In the Right Place at the Wrong Time –

The Role of Firms and Luck in Young Workers' Careers

Till von Wachter*
University of California Berkeley

JOB MARKET PAPER

November 2002

Abstract: Do early job losses permanently reduce the earnings and career prospects of young workers? Simple estimates may overstate the true effects of early displacements, especially if less able workers sort into firms with high turnover rates. The bias from initial assignment of workers *between* firms is compounded by biases from selection *within* firms, which arise if employers selectively displace their least able workers, or if workers move voluntarily to take better jobs. This paper uses longitudinal social security data on German apprentices and their training firms to obtain estimates of the long-term effects of an early job loss that account for nonrandom assignment between firms and selection within firms. I use differences over time in the fraction of graduating apprentices that are retained by the training firm as an instrument for job displacement. These should reflect exogenous changes in firm-specific labor demand that are independent of individual ability or permanent firm characteristics. Using this strategy, I find that wage losses from leaving the training firm at graduation are initially strong but fade within the first five years in the labor market. The results also confirm an important influence of voluntary mobility and of initial sorting matching trainees to firms. Both of these factors are likely to confound results of previous studies of early job mobility lacking information on the demand side.

---

* Contact: vwachter@econ.berkeley.edu. I would like to thank David Card for constant guidance and support throughout this project. I am indebted to Stefan Bender for countless hours spent compiling the data set used in the paper. I would also like to thank Ken Chay, Olivier Deschenes, David Lee, David Levine, Justin McCrary, Daniel McFadden, Tiago Ribeiro, Jesse Rothstein, Emmanuel Saez, Diane Whitmore and seminar participants at Berkeley for helpful comments and suggestions. The Institut fuer Arbeitsmarkts- und Berufsforschung (IAB) has kindly hosted me and provided access to the administrative data set. I am grateful to the Institute for Labor and Employment for financial support. All remaining errors are my own.

1. Introduction

There are two competing views of the role of early job mobility in young workers' careers. One view, shared by many economists, is that early job mobility plays an important role in early career development and wage growth. An alternative view that has informed policy proposals in the past highlights the potential costs of early job mobility. For young Americans the rate of job change is indeed very high, a fact that has been interpreted as evidence of beneficial job search (Topel and Ward 1992). But it has also been argued that the unstructured transition from school to work in the US labor market leads to excess mobility and slows the rate of human capital accumulation (Ryan 2001). In fact, we know that young workers have high displacement rates (Farber 1993) and suffer the largest wage declines in recessions (Blanchflower and Oswald 1994). Consistent with the more negative view, studies of early job displacement typically find persistent wage losses for young displaced workers (Kletzer and Fairlie 2001, Gustafson 1998).

The paper presents estimates of the long-term wage losses suffered by young German workers who leave their training firm at the end of an apprenticeship. Similar to what has been found for the United States, simple comparisons of leavers and stayers suggest that there are large permanent costs of displacement – on the order of 10 percent after 5 years. These comparisons, however, ignore two critical issues suggesting that simple estimates overstate wage losses. First, it is widely recognized that leavers may be adversely selected (Gibbons and Katz 1991). A second issue that has received less attention in the literature is that the sample of leavers is disproportionately drawn from firms with high turnover rates. To the extent that high-turnover firms attract lower-quality apprentices, or offer lower-quality training, the nonrandom nature of the displaced worker pool is a problem. A third issue – particularly important for young workers – is that leavers include both involuntary movers and those who moved voluntarily. Since voluntary movers tend to benefit from mobility, this would lead simple estimates to understate the effects of displacements.

Ideally, what is needed to identify the causal effect of displacement in this environment is exogenous variation in firm-specific demand for apprentices. As a proxy for this, I use the fraction of apprentices in the same cohort at the same firm who leave the firm at the end of training. By pooling data for several cohorts and adding firm fixed effects, the instrument represents year-to-year variation in the fraction of apprentices retained by each firm. This instrument is clearly orthogonal to permanent characteristics of the firm, and to any individual-specific demand side shocks, such as adverse selection or learning effects. It may still reflect some variation in supply side opportunities for the apprentices of a given firm in a given cohort. Thus, I consider a second instrumental variable, based on the fraction of the trainees' cohort who experience a spell of unemployment at the end of their apprenticeship. The inclusion of firm fixed effects also controls for any bias from initial sorting of workers into firms based on unobserved ability.

The sample consists of all graduates from the German apprenticeship system in the period from 1992 to 1994 who are observed working at least once in the first five years after training. In Germany, more than two thirds of recent cohorts participate in apprenticeship training programs that last on average two years and include both formal and practical training. About 35% of apprentices leave their training firm at graduation, suggesting adverse selection of workers is potentially an important problem. Initial sorting of workers into different types of firms is relevant as well, since firms provide different amounts of training and offer different career prospects as evident from variation in turnover rates. Moreover, high mobility of apprentices in the years following training suggests that some of those leaving their training firms move voluntarily. Thus, post-training mobility occurs in a rich environment with voluntary and involuntary mobility, adverse selection, and nonrandom sorting of workers into their training firm.

Using an instrumental variables (IV) estimator based on random firm-level fluctuations in retention rates, I find that involuntarily displaced trainees have initially lower wages than those who stayed, but that these losses disappear within five years after the end of training. Only the wage losses of workers leaving very large training firms have a persistent component, consistent with the presence of firm-size wage differentials or internal labor markets. Estimates accounting for nonrandom selection into and out of firms thus do not imply permanent negative effects of job losses. Understanding the discrepancy between these and the simple OLS estimates requires closer examination of the different confounding factors. Alternative estimates of wage losses given by OLS with fixed effects, IV, or IV with fixed effects address different sources of selection within firms or initial sorting between firms. Comparison between these estimates therefore helps to disentangle the separate impacts of sorting and selection. Moreover, each of these confounding factors is closely related to a different theory of job and wage mobility. Thus, the comparison of different estimators also provides a way to assess the importance of the basic models of early job mobility among young workers.

To make the comparison between estimators and theoretical implications explicit, I present a straightforward model of wage determination that captures the basic theories of early job mobility in a unified framework. Using this model to interpret the empirical results, I conclude that standard job search theory provides a good explanation for the incidence of voluntary mobility and for the patterns of wage losses observed for involuntarily displaced workers. In addition, the fact that training firm fixed effects matter for both OLS and IV estimates suggest that higher ability workers are sorted into lower-turnover firms at the start of apprenticeships. Lower ability workers are also more likely to be released by their training firms at the end of training. This could be explained by several different models, including adverse selection models or models that focus on the wage rigidities present in the German labor market.

