	(0.01)	
Log attendance combined	0.03	-0.02	72,326
	(0.05)	(0.02)	
Funds raised, in 2010 Euros	132.49	-797.98*	9,054
	(1,199.83)	(410.12)	
All outcome variables	7.06***	-1.26*	306,053
	(1.69)	(1.59)	
Parish size controls		Yes	
Pastor age controls		Yes	
Month-by-year fixed effects		Yes	
Parish fixed effects		Yes	

Notes: [1] Each cell reports the result of a dummy on pastor gender. [2] No parish size controls in the regressions where number of parish members is the dependent variable. [3] Parish size controls are the number of parish members and number of parish members squared. [4] Standard errors clustered by parish or, in the case of regressions with Z-scores as the dependent variable, by parish and year.

Table 3.9: Results of the Gelbach (2016) decomposition, single-pastor parishes

	Baseline coefficient	Decomposition				Full-specification coeffcient
		Age	Parish size	Year dummies	Parish dummies	
Number of parish members, in '000	1.2843	.0009		.04166	-.14423	.02677
Parish member exits	1.20871	-.02668	-.8107	-.06788	.22207	.52552
Adult parish member entries	.36231	-.03488	-.31433	-.30779	.29621	.00152
Adult baptisms	.24269	-.01122	-.12797	-.0766	.41389	.44079
Child baptisms	-.16861	-.04486	-.90026	1.00614	-.61475	-.72234
Confirmation candidates	.99473	-.00095	-1.03999	-.07346	-.46676	-.58644
Marriages	-.43803	-.0048	-.28168	.47612	-.37518	-.62356
Funerals	1.30991	.02095	-.74922	.28996	-.71094	.16067
Masses	1.04701	-.12184	-1.44159	1.55953	-1.30383	-.26072
Children's masses	-5.33996	.08573	-.09337	3.89963	2.07204	.62406
Youth club meetings	-.1648	.00985	-.10882	.15904	.54643	.4417
Volunteers	5.43743	-.31158	-1.80232	-6.04768	3.58126	.85711
Log attendance first Sunday in lent	-.04861	-.00518	-.0224	.03999	.05865	.02246
Log attendance first Sunday in lent (children)	-.07636	.00148	-.01917	.01636	.02739	-.05031
Log attendance Good Friday	-.08595	-.00054	-.01111	.05184	-.02183	-.06759
Log attendance Advent	-.09598	-.00727	-.02175	.02641	.06266	-.03493
Log attendance Christmas Eve	.02189	-.00288	-.02557	-.00515	-.01927	-.03098
Funds raised for the parish, in 2010 Euros	-210.394	-25.9303	-64.3564	134.901	46.6168	-388.965
Funds raised for other causes, in 2010 Euros	-1,131.7	-33.8058	-313.035	151.684	159.33	-1,167.5
Female share of exits	.0159	-.00049	-.00072	-.01397	-.00381	-.00309
Female share of entries	-.01247	-.00078	-.00192	.00458	-.03732	-.04791
Share female volunteers	.00539	-.00035	-.00202	-.00494	.00139	-.00053
Share female presbyters	.01697	.00048	.00014	-.02195	.00638	.00202
Log attendance combined	-.05304	-.0042	-.01449	.01582	.01859	-.03732
Funds raised, in 2010 Euros	-1,342	-59.7361	-377.392	16.7834	205.947	-1,556.4
All outcome variables	-3.95983	-.41963	-3.9525	2.26521	2.86002	-3.20673

Returning to our baseline specification and the finding of significant gender effects of donations, we have conducted two more tests to explore possible heterogeneities behind this result. In the first, we have sorted pastors into ten-year age cohorts and interacted the resulting cohort dummies with the gender dummy. The results, displayed in table 3.10, indicate that the effect varies with age in a U-shape. The estimated effects are strongest in the age 30-39 and 60-69 cohorts, where the difference between men and women in terms of total donations amounts to more than €3,000 on average. By contrast, results are smallest for the age 40-49 cohort, where female pastors even raise more donations for their own parishes than their male counterparts, although the results are not statistically significant.

Table 3.10: Regression results with separate effects by age, single-pastor parishes

	(1) Donations for own parish	(2) Donations for other causes	(3) Total donations
Men, aged 20-29		Baseline group	
Men, aged 30-39	559. 35	1,518.90	2,078.25*
	(481.80)	(1,054.09)	(1,191.83)
Men, aged 40-49	59.61	333.98	393.59
	(180.89)	(386.85)	(368.24)
Men, aged 50-59	33.49	245.57	279.06
	(509.32)	(347.68)	(750.27)
Men, aged 60-69	23.45	256.70	280.15
	(209.26)	(390.44)	(381.91)
Women, aged 30-39	-245.06	-680.62***	-925.68***
	(329.30)	(220.43)	(255.72)
Women, aged 40-49	346.18	-99.03	247.15
	(248.43)	(573.84)	(682.80)
Women, aged 50-59	-464.17	-604.40	-1,068.58
	(365.09)	(630.65)	(895.08)
Women, aged 60-69	-607.80	-2,424.78***	-3,032.58**
	(826.64)	(897.34)	(1,222.30)
Number of observations	3,648	3,648	3,648
Parish size controls	Yes	Yes	Yes
Pastor age controls	Yes	Yes	Yes
Month-by-year fixed effects	Yes	Yes	Yes
Parish fixed effects	Yes	Yes	Yes

Notes: Standard errors clustered by parish.

Finally, we have investigated whether gender effects are heterogeneous across the distribution of the donation variables. To this end, we have estimated the unconditional fixed effects quantile regression estimator developed by Powell (2020). The results, shown in table 3.11, indicate that results are largest at the upper end of the donation distribution, especially for donations for causes other than one's own parish. In other words, there is a sizable observed difference between the most successful male fundraisers versus the most successful female fundraisers, whereas the median female fundraiser is nearly as effective as the median male fundraiser.

Table 3.11: Unconditional quantile regression results of gender effects on donations

Quantile	(1) Donations for own parish	(2) Donations for other causes	(3) Total donations
10	3.24	-362.42	-713.12
	(196.87)	(388.47)	(471.26)
25	-13.06	-555.24	-1,168.46***
	(239.95)	(385.98)	(451.75)
50	-136.16	-465.97	-471.41
	(354.17)	433.04)	(441.74)
75	-274.15	-983.34	-1,272.56**
	(504.60)	(618.15)	(647.05)
90	-285.03	-1,507.43	-3,418.82**
	(554.15)	(1,104.09)	(1,338.83)
Number of observations	3,648	3,648	3,648
Parish size controls	Yes	Yes	Yes
Pastor age controls	Yes	Yes	Yes
Month-by-year fixed effects	Yes	Yes	Yes
Parish fixed effects	Yes	Yes	Yes

3.5 Discussion

In this paper, we have estimated the effects of female leadership for a sample of pastors across a wide range of outcomes. While we are unable to detect any meaningful differences between parishes led by male or female pastors across a wide range of outcomes, one result stands out: parishes led by female pastors raise about 7.5% fewer donations for both the own parish and other causes, on average. These results are sizable and seem to be concentrated among the youngest and oldest cohorts, as well as at the top of the distribution of fundraisers.

In our view, there are two – not mutually exclusive – possible explanations for these results. The first is that the distribution of skills or preferences is different between men and women, such that fewer women dedicate themselves towards becoming excellent fundraisers. It is well possible that these women choose instead to specialize in other areas which, despite the wide range of outcomes in our dataset, we have been unable to quantify. The second explanation is that, on average, female pastors try just as hard to raise donations as their male counterparts, but that they are less successful because their parishes discriminate against them. Indeed, such an explanation would be consistent with Perry (2013), who has documented a similar mechanism using qualitative data in the context of evangelical churches in the United States.

Chapter 4

Apprenticeship Training in Heterogeneous Firms[1]

4.1 Introduction

As the wages of lower-educated workers have stagnated in many countries, firm-based apprenticeship trainings, which are common in Germany, Austria and Switzerland, have received increasing attention. Over the course of an apprenticeship, which typically lasts three years, apprentices work in training firms part-time while attending vocational school, culminating in a standardized examination and a certificate. Apprenticeships exist in a wide range of medium-skilled occupations as diverse as shop manager, hairdresser, laboratory assistant, or paralegal. In the US and the UK, adopting and expanding continental European-style apprenticeship programs has featured prominently as a possible avenue for improving the labor market prospects for workers without tertiary education.[2] In Germany and Austria, the arrival of large numbers of refugees

[1] This chapter is joint work with Andrea Weber, Department of Economics and Business, Central European University, Quellenstraße 51, 1100 Vienna, Austria. `webera@ceu.edu`.

[2] For instance, in April of 2016, the Obama administration announced a plan to invest \$90 million into expansion of firm-based apprenticeships in the United States. Likewise, the British government released a plan to increase the quality and quantity of apprentices, explicitly pointing to Germany and Austria and models. See White House (2016) and Department for Business and Skills (2015).

with few formal qualifications has renewed interest in firm-based training programs in order to integrate these individuals into the labor market.[3]

From a theoretical standpoint, two features of apprenticeships have piqued the interest of labor economists. First, with its emphasis on vocational school attendance and the standardized final examination, the skills conferred during an apprenticeship are quite general, rather than firm-specific (Acemoglu and Pischke, 1998). Second, a series of studies have estimated that running apprenticeship programs costs firms thousands of Euros per apprentice and year, after accounting for training wages and other expenses as well as the production provided by apprentices (Bardeleben, Beicht, and Fehér, 1995; Beicht and Walden, 2002). Together, these two features present something of a puzzle: in a perfectly competitive world, the worker would earn her marginal product at every firm, and no firm would be able to recoup its training costs. Hence, in equilibrium, no firm provides general training (Becker, 1964). A number of articles have therefore appealed to market imperfections to justify why workers receive general training, ranging from search frictions (Acemoglu, 1997) to adverse selection caused by unobserved worker ability (Acemoglu and Pischke, 1998; Autor, 2001) to a firm's inability to observe worker effort (Loewenstein and Spletzer, 1998; Acemoglu and Pischke, 1999) to wage compression caused by the presence of unions (Dustmann and Schönberg, 2009) or minimum wages (Acemoglu and Pischke, 2003) to complementarities between firm- and worker-specific skills (Kessler and Lülfesmann, 2006). These papers share an assumption that firms are homogeneous: if it is advantageous for one firm to train apprentices, it should pay off for all them to do so.