The next section outlines the model of wage determination, relates it to theories of job mobility, and uses it to interpret the bias of OLS and to outline the estimation strategy. The third section describes the matched worker-firm data set, and compares the German apprenticeship system to the US labor market. The fourth section presents the basic empirical results and a detailed sensitivity analysis. The fifth section discusses the empirical findings in light of models of job and wage mobility, while the sixth presents more evidence on firm heterogeneity. The last section concludes and indicates areas for future research.

2. Estimation of Wage Losses and Theories of Job Mobility

Even though it is a well-documented feature of job mobility, standard models of the labor market do not predict that job losers experience wage declines. Several explanations have been proposed in the literature, each focusing on a separate aspect of job mobility. However, most of the mechanisms emphasized by different theories are likely to occur simultaneously in the labor market. The following statistical model of wage determination helps distinguish true causal effects of displacements from potential confounding factors in such a complex environment.

2.1. Wage Determination and Theories of Job Mobility

Consider a class of models in which young workers' real log wages are a function of their innate skills, a_i, and of their mobility status after their last job, $D_{i0} = V_{i0} + I_{i0}$. Mobility can be either voluntary ($V_{i0} = 1$) or involuntary ($I_{i0} = 1$); denote the gain or loss from voluntary and involuntary mobility t periods after a job change as δ_{Vt} and δ_{It}, respectively.[1] The goal of the analysis is to obtain an estimate of δ_{It}, the wage loss from a job displacement over time. However, as in many applications, suppose it is not known whether a job change was voluntary or not, that is, only D_{i0} is known, and neither V_{i0} nor I_{i0} is observed separately. The process determining wages t periods after a job change then is $w_{it} = \delta_{It}D_{i0} + (\delta_{Vt} - \delta_{It})V_{i0} + a_i + \varepsilon_{it}$. To capture that workers may be sorted among their *initial* employers, the firms that in the present application provide training, this can be rewritten as

$$w_{it} = \delta_{It}D_{i0} + (\delta_{Vt} - \delta_{It})V_{i0} + (a_i - \bar{a}_{j(i)}) + \bar{a}_{j(i)} + \varepsilon_{it}, \qquad (1)$$

where $\bar{a}_{j(i)}$ is average ability of workers at firm j that trained individual i, and ε_{it} is a random disturbance term. In this formulation, wages are determined by mobility status, an individual component of ability relative to the training firm's average, $a_i - \bar{a}_{j(i)}$, and a firm specific component of ability, $\bar{a}_{j(i)}$, neither of which is usually observed by the econometrician. This basic model is able to incorporate the main theories of job and wage mobility, whose main assumptions

[1] Note that both voluntary and involuntary mobility could lead to gains or losses for different workers. In this case, one could reinterpret δ_{Vt} and δ_{It} as the *average* gain or loss from voluntary and involuntary mobility, respectively.

and implications are summarized in Table 1.[2] Each theory has implications for different components of equation (1), and this will be helpful in the empirical analysis.

A widely cited theoretical explanation for wage losses of displaced workers has been the presence of adverse selection in the labor market. The basic idea, put forward by Gibbons and Katz (1991) and already present in Greenwald (1986), is that in a world in which only current employers are informed about a worker's true ability a displacement may be perceived as a negative signal about a worker by other employers. In equilibrium, the firm displaces less able workers and these workers get paid according to their lower expected ability. Thus, displaced workers suffer wage losses even if job changes themselves have no direct effect on wages (i.e., $\delta_{It} = 0$). In terms of equation (1), adverse selection implies that movers are likely to be the least able workers within a firm, i.e., $\operatorname{cov}(a_i - \overline{a}_{j(i)}, D_{i0}) < 0$. By raising the average ability of the pool of changing workers, equilibrium is sustained by the presence of individuals moving for exogenous reasons. Exogenous mobility has been featured prominently in tests of the adverse selection hypothesis as well. For example, Gibbons and Katz find that wage losses are higher for workers displaced by lay-offs compared to those displaced from plant closings, as the latter should be less selected.[3] Their basic insight can be generalized, and this will be taken up in the next section.

In a recent study, Krashinsky (2002) finds that among mature workers the differences in wage losses between workers displaced by plant closings and lay-offs found by Gibbons and Katz (1991) is partly driven by differences in firm size of pre-displacement employers.[4] Krashinsky's findings suggest that adverse selection may matter less for older workers. This is not surprising if markets

---

[2] While most of these theories could be integrated into richer models explaining a broader set of facts, the following discussion concentrates on the core contribution of each theory to obtain meaningful predictions.

[3] Using the Displaced Worker supplement from the Current Population Survey (CPS), they also find that differences in wage losses increase with tenure, claiming that this strengthens their results since for higher tenured workers firms should have more information. However, this is not implied by learning models in which for high-tenured and older workers the market should have learned more and thus the adverse selection problem is mitigated. As discussed below, adverse selection should be strongest for young and low-tenured workers.

[4] This is not surprising, since small firms are more likely to close in face of economic shocks, while larger firms are more likely to reduce their work force. Laid-off workers should thus tend to loose any wage premia or rents they earned from working at larger firms. Using the National Longitudinal Study of Youth (NLSY), Krashinsky (2002) shows that among workers not changing firm-size differences in wage-losses among workers moving due to plant-closings and lay-offs are small and not statistically significant.

continuously learn about workers' ability, as suggested by Farber and Gibbons (1996). In this case wages and career histories of older workers reflect their skills, and the information contained in any additional single signal such as a displacement is small.[5] However, it remains an important problem for young workers, since for them the market should still be learning about skills and motivation. The idea that learning and adverse selection matters most for the young is also implicit in Acemoglu and Pischke (1998), who develop a model of the German apprenticeship system in which employers' monopsony rents generated by private information about young workers encourage them to pay for general training. Since the empirical application focuses on young German apprentices, the hypothesis of adverse selection is particularly relevant for the present paper.[6]

Krashinsky's results also highlight the need to control for firm characteristics in the study of displaced workers. This point is related to a deeper issue raised by Gibbons and Katz (1991) themselves and further developed in a sequence of papers (Gibbons and Katz 1992, Gibbons, Katz, Lemieux, and Parent 2002) – namely that firms may differ systematically and that observed mobility and wage changes may be driven by a sorting process of heterogeneous workers into heterogeneous firms. Sorting is a particular problem for the study of displaced workers if less able workers are hired by firms with higher turnover rates. If workers are initially assigned to firms in this way, then not only are movers not comparable to stayers in general, but movers and stayers are not comparable across different types of firms.