However, it is now very well documented that firms are *very* heterogeneous across many dimensions, notably their productivity; see, e.g., Syverson (2011). It is natural to ask whether this heterogeneity impacts firms' decisions about training: do higher-productivity firms offer more or fewer training opportunities? Do they retain more workers after the completion of training? Do they invest more into their employees' training? To answer these questions, we

[3]For example, the German government formed the "Alliance for Initial and Further Training", which explicitly sought to improve the opportunities for young migrants to enter vocational training (Allianz für Aus- und Weiterbildung, 2014).

present a simple theoretical model in which heterogeneous firms offer apprenticeships that serve a dual purpose: they impart human capital onto the worker, and they allow the worker-firm pair to learn about the quality of their match. If the match turns out to be of sufficiently high quality, the worker-firm pair will retain it and split the surplus it generates. The prospect of enjoying the proceeds from such a surplus, in turn, is what motivates firms to provide training in the first place.

Our model predicts that, more productive firms will, on average, be better able to take advantages of good matches and will achieve bigger surpluses after training. These firms are therefore more sensitive to match quality and will be less likely to retain a worker with whom the match is mediocre; anticipating this, the more productive firms will run smaller apprenticeship programs. However, the expected surplus from a good match is larger at more productive firms, and workers can use this fact to bargain for more training. In sum, more productive firms train fewer apprentices, but they train them more intensively.

We then test some of these predictions using administrative data on the universe of Austrian private-sector apprenticeships over a 40-year period. Our empirical analysis supports the predictions that surviving matches are of higher quality and that more productive firms run smaller apprenticeship programs.

4.2 A Simple Theoretical Model

The following section presents an illustrative three-stage model of apprenticeships in order to guide our thoughts on how and why firms might pursue different strategies when it comes to apprenticeship training. Three features of the model serve as our point of departure: first, as argued forcefully by the previous literature, notably Acemoglu and Pischke (1999), the content of apprenticeship training is mostly general. Second, given that apprenticeships last a long time — typically three, sometimes as much as four years — and that most apprentices are very young when they enter (usually, between 16 and 20 years old), an apprenticeship is an opportunity for the firm to learn which skills the apprentice has and how well they fit with the skill requirements of the firm. Third, firms are heterogeneous in productivity, a crucial feature of today's economies. For example, Syverson (2004) demonstrates that, for US manufacturing firms, a firm at the 90th percentile of the productivity distribution is twice as productive as a firm at the 10th percentile, even within narrowly defined industries.

The model features firms and workers; everyone is risk-neutral and there is no discounting. All workers are ex ante identical, while firms have either high productivity γ_1 or low productivity γ_0. In the first stage, each firm learns its type γ and the number of vacancies $V \in \mathbb{Z}_+$ that it will have to fill during the coming stage. The firm then decides upon the number of apprentices $M \in \mathbb{Z}_+$ which it wishes to train. We assume that there is a sufficiently large number of potential apprentices available so that each firm can hire as many apprentices as it likes, and relegate the details of how workers and firms meet to appendix 4.C.

In the second stage, training takes place. During training, the workers produce zero output (a normalization) and accumulate a level of human capital H determined by Nash bargaining between worker and firm, where the firm's bargaining weight is given by β. The firm must pay a cost of training $c(M, H)$; for simplicity, we specify the functional form $c(M, H) = M(\epsilon + H^2)$, where ϵ is a small number.[4] Finally, each worker-firm pair learns about the quality of their match μ, which is drawn from a distribution $F(\mu)$ over the interval $[\mu_{min}, \mu_{max}]$ with expected value $\bar{\mu}$ and density $f(\mu)$. Having learned their match quality, the worker-firm pair can decide to either retain the match or separate. A worker who has separated from his training firm enters a perfectly competitive labor market populated by high-productivity firms. Should a firm retain fewer trained workers than it has vacancies available, the firm pays a cost b for each vacancy, and the vacancy remains unfilled.[5]

[4]The exact functional form is by no means essential, but the chosen form simplifies some expressions and conveys the basic intuition. More generally, $c(\cdot)$ should be weakly concave in M and strictly convex in H, with the marginal costs of acquiring human capital H starting at zero and then increasing. The purpose of ϵ is to rule out a strategy where the firm trains an infinite number of apprentices at zero costs, while the worker accumulates no human capital at all.

[5]One could try to flesh out this secondary labor market in more detail, by letting both low- and high-productivity firms fill their vacancies there. The presence of heterogeneous firms would require some labor market

In the third period, the firm and all retained workers agree upon a wage and production takes place.[6] Workers who are still with their training firm will produce $H + \mu\gamma$, while a worker who leaves for a high-productivity firm on the outside market will earn her expected marginal product $H + \bar{\mu}\gamma_1$. For the results below to hold, it is important that match quality μ and firm productivity γ are complements, such that the most productive firms are best able to take advantage of the best matches, and the multiplicative structure captures this in a simple way. The surplus that a retained worker and a training firm have available from continuing the employment relationship is therefore $\mu\gamma - \bar{\mu}\gamma_1 + b$, and it is split according to Nash bargaining. Again, the firm's bargaining weight is β. Thus, the worker's wage is

$$w = H + \beta(\bar{\mu}\gamma_1) + (1 - \beta)(\mu\gamma + b). \tag{4.1}$$

First, we are interested in the number of apprentices M which the firm trains initially. If the firm decides to train fewer than or equal to V apprentices, it will retain each trained worker who does not want to leave for the secondary market, which requires a match that turns out to be larger than $\mu^* \equiv \frac{\bar{\mu}\gamma_1 - b}{\gamma}$. The firm's maximization problem to decide upon the optimal number of apprentices in this case is

$$\max_{M} \; MP(\mu \geq \mu^*)\beta\left(E\left[\mu|\mu \geq \mu^*\right] - \mu^*\right) - c\left(M, H\right)$$
$$\text{subject to } \; 0 \leq M \leq V.$$

In words, the firm weighs the possibility that a worker might form a good enough match that they will choose to stay post-apprenticeship, and the benefits that the firm will derive from retaining the worker, against the costs of training. Having already decided to train M apprentices, the

imperfection – otherwise, no worker would choose to join a low-productivity firm – and the secondary market would become quite complicated. Since this is not our focus, we choose to stick with the simplest possible set-up.

[6] Notice that the firm has already let go of the all workers except for the ones with the V highest values of μ, provided they want to stay. This means that it cannot use the threat of hiring the less productive workers to improve its bargaining position in the wage negotiations — the firm's outside option is to pay b and leave the vacancy unfilled entirely.

net expected benefit of hiring an additional apprentice is

$$S(V,M) = P(\mu \geq \mu^*)\beta\left(E\left[\mu|\mu \geq \mu^*\right] - \mu^*\right) - H^2 - \epsilon.$$

As we will see below (see footnote 8), the optimal amount of human capital H does not change with M, so long as $M < V$. This means that training costs c are linear in M, and this problem does not have an interior solution. The firm will never find it optimal to train fewer apprentices than it needs; it will either train at least V apprentices, or no apprentices at all.

Should the firm decide to train more than V apprentices, it will choose to retain the V apprentices who turn out to have the highest match quality μ, provided the match quality μ is sufficiently high that the workers will prefer staying with the training firm over the secondary market. Let μ_i^j denote the ith-largest value of μ out of j draws. Then, the firm's problem is

$$\begin{aligned} \max_{M} \quad & -c\left(M,H\right) + \sum_{i=1}^{V} P(\mu_i^M \geq \mu^*)\beta\left(E[\mu_i^M|\mu_i^M \geq \mu^*] - \mu^*\right) \\ & \text{subject to } \; M \geq V, \end{aligned}$$

and the expected net gain from training an additional worker, having already decided to train M workers, is

$$\begin{aligned} S(V,M) = \beta\Big\{ & P(\mu_V^M \leq \mu^*) \cdot P(\mu \geq \mu^*) \cdot \left(E[\mu|\mu \geq \mu^*] - \mu^*\right) \\ & + P\left(\mu_V^M > \mu^*\right) \cdot P\left(\mu \geq \mu_V^M|\mu_V^M \geq \mu^*\right) \cdot \left(E[\mu|\mu \geq \mu_V^M \geq \mu^*] - E[\mu_V^M|\mu_V^M \geq \mu^*]\right)\Big\} \\ & - H^2 - \epsilon. \end{aligned} \tag{4.2}$$

The first line covers the case in which, out of the M workers already trained, the match quality of the worker with the Vth-highest value of μ does not surpass μ^*. In this case, the firm will not be able to hire all V workers from the pool of apprentices it has trained previously, and all an additional worker needs to do to get retained is to draw a match quality higher than μ^*. The second line addresses the case in which at least V trained workers have a match productivity higher than μ^*. Here, to be retained by the firm, the worker's match productivity must be at least as large as that of the Vth-highest worker.