That workers select themselves into firms based on ability is suggested by Abowd, Kramarz, and Margolis (1999), who find that differences in workers' ability levels explain a large fraction of

---

[5] While Farber and Gibbons (1996) concentrate on symmetric learning, a similar result holds for asymmetric learning. The point can be made formally in a model of asymmetric information where wages of retained workers are partly determined by their marginal product, such that wages are partially revealing about worker's ability (this could be motivated by rent sharing or by comparative advantage – more able workers have to be assigned to better jobs to be productive, as in Waldman 1984). If employers gradually learn about workers, wage losses due to adverse selection decline with tenure as the market realizes that past wages are increasingly a function of true ability. Moreover, the variance of the prediction declines with tenure as well, making it less probable a high-tenured worker is displaced because of a bad signal to the firm (von Wachter 2001a). The same point can be made using labor market experience as well.

[6] Adverse selection has featured centrally in the debate on why firms pay for general training in Germany. The main alternative explanation has been the role of labor market institutions such as unions (Dustmann and Schoenberg 2002) or firing costs. The present paper stays aloof of this debate and focuses instead on the fundamental determinants of job and wage mobility.

wage-differences among firms.[7] Similarly, some firms may value job stability more than others and try to structure wage incentives accordingly. Since firms with low turnover rates offer better career opportunities and have greater incentives to invest into their work force in the form of high-quality training, these firms are likely to attract the most able workers.[8] While this is a key hypothesis in several theoretical models of turnover and wage-profiles (e.g., Salop and Salop 1976, Weiss and Wang 1998), until recently little was known about differences in average mobility rates and tenure profiles between firms. Taking into account the potential bias from selection of workers into firms, Margolis (1995) shows that the return to seniority indeed varies considerably between firms. Consistent with this result, the recent literature on turnover and job-creation finds that there is considerable heterogeneity in turnover rates and growth rates among establishments (e.g., Davis and Haltiwanger 1990, Abowd, Corbel, and Kramarz 1999).[9]

If firms thoroughly screen young workers during the hiring process, better firms may be able to attract the most skilled and motivated young workers and initial assignment is perfect. This is a likely scenario for Germany, where training firms thoroughly screen potential new apprentices by school grades, internships, and entry exams. To capture the effects of efficient initial assignment, equation (1) allows for differences in average ability across training firms, $\bar{a}_{j(i)}$. This could also be interpreted as capturing differences in the quality of training across firms, or differences in firms'

---

[7] That firms differ in key characteristics and that heterogeneous workers may select into firms based on these permanent differences is not a new idea. Groshen (1991) argues that differences in average firm compensation practices can explain a large fraction of cross-sectional wage variation. That part of these differences could be explained by differences in workers' ability has been raised in the literature on inter-industry wage differentials (e.g., Gibbons and Katz 1992). This selection process introduces a bias into simple estimates of inter-firm or inter-industry wage differentials. Gibbons et al. (2002) propose an estimation procedure that takes this bias into account.

[8] Krueger and Summers (1988) suggest that some industries may pay efficiency wages to reduce costly turnover; this is what some versions of the efficiency wage model predicts. Neal (1998) argues that differences in average wages and turnover rates across industries or occupations could both be explained by sorting workers by ability if more able workers accumulate more job specific human capital. In his model, low turnover rates derive from lower quit rates, whereas in the model in the text low turnover rates are driven by lower layoff rates. However, it does not matter for the basic point made here whether it is a firm's policy with regard to turnover that attracts better workers or whether the firm offers more training to such workers; or whether it is another feature of the firm which attracts workers that are both 'low-mobility types' and of higher ability as in Neal's (1998) model.

[9] Davis and Haltiwanger (1990) document how establishment growth rates vary significantly even within industries or regions. However, they are silent about trend differences across establishments or about the extent of differences in turnover rates. Abowd, Corbel, and Kramarz (1999) show that average turnover rates vary significantly between French establishments (Table 4). Anderson and Meyer (1994) find large differences in turnover rates between firm-size classes, industries, and regions. Despite these efforts, still little is known about permanent heterogeneity in turnover rates between establishments.

wages more generally. In terms of this model, initial assignment of less-able workers into high turnover firms implies that $\text{cov}(\bar{a}_{j(i)}, D_{i0}) < 0$; in simple estimates displaced workers may then spuriously appear to obtain lower wages. As in the case of adverse selection, in this model workers' mobility status has no direct impact on wages because workers are always paid their marginal product, and thus $\delta_{It} = \delta_{Vt} = 0$. Note that if one observes a fully informative pre-displacement wage, the bias from perfect initial sorting can be eliminated by analyzing wage changes.[10] While this is often possible for mature workers, it might not be a good assumption for young workers who may receive training wages.[11]

One reason for why wages of young workers may not fully reflect ability is if firms and workers themselves only gradually learn about their abilities and preferences. The process of sorting then becomes sequential as in Gibbons and Katz (1992). Because worse workers get down-ranked over time as employers learn about their true ability, this implies that displacements are associated with wage losses even controlling for initial assignment. In terms of the wage model in equation (1), both $a_i - \bar{a}_{j(i)}$ and $\bar{a}_{j(i)}$ are correlated with the mover dummy D_{i0} since there is both negative selection and initial sorting. However, now more able workers should leave less attractive firms once their ability becomes known implying $\text{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) > 0$. Since negative selection implies $\text{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) < 0$, the bias of the simple OLS estimator is ambiguous in the case of gradual learning. The implications of this more complex model will be further discussed below. Note that since every worker is paid according to their marginal product at all times, as before job changes by themselves have no direct effect on wages, i.e. $\delta_{It} = \delta_{Vt} = 0$.

Both adverse selection and sorting among firms imply that simple estimates *overstate* earnings losses of displaced workers. Yet, in neither model is there scope for voluntary job mobility – the presence of adverse selection in particular should reduce voluntary mobility since movers cannot

---

10 This is done for example by Jacobson et al. (1993), which uses worker fixed effects.
11 Alternatively this would arise in Antel's (1985) model of job mobility, in which employers learn about workers' productivity but wages cannot be renegotiated.

signal their ability to future employers in the presence of asymmetric information.[12] Thus, these models are unable to explain the large amounts of beneficial job change observed for young workers (Topel and Ward 1992).[13] Instead, these patterns are consistent with models of job-search, in which workers are homogeneous, but repeatedly draw job offers from a distribution of wages. Over time, workers searching on the job should obtain more favorable job matches, such that their wages grow with experience even in the absence of general human capital accumulation. In terms of the basic model, job search implies $\delta_{Vt} > 0$; over time, these gains are stable, or increasing if the new job has a steeper career-profile ($\partial\delta_{Vt} / \partial t \geq 0$). However, a displacement destroys this 'search capital' because workers have to start looking for good jobs from scratch (Manning 2001, Chapter 6). Gradually, workers again find better job matches, and the initial wage losses from displacements should be temporary. Within the basic model of wages, such transitory wage losses are captured by $\delta_{It} < 0$ and $\frac{\partial\delta_{It}}{\partial t} > 0$.

The research on young displaced workers aims to focus exclusively on involuntary lay-offs. However, in an environment of high job-to-job fluctuations the distinction between involuntary and voluntary job change may be hard to draw. If measures of job displacement pool voluntary and involuntary movers, simple estimates of the earnings losses from an early job change may *underestimate* the effect of a job loss on wages. The measurement problem is particularly severe in many administrative data sets in which the only information given is whether a worker left a job, not why he did so.[14] A recent study by Neumark (1998) using the NLSY shows that this is the case even

---

[12] For example, Greenwald (1986, p.325) argues, "the marking process which accompanies any job change may impose a substantial turnover cost on workers who seek new jobs."

[13] Topel and Ward (1992) find that for young men, during the first ten years in the labor market job changes lead to 10% wage increases on average, and that 30-40% of wage growth occurs at job changes. More generally, wage changes of voluntary movers have been found to be higher than those of involuntary lay-offs (Mincer 1986, Bartel and Borjas 1981). Basic search models do not distinguish between involuntary and voluntary separations as wages are assumed to be unilaterally set by the employer (i.e., drawn from a pre-specified wage distribution) and renegotiation upon the arrival of new information is assumed to be costless. Antel (1985) and McLaughlin (1991) introduce the distinction by assuming different restrictions on the renegotiation process. Both models predict that wage changes of voluntary quits are higher than at involuntary lay-offs.

[14] A recent paper by Bender, Dustmann, Margolis, and Meghir (1995) estimating wage losses of mature displaced using administrative data from France and Germany tries to circumvent this problem by defining a displacement to have occurred when workers spend at least 30 days out of the labor force after terminating a job. By focusing on displaced

in more conventional data sets. An earlier study (Gardecki and Neumark 1997) finds that high job mobility early in a career has no impact on wages later in life,[15] and Neumark argues that such an estimate must be a lower bound due to the presence of negative selection. However, using local unemployment rates as an instrument for early job mobility, he finds that early job mobility has a significant negative impact, strongly suggesting that mobile workers in his sample are positively selected.

The models discussed so far, and standard models of career development more generally, do not imply permanent effects of job displacements on earnings.[16] Empirical papers concerned with 'scarring' effects of early job loss typically cite two reasons for such long-term effects: there may be losses of labor market experience due to unemployment, or some form of permanent negative signaling to employers.[17] Yet, as unemployment spells of young workers tend to be short, the pure effect from time spent out of work is an unlikely explanation. Similarly, the idea that early displacements or unemployment experiences irrespective of duration 'scars' young workers relies on quite strong implicit assumptions on the learning process.[18] Two alternative models of career development predicting permanent effects of a job loss are a version of 'stepping stone' human capital accumulation and the presence of internal labor markets. In a 'stepping stone' model workers accumulate human capital at the current job level that will be needed at the next higher level. Such

---

workers who became unemployed, this risks imposing part of the final outcome ex ante. To counter the same problem, Jacobson et al. (1993) construct a special 'mass-layoff' sample of workers who leave firms who experience large reduction in workforce.

[15] Mobility is measured either as the number of jobs held or the highest job tenure attained in the first five years on the job.

[16] One of the leading models of career development by Gibbons and Waldman (1999) combines employer learning and heterogeneous human capital growth with job assignment *within* firms. As innate skills and continuous learning by the market determine careers, the model does not imply path dependence.

[17] Most empirical papers discussing 'scarring' effects of early unemployment spells or displacements do not specify the precise economic mechanism behind permanent or highly persistent effects. A notable exception is Machin and Manning (1999), who discuss how true duration dependence could arise among the long-term unemployed. For workers of all ages see, e.g., Heckman and Borjas (1980), Arulampalam, Gregg and Gregory (2001). For young workers see, e.g., Ellwood (1982), Margolis, Simonnet, and Vilhuber (2000).

[18] Even under asymmetric information, continuous learning by employers would predict that eventually workers get paid their true marginal product. The idea of such strong signaling imposes long-term effects almost by assumption. That is not to say that imperfect employer learning (e.g., employers stop learning at some point) is not a realistic possibility. Machin and Manning (1999) suggest that statistical discrimination by employers could lead to such an outcome. A particular example was suggested by Blanchard and Diamond (1994) who construct a model in which employers rank workers by latest arrival into the pool of unemployed.

an approach is akin to Rosen's (1972) model, in which jobs vary in the amount of human capital accumulation they allow. If there are entry restrictions into 'career' jobs leading to different paths of human capital accumulation, early shocks have permanent effects.[19] Okun (1973) speculated that such a 'stepping stone' effect could operate at the aggregate level as well. He suggested cyclical effects could have permanent adverse or beneficial consequences at the cohort level by pushing young workers into better or worse jobs, thereby affecting their human capital accumulation. [20]

Alternatively, labor markets within larger firms are said to provide well-defined career paths. Internal labor markets are said to protect workers from external market conditions and to restrict entry to specific jobs (also referred to as "ports of entry").[21] If such entry-level jobs are scarce, or if entry into internal labor markets is restricted by age, then workers leaving large firms are likely to permanently lose career prospects or rents associated with firm size. In the notation of equation (1), institutional theories predict that $\delta_{It} < 0$ for all periods. Moreover, as firms with internal labor markets should have more stable job attachment and attract better workers, one would expect that $\text{cov}(\bar{a}_{j(i)}, D_{i0}) < 0$. Similarly, since wages in internal labor markets may be rigid, firms could displace their worse workers rather than pay them lower wages, i.e. there may be negative selection and $\text{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) < 0$. The role of large firms will be taken up again in the empirical analysis.