We now turn to the level of human capital H that the worker receives during training. This level is bargained over at the beginning of training. Since the hiring stage has already passed, the firm cannot go back to hire another apprentice instead, so its outside option is zero. Similarly, the worker's outside option is to proceed to the secondary labor market without having received any training. To simplify the notation, let $q = P\left(\mu \geq \max\{\mu^*, \mu_V^M\}\right)$ denote the ex ante probability that a match will persist and let $\tilde{\mu} = E\left[\mu | \mu \geq \max\{\mu^*, \mu_V^M\}\right] - E[\max\{\mu^*, \mu_V^M\}]$ denote the expected increase in μ over the next-best apprentice that the match will provide, if it persists. Nash bargaining now yields the maximization problem

$$\max_H \left(\beta q \tilde{\mu} - H^2 - \epsilon\right)^\beta \left(H + (1-\beta) q \left(E\left[\mu | \mu \geq \max\{\mu^*, \mu_V^M\}\right] - \mu^*\right)\right)^{1-\beta}.$$

The firm's benefit from continuing the match is given by the share β of the expected surplus of the match, $\tilde{\mu}$, which will only be realized if the match persists. This event occurs with probability q. By contrast, the costs of training $H^2 + \epsilon$ accrue for sure, and they are borne solely by the firm. Notice that the firm derives no benefit from providing training to the worker, since the worker's wage rises one-for-one with every unit of human capital. The worker's benefit from training consists of the amount of human capital H plus the worker's share $1-\beta$ of the expected surplus $q\left(E\left[\mu | \mu \geq \max\{\mu^*, \mu_V^M\}\right] - \mu^*\right)$.[7] The first-order condition is[8]

$$(1+\beta)H^2 + 2H(1-\beta)\beta q \left(E\left[\mu | \mu \geq \max\{\mu^*, \mu_V^M\}\right] - \mu^*\right) - (1-\beta)\beta q \tilde{\mu} + (1-\beta)\epsilon = 0. \quad (4.3)$$

[7] The firm is in a better bargaining position than the worker here. If negotiations break down, the worker will have no chance to benefit from a good match, while the firm will still have the remaining M apprentices and can hope that some of them turn out to be productive matches. This is why the worker's expected surplus from continuing the match is $q\left(E\left[\mu | \mu \geq \max\{\mu^*, \mu_V^M\}\right] - \mu^*\right)$, while the firm's is $q\tilde{\mu}$, a smaller number.

[8] Note that M only enters equation (4.3) via μ_V^M, the V-th largest value of μ out of M draws. This number obviously doesn't exist in the case where $M < V$, in which case expressions such as $\max\{\mu^*, \mu_V^M\}$ would simplify to μ^*. So for the case that $M < V$, the optimal level of human capital H is indeed independent of M, as we noted earlier.

It is easy to verify that the optimal amount of human capital provided is positive. This is perhaps surprising, since it is the firm that pays the training costs in its entirety, but it has no use for training, as the worker's wage rises one-for-one with each unit of human capital. The reason the firm provides training anyway is that the worker can threaten to end the match, and uses this threat to extract a positive amount of training as a concession from the firm.

Together, the optimal number of workers trained M and amount of human capital transferred during training H are given by (4.3) and the inequalities $S(V, M^* - 1) \geq 0$ and $S(V, M^*) \leq 0$, where $S(\cdot)$ is defined in (4.2). Because of the restriction that M must be an integer, writing down closed-form solutions is quite tedious and not particularly informative. We prove in the appendix that $S(V, M)$ is strictly decreasing in M and converges to a negative number. This implies that the optimal number of apprentices trained is finite and unique, except for the knife-edge case where $S(V, M)$ is exactly equal to zero for some M, and the firm is indifferent between training the last apprentice or not.

In our view, the main purpose of this model is to perform comparative statics with respect to γ. In the appendix, we prove the following result:

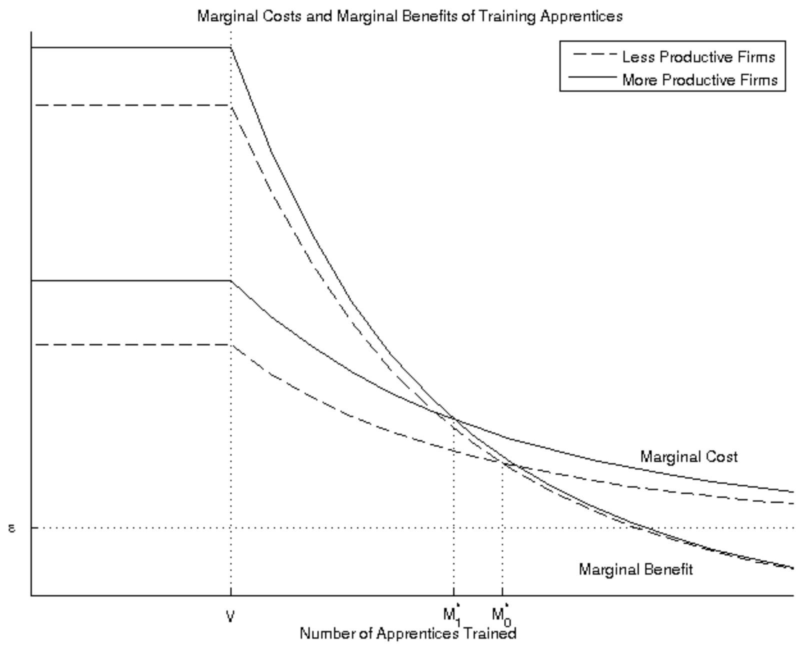
Proposition 1. *For any number of vacancies V, there exists a bound $\bar{b}$ such that, whenever $b > \bar{b}$, firms with productivity γ_1 will choose to train a weakly lower number of apprentices M than firms with productivity γ_0. Moreover, there exists a number of vacancies $\bar{V}$ and a number of apprentices trained $\bar{M}$ such that, whenever $V \geq \bar{V}$ and $M \leq \bar{M}$, firms with productivity γ_1 provide a strictly larger level of human capital H than firms with productivity γ_0.*[9]

The model's logic is summarized in Figure 4.1. The marginal benefit from training an extra apprentice is that the worker-firm match may turn out to be good enough to be retained. As long as a firm is considering to train fewer than V apprentices, each apprentice is as valuable as the next, and the marginal benefits of training an additional worker are flat. Once the number of apprentices surpasses V, marginal benefits converge to 0, because it becomes increasingly

[9]The restrictions on V and M are technical conditions required for our proof. We suspect that these conditions may not be necessary, but we have not been able to produce an alternative proof that doesn't require them.

likely that the firm has already trained enough apprentices to fill all vacancies; the benefit of an extra apprentice is merely that he may prove to form an even better match than one of the others. The marginal costs of training decrease with additional apprentices as the firm provides less training, but they can never fall below ϵ, so the optimal number of trained workers is finite. More productive firms are more sensitive to the quality of the match, so they are less inclined to retain workers with whom they form a mediocre match for the sake of avoiding the costs of a vacancy b. Therefore, training an extra worker is less attractive to the high-productivity firm, and the firm trains fewer workers but provides them with more training.

Figure 4.1: Marginal Costs and Marginal Benefits of Training Apprentices



In sum, the preceding analysis delivers three main empirical predictions, which will be tested in the following section:

- Workers who stay with the firm where they have received apprenticeship training should be in higher-quality matches, on average, than workers who have separated from their training firm. This prediction echoes Acemoglu (1997),

- Firms should never train a positive number of apprentices that is smaller than the number of vacancies they will have available; they should either hire at least V apprentices, or none at all.

- More productive firms should be running smaller training programs, have higher retention rates, and provide less human capital during training.

4.3 Data

We use data from the Austrian Social Security Database (ASSD), a matched employer-employee dataset containing the universe of private-sector employment spells in Austria between 1972 and 2012 (Zweimüller et al. (2009); see also section 2.2). Importantly, the database includes all private-sector apprenticeship spells. For each spell, we have information on spell duration and annual salary, as well as information on the location and 3-digit industry code of the firm and the age and gender of the worker. We construct our main dataset for analysis by considering all apprenticeship spells that last at least 8 quarters, are the only apprenticeship spell a worker has completed, and were begun when the apprentice was not yet 18 years old. We impose the last condition because the typical time to begin an apprenticeship is right after leaving school at approximately age 15, and we want to focus on individuals who do not have any significant pre-apprenticeship experience in the labor market. We track every individual who has completed an apprenticeship, recording his employer and earnings (if any) on a quarterly basis.[10] All told, we have information on 1,374,921 individuals, who have trained at 143,671 firms.

For parts of the analysis, we will merge this dataset with firm-fixed effects calculated from the Abowd, Kramarz, and Margolis (1999, AKM) wage decomposition described in Chapter 2.4. Since the ASSD does not include information on hours worked, a firm that offers lots of part-time work could easily be mistaken for a low-wage firm. For this reason, we exclude women from

[10] As noted above, we have wage information only on an annual basis. We convert it to a monthly basis by multiplying annual earnings by the number of days worked, multiplying by 12/365. This means that there is no within-year earnings growth for workers who remain at the same firm.

the wage decomposition, and calculate fixed effects only using data on men, for whom part-time work is rare. The AKM model includes years of experience and occupational status (blue- or white-collar worker) as covariates. Since earnings prior to 1994 are subject to censoring, this decomposition is based only on the years 1994–2012. In all, we have fixed effects for 81,273 firms, who have trained 541,459 apprentices.

While the ASSD is a rich source of employment and wage dynamics, its use has three drawbacks. First, we do not observe whether the worker passes the centralized exam at the end of the apprenticeship. This might introduce some error, but we are confident that it is not too severe, as approximately 90% of candidates pass the exam (WKO, 2016). Second, the dataset does not contain information on firm productivity. Below, we will use a series of proxies for firm productivity suggested by the previous literature. Third, the ASSD does not allow us to observe the content of training received, so we cannot test the prediction that more productive firms impart more human capital to their apprentices during training.

4.4 Empirical Analysis

The model's first prediction is that workers who remain in their training firm should be in a better match than workers who move. A first check on two proxies for match quality bears this out. We consider the first job after completion of an apprenticeship and record whether this job is with the training firm or some other firm.[11] In Table 4.1, we report results from regressions of log earnings on a dummy variable for staying with the training firm, controlling for experience, experience squared, and gender interacted with year effects.[12] The second column also controls for industry-fixed effects, but our preferred estimate is in column 3, which includes *firm*-fixed

[11] There is compulsory military service for men in Austria, and it typically takes place sometime after completion of the apprenticeship. While we do not observe military service directly, we do observe conspicuous gaps in the employment history for a large fraction of men right around the end of the apprenticeship. For the analysis of earnings and job durations, we have excluded all men who leave the labor force for a period between 180 and 550 days.