---

19 In this model a promotion in itself would have a true causal effect on wage growth and future promotions. This is also emphasized by Jovanovic and Nyarko (1997) who contrast a 'stepping stone' model with a model of sequential sorting of more able workers into jobs requiring more complex skills. In Gibbons and Waldman's career model on the other hand, promotions and wage growth are based entirely on workers innate ability interacted with a deterministic experience variable; thus, promotions are correlated with future promotions and wage growth but are not causing them in statistical sense.

20 Other effects may arise if at the cohort level as well. For example, prolonged periods of recurrent unemployment of a cohort may lead to a true loss in job-experience as predicted by a general human capital model. Alternatively, unemployment rates at the time of hiring may influence wage contracts workers obtain. Baker, Gibbs and Holmstrom (1994) find such firm-specific cohort effects, and Beaudry and DiNardo (1992) test this hypothesis directly in a model of insurance within the firm.

21 Internal labor markets have been defined by Doeringer and Piore (1971), loosely speaking, as a set of institutions within firms that determine wages (not necessarily by productivity), shelter workers from outside market conditions, and have a clear connection to that labor markets through well-defined jobs ('ports of entry'). Baker, Gibbs and Holmstrom (1994) test several predictions of the internal labor market paradigm.

2.2. Estimates of Wage Losses and Confounding Factors

The empirical application of this paper is concerned with the wage losses of young German apprentices leaving their training firm at graduation. In this particular application, it is unknown whether moves are voluntary or involuntary. In addition, training wages are not informative about workers' ability. Thus, one cannot simply control for initial assignment by taking first differences as one would do for older workers, nor can one concentrate the analysis specifically on displaced workers. Each of the confounding factors is potentially present in the German case, too: there is a high degree of beneficial job mobility; individuals differ in ability; firms differ in their retention rates of young trainees and they actively try to screen among applicants to their apprentice programs; firms acquire disproportionate information about them during training and try to retain their best workers. Thus, the statistical process generating wages can be represented by the model in equation (1).

If one ignores these problems and estimates a simple OLS regression of log real wages on a dummy D_{i0} for moving out of the training firm (leaving out other control variables for simplicity), the estimated effect on wages of a move out of the training firm at the end of training after t years is

$$\hat{\delta}_{It}^{OLS} = \frac{\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})} + \frac{\operatorname{cov}(\bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})} + \delta_{It} + (\delta_{Vt} - \delta_{It}) \frac{\operatorname{cov}(V_{i0}, D_{i0})}{\operatorname{var}(D_{i0})}.$$

Negative selection implies that $\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) < 0$, whereas initial assignment implies that $\operatorname{cov}(\bar{a}_{j(i)}, D_{i0}) < 0$. In both cases OLS tends to be biased toward finding a negative effect even if $\delta_{It} = 0$. On the other hand, search models predict that $\delta_{Vt} - \delta_{It} > 0$. Since $\operatorname{cov}(V_{i0}, D_{i0}) > 0$, this implies that OLS would tend to underestimate the true effect of an involuntary move from the training firm. Together with these confounding elements, the OLS estimate may also pick up true negative effects of a displacement implied by job search, sequential human capital accumulation or institutional models. Clearly, without further information it is hopeless to disentangle the various pieces of information contained in the OLS-estimate and obtain the true effect of mobility.

Two fundamental problems affect the simple OLS estimator. First, there may be both negative and positive selection of workers into moving *within* firms. Second, initial assignment implies that workers differ systematically *between* firms. To solve this problem, the paper uses firm level data on the training firms of young German apprentices to implement the following two-tiered strategy. First, it uses firm fixed effects to control for systematic differences of workers between firms. Thereby, non-displaced workers *at the same training* firm function as comparison group for the wage-outcomes of displaced workers. Second, the paper uses firms' retention rates of other young graduates finishing apprenticeship in the same year as a displaced worker as an instrument for the probability of a displacement. To account for the fact that workers may sort into firms based on average retention rates (i.e., the career prospects firms offer), the preferred instrument will be the deviation of the retention rate from the firm's average. This should isolate as closely as possible the group of workers who would not have moved under normal business conditions and thereby best approximate an exogenous displacement.

The retention rate of a firm is measured by the fraction of workers *other* than the young trainee in question that finished apprenticeship training in the same year who left the training firm. Thereby, I use the mobility behavior of other graduates in the same firm as a proxy for the individual trainee's probability of moving. Let D_{ijc} be a dummy variable denoting the event that worker i in graduating in cohort c leaves firm j. Then for each worker the fraction movers among other trainees graduating from the same firm during the same year is $z_{ijc} = m_{jc(-i)} / (n_{jc} - 1)$, where n_{jc} is the number of graduates at firm j in cohort c and $m_{jc(-i)} = \sum_{l \neq i}^{n_{jc}} D_{ljc}$ is the number of movers among a young graduate's peers. Since a key point of the paper is that z_{ijc} may systematically differ across firms, the final instrument used for the probability of moving will be the deviation of z_{ijc} from its firm specific average $\tilde{z}_{ijc} = z_{ijc} - \bar{z}_j$, where $\bar{z}_j = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{n_{jc}}\sum_{i=1}^{n_{jc}} z_{ijc}$.[22]

The basic idea is to use within-firm changes in labor demand for young apprentices as an instrument for involuntary mobility. That changes in plant-level employment demand are frequent, large, and heterogeneous has been suggested by Davis and Haltiwanger (1990), who show that at each point in time there is a large degree of heterogeneity in employment growth rates between establishments.[23] Some of these employment changes might be expected, but part of it is likely to be due to unexpected changes in labor demand. Other studies have used plant-level changes in employment to identify unexpected shocks to labor demand. Jacobson et al. (1993) use the fraction laid-off at a given firm as determinant of who is counted as displaced worker. Similarly, Gibbons and Katz (1991) use plant closing as an instrument for displacement.[24] However, neither study controls for the 'normal' level of turnover at a firm, and therefore these papers do not directly control for sorting that occurred prior to any 'shocks' to firms' employment. Here I not only introduce firm fixed effects to control for average retention rates, but I am also able to compare instrumental variable (IV) estimates with and without firm fixed effects to learn about the role of initial assignment. To see this, note that implicitly the IV estimator reweights the data such that more weight is put on workers coming from firms with higher fractions 'other' movers (or its deviation from firm means).[25] If there is initial sorting based on firms' retention rates, IV in levels will differ from IV with fixed effects, because the former will put higher weight on less able workers

[22] The average is taken across cohorts and workers. In a full sample, this will be exactly equal to the average retention rate of the firm across cohorts. In the final sample, this won't hold exactly due to sample restrictions.

[23] These findings have been confirmed for other countries, e.g., see Bauer and Bender (2002) for Germany and Abowd, Corbel, and Kramarz (1999) for France.

[24] Since firm-specific demand shocks generate exogenous mobility, models of adverse selection in the labor market more generally can be used to motivate the approach. In Greenwald's words (1986, p. 329), 'random quits' are "workers who leave a firm despite ability levels that, from the point of view of the market at large, should have kept them with their initial employers. This second category includes workers who have special skills, workers whose skills are in excess supply at their current firm and workers whose abilities have been misjudged by their current employers."

[25] The IV estimate in levels can be rewritten as $\hat{\delta}_{It}^{IV} = \sum_{n=1}^{N} z_{ij} w_{it} / \sum_{n=1}^{N} z_{ij} D_{i0} = \sum_{n=1}^{N} \omega_i w_{it}$, where $\omega_i \equiv z_{ij} / \sum_{m=1}^{M} z_{ij}$.

who were assigned to firms with high turnover, while the latter puts more weight on 'random' movers.

If variation in the instrument is driven by firms' demand shocks, the correlation of the instrument with the probability of leaving the training firm at the end of training should be positive. However, if external labor market conditions vary, there may be changes in the fraction of workers leaving voluntarily as well, inducing the opposite correlation. This is particularly relevant for firms with very few apprentices where the mobility of an apprentice might be directly influenced by the mobility decisions of individual colleagues. In part, the question will be resolved by considering the first stage. In addition, I restrict my sample to firms with a minimal number of graduating apprentices. I also consider a second instrument (henceforth IV2), which treats as 'movers' only those workers who have a spell of unemployment of at least 30 days at the end of training. Since I am certain to exclude most voluntary movers, involuntary movers should drive most of the variation in the second instrument. Thereby, it should yield a valid second set of estimates and a useful sensitivity check on the approach.

Thus, using firm level data, I have at least three additional estimators of the wage loss from displacement at my disposition to estimate the true effect of a displacement from the training firm: OLS with firm fixed effects (OLSFE), instrumental variables (IV1), and IV with firm fixed effects (IVFE1). Moreover, I present two additional IV estimates based on my second instrument (IV2 and IVFE2). Each of these estimators helps to address some of the biases affecting OLS. Suppose again that all confounding factors are present in the data, such that the process determining wages can be captured by equation (1). The simplest of the alternative estimates, OLS with firm fixed effects (OLSFE) is identified by deviations from firm averages. The resulting estimate is

$$\hat{\delta}_{It}^{OLSFE} = \delta_{It} + (\delta_{Vt} - \delta_{It}) \frac{\mathrm{cov}(V_{i0}, D_{i0})}{\mathrm{var}(D_{i0})} + \frac{\mathrm{cov}(a_i - \bar{a}_{j(i)}, D_{i0})}{\mathrm{var}(D_{i0})}.$$

By only comparing workers who graduated at the same training firm, OLSFE accounts for the bias from initial assignment. Yet, it is still affected by negative selection and by voluntary mobility.

The next more sophisticated estimator is IV using the fraction of 'other' movers as an instrument (IV1). Implicit in the IV approach is the assumption that $\operatorname{cov}(a_i - \bar{a}_{j(i)}, z_{ij}) = 0$ and $\operatorname{cov}(V_{i0}, z_{ij}) = 0$, i.e., workers induced to move by the instrument are neither negatively selected nor voluntary movers (where I have dropped the cohort sub-scripts from the instrument). If there is no initial sorting, IV in levels identifies the true effect of involuntary mobility δ_{It}. However, if the least able workers are sorted into the firms with the lowest retention rate (the highest fraction 'other' movers), then we have that $\operatorname{cov}(\bar{a}_{j(i)}, z_{ij}) < 0$. The resulting IV estimator is biased, i.e.,

$$\hat{\delta}_{It}^{IV} = \delta_{It} + \frac{\operatorname{cov}(\bar{a}_{j(i)}, z_{ij})}{\operatorname{cov}(D_{i0}, z_{ij})}.$$

Note that since the denominator is now smaller than in the case of OLS ($\operatorname{cov}(D_{i0}, z_{ij}) < \operatorname{var}(D_{i0})$), initial sorting could imply that $\hat{\delta}_{It}^{IV} < \hat{\delta}_{It}^{OLS}$, i.e., the IV estimator can be *more* negative than OLS. Alternatively, this could occur if the effect of negative selection is more than offset by the positive bias from voluntary mobility.

To account for the remaining bias, the last step is to introduce firm fixed effects into the basic IV regression (IVFE). By using firm fixed effects, IVFE is identified by within-firm wage losses of workers moving because the retention rate at their firm was lower than average.[26] In this case, the assumptions necessary for the instrumental variable approach are $\operatorname{cov}(a_i - \bar{a}_{j(i)}, z_{ij} - \bar{z}_{j(i)}) = 0$ and $\operatorname{cov}(V_{i0}, z_{ij} - \bar{z}_{j(i)}) = 0$ (where again I have dropped the cohort sub-script). Since firm fixed effects now control for initial assignment the IV estimate should yield a consistent and unbiased estimate of the true effect of involuntary displacement, i.e.,

$$\hat{\delta}_{It}^{IVFE} = \delta_{It}.$$

[26] The formula for the IV estimator from a within regression is $\hat{\delta}_{It}^{IVFE} = \frac{\sum (w_{it} - \bar{w}_{j(i)})(z_{ij} - \bar{z}_{j(i)})}{\sum (D_{i0} - \bar{D}_{j(i)})(z_{ij} - \bar{z}_{j(i)})}$, where all averages are taken across cohorts and thus do not have a time subscript.