[12] Since wage information is only available annually, we restrict this analysis to quarters at least one year post-apprenticeship. For the earlier quarters, stayer earnings would reflect a mix of post-training earnings and the lower earnings paid during training.

Table 4.1: Log Earnings Regression at First Employment Spell

	(1)	(2)	(3)
Stayer Dummy	0.0488	0.0389	0.0653
	(0.0067)	(0.0072)	(0.0032)
Industry FE		Yes	
Firm FE			Yes
N	4,239,349	4,220,096	4,239,349
R^2	0.49	0.50	0.79

Each column reports results from a linear regression of log earnings on a dummy variable indicating whether the employment is still at the training firm. The unit of observation is a worker-quarter. Control variables include experience, experience squared, and gender interacted with year effects. Heteroskedasticity-robust standard errors are clustered at the firm level.

Table 4.2: Durations of First Earning Spell

	(1)	(2)	(3)
Stayer Dummy	0.8624	0.8823	0.8292
	(0.0102)	(0.0049)	(0.0167)
Stratified by Industry		Yes	
Stratified by Firm			Yes
N	745,526	739,370	745,526

Each column reports Hazard Ratios from a Cox Proportional Hazard model including a dummy variable indicating whether the employment is still at the training firm. The unit of observation is an employment spell. All models allow for different baseline hazards based by gender and year the spell began. Heteroskedasticity-robust standard errors are clustered at the firm level and transformed from coefficient estimates to hazard ratios via the Delta Method.

effects. Here, the effect is identified off recent graduates working for the same firm, with some having trained at the firm, and others elsewhere. The estimated effect of staying with the training is statistically significant and ranges from 3.9 to 6.5 per cent.

As a second proxy for match quality, we consider job duration. In Table 4.2, we present Hazard Ratios from a series of Cox Proportional Hazard models. All models are stratified to allow for different baseline hazards by gender and year; columns 2 and 3 further allow for different baseline hazards by industry and firm, respectively. The estimated effect is large and statistically significant; job spells at the training firms are estimated to have a 17% lower chance of ending in a given period.

To check whether firms indeed train at least as many apprentices as they anticipate having vacancies, we next focus on the subsample of firms for whom we have AKM fixed effects available. We also restrict attention to those apprentices who are found to be working one year after completion of the apprenticeship. For each firm, we have calculated the number of workers who have finished an apprenticeship with the firm in a given year, and the number of apprenticeship graduates from that cohort employed at the firm one year later. If our theory is correct, some firms should train a large number of apprentices but end up hiring only a few, whereas no firm should train just a few apprentices but then hire a lot. Table 4.3, which tabulates the number of recent graduates hired versus apprentices trained, provides support for the theory. Of interest are the off-diagonal entries. For instance, in over 16,000 firm-years, a firm would train between 2 and 4 apprentices but hire just one; by contrast, in fewer than 10,000 cases, a firm would hire between 2 and 4 apprentices even though it had trained exactly one. This asymmetry, present in all of the off-diagonal entries, is consistent with firms planning the number of apprentices in the way our model predicts. The fact that some firms do end up hiring more former apprentices than they had trained is most likely because the firm failed to hire as many apprentices as it would have liked, because some apprentices failed to complete their apprenticeships, or because a shock to the firm created more vacancies than anticipated.

The findings from Table 4.3 are visualized in Figure 4.2. To construct this figure, we have extracted from Table 1 only the firms who train a nonzero number of apprentices and those who employ a nonzero number of graduates one year later. The blue line shows, for each number of *apprentices trained*, the share of firms who employ a smaller number of graduates one year later. The red line shows, for each number of *graduates hired*, the share of firms who have trained a smaller number of apprentices one year earlier. We interpret the fact that the red line lies consistently below the blue line as evidence in line with our theory. Note also that the lines get quite noisy for larger numbers on the x-axis, as few firms train more than 5 apprentices in a given year. Below, when we analyze subgroups of firms, we therefore focus on firms with 5 or fewer apprentices trained and graduates hired.

Table 4.3: Number of apprentices trained and recent apprentices hired one year later. Firm-years from 1994 to 2012.

		Apprentices Trained							
		0	1	2-4	5-10	11-20	21-50	51+	Total
Graduates Hired	0	0	66094	4318	210	52	32	11	70717
	1	66474	91273	8174	100	16	7	0	166044
	2-4	6107	4572	17743	1000	17	5	0	29444
	5-10	492	121	605	2048	150	5	1	3422
	11-20	84	16	30	159	348	45	3	685
	21-50	29	3	8	6	39	115	3	203
	51+	7	0	0	0	0	2	23	32
	Total	73193	162079	30878	3523	622	211	41	270547

Table 4.4: Number of apprentices trained and recent apprentices hired one year later. Small firms with 5-10 non-apprentice workers only. Firm-years from 1994 to 2012.

		Apprentices Trained			
		0	1	2-4	Total
Graduates Hired	0	0	15722	480	16202
	1	10028	18226	713	28967
	2-4	293	276	470	1039
	Total	10323	34224	1663	46210

Table 4.5: Number of apprentices trained and recent apprentices hired one year later. Large firms with more than 50 non-apprentice workers only. Firm-years from 1994 to 2012.

		Apprentices Trained							
		0	1	2-4	5-10	11-20	21-50	51+	Total
Graduates Hired	0	0	10447	1717	172	50	30	11	12427
	1	20301	21829	3595	94	15	7	0	45841
	2-4	4270	2822	12032	970	17	5	0	20116
	5-10	456	118	596	2038	150	5	1	3364
	11-20	83	16	30	159	348	45	3	684
	21-50	29	3	8	6	39	115	3	203
	51+	7	0	0	0	0	2	23	32
	Total	25146	35235	17978	3439	619	209	41	82667

Figure 4.2: Number of Apprentices vs. Number of Graduates

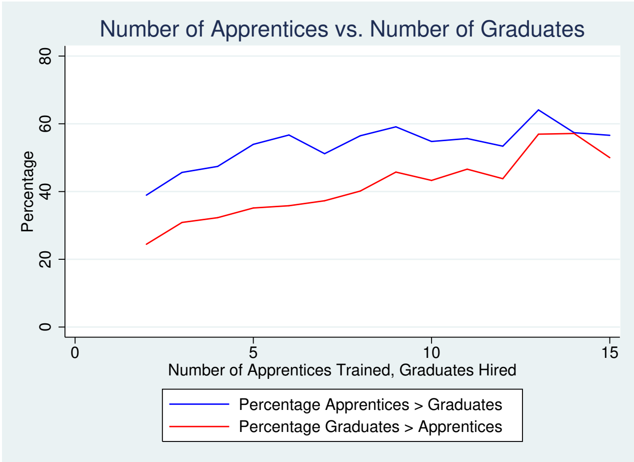
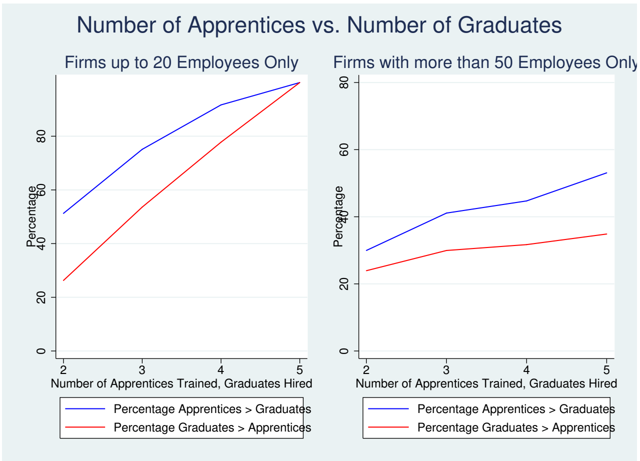


Figure 4.3: Number of Apprentices vs. Number of Graduates, Small and Large Firms



The model also predicts that unproductive firms should train more apprentices than they will have vacancies, while productive firms should train closer to the number of vacancies they will have. At the latter type of firms, training should be more intensive and more costly to the firm. While we do not have a good measure of firm productivity, we do have three proxies, the first of which is firm size. The literature on the dynamics of firm size has shown that large firms have grown large in the first place in part because of the use of more productive technologies than their competitors (Bartelsman, Haltiwanger, and Scarpetta, 2013). In line with this finding, Tables 4.4 and 4.5 reproduce Table 4.3 for small and large firms, respectively. The asymmetry

Table 4.6: Number of apprentices trained and recent apprentices hired one year later. Firms with an AKM firm effect below the median only. Firm-years from 1994 to 2012.

		Apprentices Trained							
		0	1	2-4	5-10	11-20	21-50	51+	Total
Graduates Hired	0	0	32752	2192	108	27	16	2	35097
	1	30270	40077	3711	61	7	6	0	74132
	2-4	2385	1774	5642	343	7	3	0	10154
	5-10	161	46	221	463	31	4	0	926
	11-20	19	7	13	80	71	14	2	206
	21-50	10	0	5	1	15	19	0	50
	51+	.	.	.	.	.	.	.	.
	Total	32845	74656	11784	1056	158	62	4	120565

in the off-diagonal entries is very pronounced for small firms; e.g., nearly three times as many firms train 2-4 apprentices but hire just one, than the other way around. By contrast, this asymmetry has all but disappeared for large firms. Figure 4.3 visualizes this difference: the blue and red lines are much further apart for small than for large firms, at least for the case of training 2 workers, which is the only one observed in large numbers for the small firms. This evidence supports our theory that small firms run relatively large training programs, while large firms only train as many apprentices as they will have vacancies to fill.