If there are no confounding factors, then OLS, OLSFE, IV, and IVFE should all yield similar estimates of the effect of moving out of the training firm. However, in presence of selection and sorting at the firm level, only the IV estimator with firm fixed effects will yield an unbiased estimate of wage losses from an early displacement.

Each theory has separate implications regarding initial assignment and various forms of selection. Since these predictions translate differently into the various estimators, the stepwise estimation procedure yields more than just an unbiased estimate – it gives a way to assess the relative importance of various mechanisms underlying wage and job changes. The detailed implications of the different theories discussed above for the main estimators are graphically represented in Figure 1 and summarized in Table 12.[27] Since most predictions follow readily from the discussion of theories and estimators above, below I briefly summarize the main points.

First, if adverse selection is an important determinant of wage losses in this sample, one expects that estimates of wage differences among movers and stayers by either IV or IV with firm fixed effects (IVFE) should be considerably smaller than the OLS estimate in all periods after exit from the training firm. This is shown in Panel A of Figure 1 (see also Table 12). Second, initial assignment is the only model predicting a strong role for firm fixed effects. If initial assignment is perfect, OLS with firm fixed effects (OLSFE) should remove all differences between movers and stayers. If sorting occurs along the lines of retention rates, then IV is likely to be more negative than OLS, and IV with firm fixed effects should be zero at all time periods (Panel B, Figure 1). Third, job search is the only model predicting a true but *temporary* decline in wages following an involuntary job change. Moreover, job search models predict an important role for voluntary mobility and thereby for positive selection among movers. Thus, IV estimates should be more negative than OLS (Panel C, Figure 1). Fourth, if gradual learning is important in addition to initial assignment, then better workers move out of high turnover firms; if initial assignment is not too strong, then sequential sorting predicts that IV is less negative than OLS. OLSFE should be less negative than OLS and the

[27] Mathematical derivations are presented in Appendix A.

extent of the difference will again depend on the degree of initial sorting.[28] IVFE should yield a zero estimate, since a random displacement has no causal impact in this model (Panel D, Figure 1). Last, the key insight of the internal labor market paradigm or models of sequential human capital accumulation is that even under IVFE there should be a permanent negative effect of involuntarily moving out of the training firm. These predictions will be used at the end of the empirical section to interpret the main results.

3. Data and Institutional Background

3.1. The German Apprenticeship system

Two-thirds of young Germans follow an apprenticeship in the German "Dual System", during which they receive both formal state-sponsored schooling as well as training on the job. Most apprentices start training right after junior-high school, and the majority fully participates in the labor force at the end of the apprenticeship.[29] Apprenticeships last on average two and a half years, after which about 40% of workers leave the training firm immediately. Training is mainly general and employment rates of graduating apprentices are very high, and these are the two features most often cited by proponents of large-scale apprentice systems in other countries. The institutional structure of the German apprenticeship system is ideal to analyze the persistence of early labor market shocks, since we can study the effects of a well-defined event (transition from training into the labor market at the same or at a different firm) for a group of workers with relatively homogeneous labor market experience and background. However, as mentioned above, all three of potential confounding factors are potentially present. Firms are likely to learn about workers and try to retain only the best of them (Acemoglu and Pischke 1998). Mobility is high even for workers who

---

[28] If OLSFE remains negative, sequential sorting suggests that negative selection is stronger than positive selection. In terms of the discussion in the theory section, it implies that $\mathrm{cov}(a_i - \bar{a}_{j(i)}, D_{i0})$ is negative rather than positive. This is further discussed in the Appendix A.

[29] A detailed account of the German education and apprenticeship system can be found in Franz et al. (2000). Using the German national account system for education (Bildungsgesamtrechnung) they describe the flows of workers between various forms of training, schooling, employment, unemployment and labor force participation. Winkelmann (1996) gives a concise overview of the German education and training system. More information on the German apprenticeship system and apprentices is also available from the Employment-Qualification Report (Berufsbildungsbericht 2001).

stay at the training firm, suggesting that young graduates from the Dual System have other options (Euwals and Winkelmann 1996). Moreover, firms differ in their turnover rates and possibly in the quality of training they provide (Buechtermann 1989).[30] Thus, the application of the theoretical and conceptual framework outlined above is appropriate for the German case. However, it is the availability of the relevant establishment-level data that makes this exercise most exciting

Before describing the data, the paper will benefit from briefly establishing that, while Germany does clearly differ in its institutions, enough basic similarities in the labor market for young workers exist to make the results relevant for the understanding of American labor markets.[31] To do so, I limit myself to noting a few results from a comparison of the German labor market with a data set exactly replicating Topel and Ward's (1992) seminal study on career patterns of young American workers. As in Topel and Ward's study, the sample consists of all men between 18 and 34 years who are in stable employment.[32] Table 2 summarizes the basic findings. First, labor force attachment is slightly higher among Germans over the first two years of potential experience (the period during which a large fraction of German workers participates in apprenticeship training), but evolves similarly in Germany and the U.S. afterwards (Panel A of Table 2 and Appendix Figure C1). Second, while Panel A of Table 2 shows Americans do transit through more jobs in the first years of the labor market, Panel B shows how job attachment is similar for jobs lasting at least six quarters.[33] Third, Panels B and C of Table 2 show that both wage growth within and between employers is important and of similar magnitude in Germany and the US.[34] Fourth, it is the first years in the labor