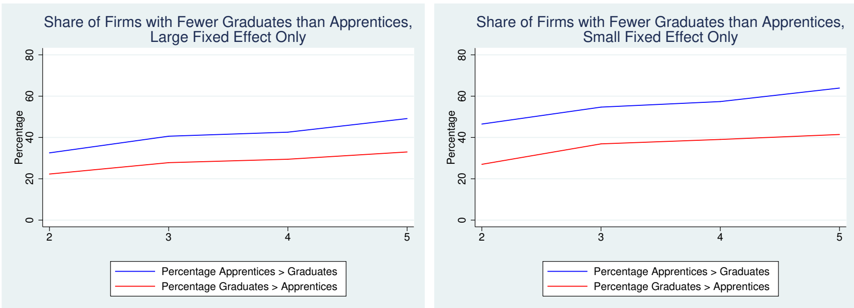
As a second proxy for productivity, we have calculated Abowd, Kramarz, and Margolis (1999) person- and firm-fixed effects in a wage regression. As Abowd, Kramarz, and Margolis (1999) show, high-wage firms also tend to be more productive and more profitable. Tables 4.6 and 4.7 split firms according to whether their firm effect is above or below the median. The patterns are similar to the large/small firm split: high-fixed effect firms are much less likely to train more graduates than they will hire. Again, Figure 4.4 confirms this impression.

A final proxy for productivity we consider is whether the firm is growing or shrinking. Growing firms have been shown to be more productive, on average, than shrinking firms (Foster, Haltiwanger, and Syverson, 2008). We have calculated, for each firm, the change in the non-apprentice workforce during the five-year window leading up to the completion of the apprenticeship, and classified firms according to whether the change in employment was positive

Table 4.7: Number of apprentices trained and recent apprentices hired one year later. Firms with an AKM firm effect above the median only. Firm-years from 1994 to 2012.

		Apprentices Trained							
		0	1	2-4	5-10	11-20	21-50	51+	Total
Graduates Hired	0	0	21407	1768	95	25	15	8	23318
	1	29577	41844	4011	38	9	1	0	75480
	2-4	3529	2601	11802	655	10	2	0	18599
	5-10	328	74	382	1584	119	1	1	2489
	11-20	65	9	17	79	277	31	1	479
	21-50	19	3	3	5	24	96	3	153
	51+	7	0	0	0	0	2	23	32
	Total	33525	65938	17983	2456	464	148	36	120550

Figure 4.4: Number of Apprentices vs. Number of Graduates, High- and Low-Fixed Effect Firms



or negative. In Tables 4.8 and 4.9, we see that the pattern of training a larger number of workers than are ultimately employed is more pronounced for the shrinking firms, as visualized in Figure 4.5. Once again, the evidence lends support to our theory.

Figure 4.5: Number of Apprentices vs. Number of Graduates, Growing and Shrinking Firms

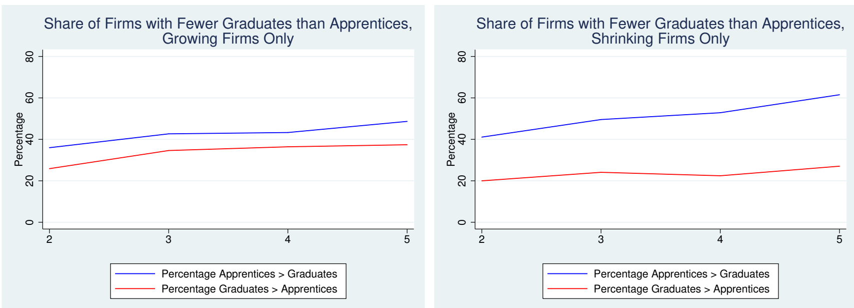


Table 4.8: Number of apprentices trained and recent apprentices hired one year later. Firms that have lost non-apprentice workers in the five years prior only. Firm-years from 1994 to 2012.

		Apprentices Trained							
		0	1	2-4	5-10	11-20	21-50	51+	Total
Graduates Hired	0	0	22735	1793	108	38	26	9	24709
	1	13779	28009	2822	44	9	3	0	44666
	2-4	994	1143	5781	403	6	1	0	8328
	5-10	58	14	134	625	74	3	0	908
	11-20	12	4	3	9	106	25	1	160
	21-50	1	1	2	0	2	37	1	44
	51+	0	0	0	0	0	0	10	10
	Total	14844	51906	10535	1189	235	95	21	78825

Table 4.9: Number of apprentices trained and recent apprentices hired one year later. Firms that have added non-apprentice workers in the five years prior only. Firm-years from 1994 to 2012.

		Apprentices Trained							
		0	1	2-4	5-10	11-20	21-50	51+	Total
Graduates Hired	0	0	25659	1788	69	9	5	1	27531
	1	27604	42897	4168	48	6	4	0	74727
	2-4	3312	2812	10206	524	6	3	0	16863
	5-10	298	96	432	1321	64	1	0	2212
	11-20	53	12	27	144	221	18	1	476
	21-50	17	2	6	6	37	76	2	146
	51+	2	0	0	0	0	1	11	14
	Total	31286	71478	16627	2112	343	108	15	121969

4.5 Conclusion

If training is not firm-specific, does it matter in which firm apprentices choose to receive their training? In this paper, we have constructed a simple model of general training, which differs from existing models in the literature in that it allows for firms that are heterogeneous by productivity. Using this framework, we are able to derive several testable predictions that imply that high-productivity firms will structure their apprenticeships differently: they will train fewer workers while providing them with more training. Using Austrian administrative data, we can test at least some of these predictions and find support for them. Notably, more productive firms indeed seem to run smaller training programs and retain more of their recent graduates, whereas less productive firms, if they train at all, have a greater trendency to train more apprentices in the knowledge that some of them will go elsewhere after the end of the apprenticeship.

Because the dataset used in this study does not contain information on the amount of human capital apprentices received during their training, we cannot test the prediction that more productive firms provide more intensive training. However, there is evidence that large firms provide more intensive training (Bardeleben, Beicht, and Fehér, 1995), and large firms tend to be more productive (Bartelsman, Haltiwanger, and Scarpetta, 2013). This evidence is consistent with our theory, but investigating the link between firm productivity and the level of training provided during an apprenticeship more fully would be a fruitful avenue for further research.

4.A Proof that the optimal number of apprentices trained is finite and unique

The proof proceeds in two steps. First, we prove that the benefits from training one additional apprentice $S(V, M)$ strictly decrease in the number of apprentices trained M. Second, we show that $S(V, M)$ must be negative for M sufficiently large. Together, these two conditions imply that the optimal number of apprentices to train must be finite and unique, except for the knife-edge case where $S(V, M) = 0$ for some M, such that a firm is exactly indifferent between training one more apprentice or not.

4.A.1 Proof that $S(V, M)$ is decreasing in M

To simplify the notation, let $S^+(V, M)$ denote the first two lines of (4.2), that is, the marginal benefit to the firm of training another apprentice. Then $S(V, M) = S^+(V, M) - H^2 - \epsilon$. The proof consistst of two steps: First, to show that $\frac{\Delta S^+(V,M)}{\Delta M} < 0$, and second, to show that $\left|\frac{\Delta H^2}{\Delta M}\right| < \left|\frac{\Delta S^+(V,M)}{\Delta M}\right|$. To show the first statement, define the function $g(x) = P(\mu \geq x) \cdot (E[\mu|\mu \geq x] - E[x])$, and note that we may write

$$S^+(V, M) = \beta g\left(\max\{\mu^*, \mu_V^M\}\right).$$

$g(\cdot)$ is decreasing in the following sense: let x_1 and x_2 both be random variables with associated distribution functions G_1 and G_2, respectively. Let $x_1 > x_2$ in the sense of first-order stochastic dominance; we are claiming that this implies $g(x_2) - g(x_1) > 0$. The proof is as follows:

$$
\begin{aligned}
g(x_2) - g(x_1) &= P(\mu \geq x_2) \cdot (E[\mu|\mu \geq x_2] - E[x_2]) - P(\mu \geq x_1) \cdot (E[\mu|\mu \geq x_1] - E[x_1]) \\
&= P(\mu \geq x_2)E[\mu|\mu \geq x_2] - P(\mu \geq x_1)E[\mu|\mu \geq x_1] + P(\mu \geq x_1)E[x_1] - P(\mu \geq x_2)E[x_2] \\
&= P(\mu \geq x_2)\frac{\int_{-\infty}^{\infty} \mu f(\mu)G_2(\mu)d\mu}{P(\mu \geq x_2)} - P(\mu \geq x_1)\frac{\int_{-\infty}^{\infty} \mu f(\mu)G_1(\mu)d\mu}{P(\mu \geq x_1)} \\
&+ \int_{-\infty}^{\infty} f(\mu)G_1(\mu)d\mu \cdot E[x_1] - \int_{-\infty}^{\infty} f(\mu)G_2(\mu)d\mu \cdot E[x_2] \\
&= \int_{-\infty}^{\infty} (\mu f(\mu)G_2(\mu) - \mu f(\mu)G_1(\mu))d\mu + \int_{-\infty}^{\infty} (f(\mu)G_1(\mu)E[x_1] - f(\mu)G_2(\mu)E[x_2])d\mu \\
&= \int_{-\infty}^{\infty} f(\mu)[G_2(\mu)(\mu - E[x_2]) - G_1(\mu)(\mu - E[x_1])]d\mu \\
&> 0,
\end{aligned}
$$

where the last line follows because $G_2(\mu) \geq G_1(\mu)$ and $E[x_2] < E[x_1]$. Since $\max\{\mu^*, \mu_V^M\}$ first-order stochastically dominates $\max\{\mu^*, \mu_V^{M-1}\}$, $g(\cdot)$ being decreasing implies that $\frac{\Delta S^+(V,M)}{\Delta M} = \beta\frac{\Delta g\left(\max\{\mu^*,\mu_V^M\}\right)}{\Delta M} < 0$.

So show the second statement, define $g'(x) = P(\mu \geq x) \cdot (E[\mu|\mu \geq x] - \mu^*)$, and apply the quadratic formula to (4.3) to obtain

$$
\begin{aligned}
H =& -\frac{1-\beta}{1+\beta}\beta g'\left(\max\{\mu^*, \mu_V^M\}\right) \\
&+ \underbrace{\sqrt{\left(\frac{1-\beta}{1+\beta}\right)^2 \beta^2 \left[g'\left(\max\{\mu^*, \mu_V^M\}\right)\right]^2 + \frac{1-\beta}{1+\beta}\beta g\left(\max\{\mu^*, \mu_V^M\}\right) - \frac{1-\beta}{1+\beta}\epsilon}}_{\equiv A} \qquad (4.A.1) \\
H^2 =& 2\left(\frac{1-\beta}{1+\beta}\right)^2 \beta^2 \left[g'\left(\max\{\mu^*, \mu_V^M\}\right)\right]^2 + \frac{1-\beta}{1+\beta}\beta g\left(\max\{\mu^*, \mu_V^M\}\right) \\
&- \frac{1-\beta}{1+\beta}\epsilon - 2\frac{1-\beta}{1+\beta}\beta g'\left(\max\{\mu^*, \mu_V^M\}\right) A.
\end{aligned}
$$

What we wish to show is that

$$\frac{\Delta H^2}{\Delta M} = \frac{\Delta H^2}{\Delta g'\left(\max\{\mu^*, \mu_V^M\}\right)} \frac{\Delta g'\left(\max\{\mu^*, \mu_V^M\}\right)}{\Delta M} + \frac{\Delta H^2}{\Delta g\left(\max\{\mu^*, \mu_V^M\}\right)} \frac{\Delta g\left(\max\{\mu^*, \mu_V^M\}\right)}{\Delta M}$$

is not too large. Applying the same logic that we used above to show that $g(\cdot)$ is decreasing, we may see that $g'(\cdot)$ is also decreasing. Moreover, inspection of (4.A.1) reveals that H, and by extension H^2, is decreasing in $g'\left(\max\{\mu^*, \mu_V^M\}\right)$. Therefore

$$\begin{aligned}
\frac{\Delta H^2}{\Delta M} &> \frac{\Delta H^2}{\Delta g\left(\max\{\mu^*, \mu_V^M\}\right)} \frac{\Delta g\left(\max\{\mu^*, \mu_V^M\}\right)}{\Delta M} \\
&= \left(\underbrace{\frac{1-\beta}{1+\beta}\beta}_{<\beta} - \underbrace{\frac{\Delta 2\frac{1-\beta}{1+\beta}\beta g'\left(\max\{\mu^*, \mu_V^M\}\right) A}{\Delta g\left(\max\{\mu^*, \mu_V^M\}\right)}}_{>0} \right) \frac{\Delta g\left(\max\{\mu^*, \mu_V^M\}\right)}{\Delta M} \\
&> \beta \frac{\Delta g\left(\max\{\mu^*, \mu_V^M\}\right)}{\Delta M} \\
&= \frac{\Delta S^+(V, M)}{\Delta M}.
\end{aligned}$$

So H^2 decreases in response to an increase in M, but $\Delta S^+(V, M)$ decreases by even more.

4.A.2 Proof that $S(V, M)$ must be negative for M sufficiently large

Since the marginal costs of training are bounded away from zero by the presence of ϵ, a sufficient condition is that the marginal benefits of training converge to zero as M approaches infinity: $\lim_{M\to\infty} S^+(V, M) = 0$. In turn, two sufficient conditions are that $\lim_{M\to\infty} P(\mu_V^M \leq \mu^*) = 0$

and that $\lim_{M\to\infty} E[\mu_V^M|\mu_V^M \geq \mu^*] = \mu_{max}$. If they hold, then

$$
\begin{aligned}
\lim_{M\to\infty} S^+(V,M) = {} & \underbrace{\lim_{M\to\infty} P(\mu_V^M \leq \mu^*)}_{=0} \cdot P(\mu \geq \mu^*) \cdot \beta\left(E[\mu|\mu \geq \mu^*] - \mu^*\right) \\
& + \underbrace{\lim_{M\to\infty} P(\mu_V^M \geq \mu^*)}_{=1} \cdot \underbrace{\lim_{M\to\infty} P(\mu \geq E[\mu_V^M|\mu_V^M \geq \mu^*])}_{=0} \\
& \cdot \beta \underbrace{\lim_{M\to\infty}\left(E[\mu|\mu \geq E[\mu_V^M|\mu_V^M \geq \mu^*]] - E[\mu_V^M|\mu_V^M \geq \mu^*]\right)}_{=\mu_{max}-\mu_{max}=0} \\
& = 0.
\end{aligned}
$$

To show the first condition, i.e. $\lim_{M\to\infty} P(\mu_V^M \leq \mu^*) = 0$, we first note that $P(\mu_V^M \leq \mu^*) = \sum_{i=0}^{V-1} F(\mu^*)^{M-i}(1-F(\mu^*))^i \binom{M}{i}$ (David and Nagaraja, 2005, p. 9). The sum on the right-hand side converges to zero if each summand also converges to zero. A sufficient condition for convergence is the ratio test, $\lim_{M\to\infty} \frac{P(\mu_V^{M+1} \leq \mu^*)}{P(\mu_V^M \leq \mu^*)} = L < 1$. Each summand passes this test:

$$
\begin{aligned}
\lim_{M\to\infty} \frac{F(\mu^*)^{M+1-i}(1-F(\mu^*))^i\binom{M+1}{i}}{F(\mu^*)^{M-i}(1-F(\mu^*))^i\binom{M}{i}} &= \lim_{M\to\infty} F(\mu^*)\frac{\binom{M+1}{i}}{\binom{M}{i}} \\
&= \lim_{M\to\infty} F(\mu^*)\frac{\frac{(M+1)!}{i!(M+1-i)!}}{\frac{M!}{i!(M-i)!}} \\
&= \lim_{M\to\infty} F(\mu^*)(M+1)\frac{(M-i)!}{(M+1-i)!} \\
&= \lim_{M\to\infty} F(\mu^*)\frac{M-i}{M+1-i} \\
&= F(\mu^*) < 1.
\end{aligned}
$$

To show the second condition, i.e. that $\lim_{M\to\infty} E[\mu_V^M|\mu_V^M \geq \mu^*] = \mu_{max}$, let $Q^M(x) = \sum_{i=0}^{V-1} F(x)^{M-i}(1-F(x))^i\binom{M}{i}$ be the CDF of μ_V^M. Note that $\lim_{M\to\infty} Q^M(x) = 0$ for all $x < \mu_{max}$, by the ratio test argument above. Of course, since $Q^M(x)$ is a distribution function,

$Q^M(\mu_{max}) = 1$. For the inverse function of Q^M, $\left(Q^M\right)^{-1}$, it follows that $\lim_{M\to\infty}\left(Q^M(x)\right)^{-1} = \mu_{max}$ for all $0 < x \leq 1$. This can be used to show the desired claim:

$$\begin{aligned}
\lim_{M\to\infty} E\left[\mu_V^M | \mu_V^M \geq \mu^*\right] &\geq \lim_{M\to\infty} E\left[\mu_V^M\right] \\
&= \lim_{M\to\infty} \int_{-\infty}^{\infty} x dQ^M(x)
\end{aligned}$$

Let $z = Q^M(x)$, so that $x = \left(Q^M(z)\right)^{-1}$. By a change of variables:

$$= \lim_{M\to\infty} \int_0^1 \left(Q^M(z)\right)^{-1} dz$$

Using the Dominated Convergence Theorem, we can pull the limit inside the integral:

$$\begin{aligned}
&= \int_0^1 \lim_{M\to\infty} \left(Q^M(z)\right)^{-1} dz \\
&= \int_0^1 \mu_{max} dz \\
&= \mu_{max}.
\end{aligned}$$

4.B Proof of Proposition 1

Let μ_1^* and μ_0^* denote the cut-off values μ^* for high- and low-productivity firms, respectively, and recall that $\mu^* = \frac{\bar{\mu}\gamma_1 - b}{\gamma}$. Therefore,

$$\begin{aligned}
\mu_1^* - \mu_0^* &= \bar{\mu} - \frac{b}{\gamma_1} - \frac{\bar{\mu}\gamma_1 - b}{\gamma_0} \\
\mu_1^* - \mu_0^* &= \bar{\mu}\left(1 - \frac{\gamma_1}{\gamma_0}\right) + b\left(\frac{1}{\gamma_0} - \frac{1}{\gamma_1}\right).
\end{aligned} \tag{4.B.1}$$

Now let $\bar{b} = \frac{\bar{\mu}\left(\frac{\gamma_1}{\gamma_0} - 1\right)}{\frac{1}{\gamma_0} - \frac{1}{\gamma_1}}$, which ensures that b is sufficiently large for the right-hand side of

(4.B.1) to be positive, and it follows that $\mu_1^* > \mu_0^*$.

To show the first part of Proposition 1, recall from appendix 4.A.1 that the marginal benefits of training an extra apprentice may be written as $S^+(V,M) = \beta g\left(\max\{\mu^*, \mu_V^M\}\right)$, and that we have shown the function $g(\cdot)$ to be decreasing. Since $\max\{\mu_1^*, \mu_V^M\} > \max\{\mu_0^*, \mu_V^M\}$ in expectation, it follows that $S^+(V,M)$ is smaller for high-productivity firms than for low-productivity firms. We also show below that more productive firms provide more human capital H than small firms. Therefore, $S(V,M) = S^+(V,M) - H^2 - \epsilon$ must be smaller for high-productivity firms than for low-productivity firms.

To show the second part of Proposition 1, recall from equation (4.A.1) that the optimal level of human capital is given by

$$
\begin{aligned}
H = &- \frac{1-\beta}{1+\beta}\beta g'\left(\max\{\mu^*, \mu_V^M\}\right) \\
&+ \sqrt{\left(\frac{1-\beta}{1+\beta}\right)^2 \beta^2 \left[g'\left(\max\{\mu^*, \mu_V^M\}\right)\right]^2 + \frac{1-\beta}{1+\beta}\beta g\left(\max\{\mu^*, \mu_V^M\}\right) - \frac{1-\beta}{1+\beta}\epsilon},
\end{aligned}
$$

which we may rewrite as

$$
H = - C + \sqrt{C^2 + D},
$$

where $C = \frac{1-\beta}{1+\beta}\beta g'\left(\max\{\mu^*, \mu_V^M\}\right)$ and $D = \frac{1-\beta}{1+\beta}\beta g\left(\max\{\mu^*, \mu_V^M\}\right) - \frac{1-\beta}{1+\beta}\epsilon$. Then,

$$
\begin{aligned}
\frac{\partial H}{\partial \mu^*} &= - \frac{\partial C}{\partial \mu^*} + \frac{C\frac{\partial C}{\partial \mu^*} + \frac{1}{2}\frac{\partial D}{\partial \mu^*}}{\sqrt{C^2 + D}} \\
&= \frac{-H}{\sqrt{C^2 + D}}\frac{\partial C}{\partial \mu^*} + \frac{\frac{1}{2}}{\sqrt{C^2 + D}}\frac{\partial D}{\partial \mu^*} \\
&\approx \frac{\frac{1}{2} - H}{\sqrt{C^2 + D}}\frac{\partial C}{\partial \mu^*},
\end{aligned}
$$

where the last line follows because

$$
\begin{aligned}
\frac{\partial C}{\partial \mu^*} =& \frac{1-\beta}{1+\beta}\beta \frac{\partial g'\left(\max\{\mu^*, \mu_V^M\}\right)}{\partial \mu^*} \\
=& \frac{1-\beta}{1+\beta}\beta \frac{\partial P(\mu \geq \max\{\mu^*, \mu_V^M\}) \cdot \left(E[\mu|\mu \geq \max\{\mu^*, \mu_V^M\}] - \mu^*\right)}{\partial \mu^*},
\end{aligned}
$$

and under the assumption that M and V are both large, such that $\mu* \approx E\left[\max\{\mu^*, \mu_V^M\}\right]$,

$$
\begin{aligned}
\approx& \frac{1-\beta}{1+\beta}\beta \frac{\partial P(\mu \geq \max\{\mu^*, \mu_V^M\}) \cdot \left(E[\mu|\mu \geq \max\{\mu^*, \mu_V^M\}] - E\left[\max\{\mu^*, \mu_V^M\}\right]\right)}{\partial \mu^*} \\
=& \frac{1-\beta}{1+\beta}\beta \frac{\partial g\left(\max\{\mu^*, \mu_V^M\}\right)}{\partial \mu^*} \\
=& \frac{\partial D}{\partial \mu^*}.
\end{aligned}
$$

Therefore, $\frac{\partial H}{\partial \mu^*} > 0$ as long as $H < \frac{1}{2}$. But H cannot be greater than $\frac{1}{2}$. To see this, assume $H \geq \frac{1}{2}$. Then:

$$
\begin{aligned}
0.5 \leq H &= -C + \sqrt{C^2 + D} \\
0.5 + C &\leq \sqrt{C^2 + D} \\
0.25 + C^2 &\leq C^2 + D \\
D &\geq C + 0.25.
\end{aligned}
$$

But

$$
\begin{aligned}
D &= \frac{1-\beta}{1+\beta}\beta g\left(\max\{\mu^*, \mu_V^M\}\right) - \frac{1-\beta}{1+\beta}\epsilon \\
&= \frac{1-\beta}{1+\beta}\beta P(\mu \geq \max\{\mu^*, \mu_V^M\}) \cdot \left(E[\mu|\mu \geq \max\{\mu^*, \mu_V^M\}] - E\left[\max\{\mu^*, \mu_V^M\}\right]\right) - \frac{1-\beta}{1+\beta}\epsilon \\
&\leq \frac{1-\beta}{1+\beta}\beta P(\mu \geq \max\{\mu^*, \mu_V^M\}) \cdot \left(E[\mu|\mu \geq \max\{\mu^*, \mu_V^M\}] - \mu^*\right) \\
&= \frac{1-\beta}{1+\beta}\beta g'\left(\max\{\mu^*, \mu_V^M\}\right) \\
&= C,
\end{aligned}
$$

a contradiction.

We have shown that H is increasing in μ^*. Since $\mu_1^* > \mu_0^*$, this implies that high-productivity firms provide larger amounts of training.

4.C Details of the Initial Meeting Stage

So far, we have neglected how apprentices and firms actually meet before the training stage. This is a potentially serious omission, since high-productivity firms provide lower levels of training, but a higher chance of remaining with the firm (by virtue of their smaller apprenticeship cohorts), and the production of larger rents of which workers can capture a fraction. Even within a firm type, different firms may be attractive to different degrees for prospective trainees because of differences in the number of vacancies available V. It would therefore seem reasonable that a firm which offers a less attractive training program should have to compensate its apprentices. The following model of the meeting stage captures this in a simple way.

Consider the case where there are an equal number of workers and vacancies.[13] We model

[13] The case in which there are more workers than vacancies is not materially different, as it merely reduces the value of the worker's outside option during bargaining. However, in case where there are more vacancies than workers, we have the additional complication that the firm knows that it might not fill all of its vacancies, which changes the expected value of each vacancy. We do not pursue this further.

the meeting stage as consisting of T steps, and focus on the case $T = 2$ here; the generalization to larger T follows below. In the final step, any workers still unmatched are randomly assigned to a vacancy. The worker and the firm now bargain over the expected surplus from the match. Since there is no more opportunity for matching past period T, the outside option for the firm is to hire one fewer apprentice than it had intended, while the worker's outside option is to enter the secondary market without any training.

With a positive expected net benefit from each apprenticeship, the worker and the firm will always agree to match, and all vacancies will be filled. Knowing that all other vacancies will be filled, the firm's surplus from the match is $S(V, M^* - 1)$, the marginal benefit of the last apprentice hired. The worker's surplus is the expected wage gains that he will experience by receiving training, namely

$$R(V, M^* - 1, \gamma) = H + \beta q(\tilde{\mu} - \mu^*).$$

Nash bargaining implies that, at the time the worker and firm meet, they agree to split the expected surplus, with the worker receiving

$$w_T = (1 - \beta)S(V, M^* - 1) - \beta(H + \beta q(\tilde{\mu} - \mu^*))$$

Since all agents are risk-neutral, there is no discounting, and contracts are perfectly enforcable, the model is silent on the time at which the surplus is split. For example, the apprentice might receive her share as a lump-sum payment before the apprenticeship starts, as renumeration during the apprenticeship, or as a promise of a raise should the match persist.

During stage $T - 1$, a fraction α of workers is randomly chosen to meet a firm that has a vacancy. Again, since it is clear that all other vacancies will be filled, the surplus that worker and firm have available to split is as in period T. As an outside option, each side can opt to pay a cost b and proceed to the final stage. Nash bargaining yields the worker's share

$$w_{T-1}(V, M^* - 1) = b + \beta(\bar{R} - R(V, M^* - 1, \gamma))$$

where $\bar{R}$ is the average worker's surplus of the last apprentice hired at all firms, weighted by the number of vacancies. There are two caveats. First, a necessary condition is that $R(V, M^* - 1, \gamma) + 2 \geq \bar{R}$, so that the worker does not find it advantageous to pass on the match and wait for the final period. Second, it must be the case that $\alpha\,(S(V, M^* - 1) - w_{T-1}) + (1 - \alpha)\,(S(V, M^* - 1) - w_T - b) \geq 0$, so that, ex ante, the firm can expect to retain some surplus after paying out the worker. We assume that both conditions holds for all (V, M^*).

Specifying the meeting stage in this way preserves the structure of the model while assuring that all vacancies are filled. It also allows for partial compensation for workers who are matched with firms with below-average expected surplus R.

Bibliography

Abowd, John, Robert Creecy, and Francis Kramarz (2002). *Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data.* Longitudinal Employer-Household Dynamics Technical Paper 2002-06. Center for Economic Studies, U.S. Census Bureau.

Abowd, John, Francis Kramarz, and David Margolis (1999). "High Wage Workers and High Wage Firms". In: *Econometrica* 67.2, pp. 251–333.

Abowd, John, Kevin McKinney, and Ian Schmutte (2019). "Modeling Endogenous Mobility in Earnings Determination". In: *Journal of Business and Economic Statistics* 37.3, pp. 405–418.

Acemoglu, Daron (1997). "Training and Innovation in an Imperfect Labour Market". In: *Review of Economic Studies* 64.3, pp. 445–464.

Acemoglu, Daron and Jörn-Steffen Pischke (1998). "Why Do Firms Train? Theory and Evidence". In: *Quarterly Journal of Economics* 113.1, pp. 79–119.

— (1999). "The Structure of Wages and Investment in General Training". In: *Journal of Political Economy* 107.3, pp. 539–572.

— (2003). "Minimum Wages and On-the-Job Training". In: *Worker Well-Being and Public Policy.* Ed. by Solomon Polachek. Research in Labor Economics 22. Emerald Group. Chap. 5, pp. 159–202.

Adams, Renée and Daniel Ferreira (2009). "Women in the boardroom and their impact on governance and performance". In: *Journal of Financial Economics* 94.2, pp. 291–309.

Allianz für Aus- und Weiterbildung (2014). *Alliance for Initial and Further Training 2015 – 2018.*

Altonji, Joseph, Anthony Smith, and Ivan Vidangos (2013). “Modeling Earnings Dynamics”. In: *Econometrica* 81.4, pp. 1395–1454.

Andrews, Martyn et al. (2008). “High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?” In: *Journal of the Royal Statistical Society. Series A* 171.3, pp. 673–697.

— (2012). “High Wage Workers Match With High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias”. In: *Economics Letters* 117.3, pp. 824–827.

Autor, David (2001). “Why Do Temporary Help Firms Provide Free General Skills Training?” In: *Quarterly Journal of Economics* 116.4, pp. 1409–1448.

Bagger, Jesper, François Fontaine, et al. (2014). “Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics”. In: *American Economic Review* 104.6, pp. 1551–1596.

Bagger, Jesper, Kenneth Sørensen, and Rune Vejlin (2013). “Wage Sorting Trends”. In: *Economics Letters* 118.1, pp. 63–67.

Bardeleben, Richard von, Ursula Beicht, and Kálmán Fehér (1995). *Betriebliche Kosten und Nutzen der Ausbildung. Repräsentative Ergebnisse aus Industrie, Handel und Handwerk.* Berichte zur Beruflichen Bildung 187. Bielefeld: Bertelsmann.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta (2013). “Cross-Country Differences in Productivity: The Role of Allocation and Selection”. In: *American Economic Review* 103.1, pp. 305–334.

Becker, Gary (1964). *Human Capital: A Theoretical And Empirical Analysis, With Special Reference To Education.* Chicago: University of Chicago Press.

Beicht, Ursula and Günter Walden (2002). “Wirtschaftlichere Durchführung der Berufsausbildung - Untersuchungsergebnisse zu den Ausbildungskosten der Betriebe”. In: *Berufsbildung in Wissenschaft und Praxis* 6, pp. 38–43.

Bertrand, Marianne and Antoinette Schoar (2003). "Managing with Style: The Effect of Managers on Firm Policies". In: *Quarterly Journal of Economics* 118.4, pp. 1169–1208.

Bhuller, Manudeep, Magne Mogstad, and Kjell Salvanes (2017). "Life-Cycle Earnings, Education Premiums, and Internal Rates of Return". In: *Journal of Labor Economics* 35.4, pp. 993–1030.

Blau, Francine and Lawrence Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". In: *Journal of Economic Literature* 55.3, pp. 789–865.

Burdett, Kenneth and Dale Mortensen (1998). "Wage Differentials, Employer Size, and Unemployment". In: *International Economic Review* 39.2, pp. 257–273.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek (2014). "Gender, Competitiveness, and Career Choices". In: *Quarterly Journal of Economics* 129.3, pp. 1409–1447.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016). "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women". In: *Quarterly Journal of Economics* 131.2, pp. 633–686.

Card, David, Jörg Heining, and Patrick Kline (2013). "Workplace Heterogeneity and the Rise of West German Wage Inequality". In: *Quarterly Journal of Economics* 128.3, pp. 967–1015.

Croson, Rachel and Uri Gneezy (2009). "Gender Differences in Preferences". In: *Journal of Economic Literature* 47.2, pp. 448–474.

David, Herbert and Haikady Nagaraja (2005). *Order Statistics*. 3rd ed. Wiley & Sons.

Department for Business, Innovation and Skills (2015). *English Apprenticeships: Our 2020 Vision*. URL: `https://www.gov.uk/government/publications/apprenticeships-in-england-vision-for-2020`.

Dustmann, Christian and Uta Schönberg (2009). "Training and Union Wages". In: *Review of Economics and Statistics* 91.2, pp. 363–376.

Eeckhout, Jan and Philipp Kircher (2011). "Identifying Sorting – In Theory". In: *Review of Economic Studies* 78.3, pp. 872–906.

Fink, Martina et al. (2010). *Extracting Firm Information from Administrative Records: The ASSD Firm Panel*. Working Paper 2010-04. NRN.

Foster, Lucia, John Haltiwanger, and Chad Syverson (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" In: *American Economic Review* 98.1, pp. 394–425.

Gagliarducci, Stefano and Daniele Paserman (2015). "The Effect of Female Leadership on Establishment and Employee Outcomes: Evidence from Linked Employer-Employee Data". In: *Gender Convergence in the Labor Market*. Ed. by Solomon Polachek, Konstantinos Tatsiramos, and Klaus Zimmermann. Bingley, UK: Emerald Publishing, pp. 343–375.

Gelbach, Jonah (2016). "When Do Covariates Matter? And Which Ones, And How Much?" In: *Journal of Labor Economics* 34.2, pp. 509–534.

Goux, Dominique and Eric Maurin (1999). "Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data". In: *Journal of Labor Economics* 17.3, pp. 492–533.

Gruetter, Max and Rafael Lalive (2009). "The Importance of Firms in Wage Determination". In: *Labour Economics* 16.2, pp. 149–160.

Henderson, Charles et al. (1959). "The Estimation of Environmental and Genetic Trends from Records Subject to Culling". In: *Biometrics* 15.2, pp. 192–218.

Jackson, Kirabo (2013). "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers". In: *Review of Economics and Statistics* 95.4, pp. 1096–1116.

Jiang, Jiming (1996). "REML Estimation: Asymptotic Behavior and Related Topics". In: *Annals of Statistics* 24.1, pp. 255–286.

Jovanovic, Boyan (1979). "Job Matching and the Theory of Turnover". In: *Journal of Political Economy* 87.5, pp. 972–990.

Kessler, Anke and Christoph Lülfesmann (2006). "The Theory of Human Capital Revisited: on the Interaction of General and Specific Investiments". In: *Economic Journal* 116, pp. 903–923.

Lentz, Rasmus and Dale Mortensen (2010). "Labor Market Models of Worker and Firm Heterogeneity". In: *Annual Review of Economics* 2, pp. 577–602.

Loewenstein, Mark and James Spletzer (1998). "Dividing the Costs and Returns to General Training". In: *Journal of Labor Economics* 16.1, pp. 142–171.

Lopes de Melo, Rafael (2018). "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence". In: *Journal of Political Economy* 126.1, pp. 313–346.

Matsa, David and Amalia Miller (2013). "A Female Style in Corporate Leadership? Evidence from Quotas". In: *American Economic Journal: Applied Economics* 5.3, pp. 136–169.

Meghir, Costas and Luigi Pistaferri (2004). "Income Variance Dynamics and Heterogeneity". In: *Econometrica* 72.1, pp. 1–32.

Mincer, Jacob (1974). *Schooling, Experience, and Earnings.* New York: National Bureau of Economic Research.

Mortensen, Dale and Christopher Pissarides (1994). "Job Creation and Job Destruction in the Theory of Unemployment". In: *Review of Economic Studies* 61.3, pp. 397–415.

Perry, Samuel (2013). "She Works Hard(er) for the Money: Gender, Fundraising, and Employment in Evangelical Parachurch Organizations". In: *Sociology of Religion* 74.3, pp. 392–415.

Powell, David (2020). "Quantile Treatment Effects in the Presence of Covariates". In: *Review of Economics and Statistics* 102.5, pp. 994–1005.

Romano, Joseph, Azeem Shaikh, and Michael Wolf (2010). "Multiple Testing". In: *The New Palgrave Dictionary of Economics.* Ed. by Steven Durlauf and Lawrence Blume. Basingstoke: Palgrave Macmillan.

Searle, Shayle, George Casella, and Charles McCulloch (1992). *Variance Components.* Hoboken, NJ: Wiley.

Silverman, Irwin (2003). "Gender Differences in Delay of Gratification: A Meta-Analysis". In: *Sex Roles* 49.9, pp. 451–463.

Smith, Nina, Valdemar Smith, and Mette Verner (2006). “Do women in top management affect firm performance? A panel study of 2,500 Danish firms”. In: *International Journal of Productivity and Performance Management* 55.7, pp. 569–593.

Sørensen, Torben and Rune Vejlin (2013). “The Importance of Worker, Firm and Match Effects in the Formation of Wages”. In: *Empirical Economics* 45.1, pp. 435–464.

Syverson, Chad (2004). “Product Substitutability and Productivity Dispersion”. In: *Review of Economics and Statistics* 86.2, pp. 534–550.

— (2011). “What Determines Productivity?” In: *Journal of Economic Literature* 49.2, pp. 326–365.

Topel, Robert (1991). “Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority”. In: *Journal of Political Economy* 99.1, pp. 145–176.

Topel, Robert and Michael Ward (1992). “Job Mobility and the Careers of Young Men”. In: *Quarterly Journal of Economics* 107.2, pp. 439–479.

Weber, Andrea and Christine Zulehner (2010). “Female Hires and the Success of Start-up Firms”. In: *American Economic Review* 100.2, pp. 358–361.

White House (2016). *Investing $90 Million through ApprenticeshipUSA to Expand Proven Pathways into the Middle Class.* URL: `https://www.whitehouse.gov/the-press-office/2016/04/21/fact-sheet-investing-90-million-through-apprenticeshipusa-expand-proven`.

Witkofsky, Viktor (2012). “Estimation, Testing, and Prediction Regions of the Fixed and Random Effects by Solving Henderson’s Mixed Model Equations”. In: *Measurement Science Review* 12.6, pp. 234–248.

WKO (2016). *Über 90 Prozent der Lehrabsolventen mit erfolgreicher Lehrabschlussprüfung.* URL: `https://www.wko.at/Content.Node/Service/Bildung-und-Lehre/Lehre/Lehrlingsstatistik/Ueber-90-Prozent-der-Lehrabsolventen-mit-erfolgreicher-Le.html`.

Wolfers, Justin (2006). “Diagnosing Discrimination: Stock Returns and CEO Gender”. In: *Journal of the European Economic Association* 4.2, pp. 531–541.

Woodcock, Simon (2008). “Wage Differentials in the Presence of Unobserved Worker, Firm, and Match Heterogeneity”. In: *Labour Economics* 15.4, pp. 771–793.

— (2015). “Match Effects”. In: *Research in Economics* 69.1, pp. 100–121.

Zweimüller, Josef et al. (2009). *The Austrian Social Security Database.* NRN Working Paper 0903.

Jasper Haller

Education

August 2021	Postgraduate Diploma in EU Competition Law **King's College London**, UK Grade: Pass with Merit
Since September 2013	Ph.D. Student in Economics, **University of Mannheim**, Germany Advisor: Prof. Dr. Andrea Weber
August 2013	Master of Science in Economic Research **University of Mannheim**, Germany Thesis: "Labor Market Responses to a Parental Benefit Reform" Advisor: Prof. Dr. Andrea Weber GPA: 1.6 (on a scale from 1 to 6, with 1 being best)
May 2011	Bachelor of Arts in Political Economy, *magna cum laude* **Williams College**, USA Thesis: "Medicaid Eligibility Policy in the States" Advisor: Prof. Lara Shore-Sheppard GPA: 3.89 (on a scale from 1 to 4, with 4 being best)
Fall 2009	Exchange Semester at the **West African Research Center** Dakar, Senegal
May 2006	Abitur, **Fichtenberg-Oberschule**, Berlin, Germany GPA: 1.1 (on a scale from 1 to 6, with 1 being best)

Scholarships and Awards

Sep 2010	Prof. Fred Greene Scholarship, Williams College