[30] The variation in training is said to occur by size (Winkelmann 1996) and across sectors (Harhoff and Kane 1997), but little direct evidence exists.
[31] Critics of policy proposals based on the German apprenticeship scheme argue that the system is fundamentally linked to German labor market institutions absent in the US. For an overview of the arguments for and against establishing large-scale apprentice programs in the US, see Stern et al. (1994) or Heckman et al. (1996). Blau and Kahn (1997) compare the labor markets for disadvantaged young workers and women in the United States and Germany.
[32] The study is based on a 1% sample of the data set used in the main analysis provided by the Institut fuer Arbeitmarkt- und Berufsforschung (IAB). The data, sample used, and approach are described in the notes to the tables and figures in Appendix C.
[33] As the first graph in Figure C2 shows, the difference in the overall hazard of job leaving is concentrated in the first six quarters of job attachment. The lower graph demonstrates how this patterns are driven from differences in job-to-job transition rates; job-to-non-employment transition rates are similar. Similar differences in mobility rates among young workers are reported by Ryan (2001).
[34]Average within job wage growth for a restricted sample with at least six quarters of tenure in the US is 7.1%, average completed job duration is 7 years, and average rate of job change is 28% (Topel and Ward 1992, Table VI). The

market that matters most in both countries; in the US, two thirds of earnings growth occurs in the first ten years of the career (Murphy and Welch 1992). In Germany, counting the apprentice period, 41% of earnings growth occurs in the first 5 years, and 80% of earnings growth occurs in the first ten years. Last, as will be seen in the next section, raw wage losses from leaving the training firm are a similar order of magnitude of wage losses of young displaced workers in the US.[35] In addition, Harhoff and Kane (1997) and Blau and Kahn (1997) find that apprentices occupy a similar position in the wage distribution, have similar wages, and similar wage-experience profiles as high-school graduates in the US.[36] Overall, in the sense of Ryan (2001), it is reasonable to suppose that the 'fundamental economic mechanisms' operating in the labor market for young workers in the two countries bear some important similarities.

3.2. German Social Security Data

The data used in this paper is drawn from the German employment register containing information on all employees covered by social security, representing around 80% of the German workforce.[37] The notification procedure for social security requires employers to record any permanent or temporary change of employment relationships, and in addition takes stock of existing employees at each establishment twice a year. Therefore, the employment register contains detailed

corresponding values in Germany are 9%, 7.3 years, and 25% (for more detail see Table C1, Panel A, which exactly reproduces Table VI). Average wage growth occurring between jobs (controlling for experience) is 19.9% in the US (TW, Table VII) and 20.6% in Germany (Table C1, Panel B, which exactly reproduces Table VII).

[35] Note that if we exclude apprentices from the sample, wage growth after 10 years is 70% of overall wage growth – apprentices tend to have higher initial wages but lower wage growth with experience. No estimates for *young* displaced workers exist for Germany. As shown by Couch (1996), wage losses for mature displaced workers are similar to wage losses of workers in the US.

[36] See Harhoff and Kane (1997) Table 5 and Figures 3 and 4, or Blau and Kahn (1997) Figures 1b and 2b. The main difference between the US and Germany is the labor force participation rate and unemployment rate of very young workers (Ryan 2001, Blau and Kahn 1997). These results are consistent with the idea that the main impact of the German apprenticeship system is to reduce unemployment rate among young workers.

[37] The employment register was established in 1973 to integrate the notification procedures for health insurance, pensions and employment insurance. Currently, only the data from 1990 to 1998 is accessible. The data is described briefly in Bender, Haas, and Klose (2000) and in more detail in Bender et al. (1997). Coverage includes full- and part-time employees of private enterprises, apprentices, and other trainees, as well as temporarily suspended employment relationships. The self-employed, civil servants, and students are excluded.

histories for each worker's time in covered employment.[38] Besides period of coverage, the key information contained in the register for administrative purposes (and therefore the most reliable) are gross daily wages subject to social security contributions. Contributions have to be paid only up to a limit, but top coding is very rare for younger workers. In addition, the data contain basic demographic information (gender, age, nationality) as well as information on occupation, industry, job-status, and education.[39] Most important for the present purpose, the data also contain unique establishment identifiers. These were used to create a separate data set of establishment characteristics that were aggregated up from the employment register and merged back onto the individual level data. Characteristics include among others establishment size, employment growth, number of graduating apprentices, and average wages.[40]

The working sample consists of information on the universe of trainees graduating from an apprenticeship in 1992 to 1994 in West Germany drawn from the employment register. The timeframe is chosen such that several cohorts can be observed for at least 5 years after entering the labor market. (Before 1992 an apprentice is not clearly identifiable in the data.) An important caveat is that for the 1994 cohort the 5th year is not available since 1999 is currently not in the sample. As the cohort dimension is important to calculate average retention rates of employees by firms, workers graduating in 1994 are nevertheless kept in the sample. Moreover, for purely idiosyncratic administrative reasons only 80% of notifications are available for 1998. For each graduating apprentice the sample contains information on the establishment where training takes place (size, employment growth rate, number of apprentices graduating, number of graduating apprentices staying at the firm, total number of apprentices, average training wage, average overall wage), on training itself (duration in days, training wage, occupation, industry), and basic demographic characteristics (gender, age, education prior to training). Moreover, for each apprentice the sample

---

[38] As it stands, information on receipt of unemployment benefits, pensions, or health insurance benefits is not linked to the core data. This information has been linked to the IAB employment sub-sample (a 1% draw from the main register), a scientific use data set.

[39] The entity reporting is the establishment for which an employee works and can thus change over time. This can lead to mistakes in the coding of some demographic variables (e.g., nationality or marital status) and in particular education (which tends to reflect required rather than actual qualification).

[40] Unfortunately, it is currently not possible to link establishments that belong to a common parent firm.

contains daily gross wages at exactly the first, third, and fifth year of potential labor market experience after the end of training. We also know whether apprentices have spells of non-employment and will use it to make further sample restrictions in the sensitivity analysis.

To ensure the sample consists of 'core' apprentices, it is restricted to occupations participating in the "Dual System" (i.e., other vocational training is not included), it requires a minimum length of continuous training of 450 days, and it excludes workers who have prior labor market experience, who have more than one apprenticeship spell, and who are older than 30 at the end of training.[41] To make the study of firm's retention rates useful, an additional crucial restriction is that firms with less than 50 covered employees and less than 5 graduating apprentices in a given year are excluded from the sample. A large fraction of apprenticeships occur at very small establishments, such that this restriction reduces our sample by roughly 50%. This limits the representativeness of the sample with respect to the German apprenticeship system as a whole. On the other hand, very small training firms seem to follow different incentives than larger firms (Winkelmann 1996), and the economic mechanisms studied in the paper are likely to apply more to larger firms.[42] Finally, workers are required to have a minimal amount of attachment to covered employment (i.e., they must have at least one appearance in covered employment after their third year of potential experience) and daily real wages are required to be above 30DM in 1996 prices (about $15).

Table 3 shows the basic characteristics of the sample for all graduating apprentices in the final sample (Column 1) as well as for those with valid wage observations in the first and fifth year of potential experience. The main sample consists of 218880 observations on graduating apprentices. Since it is restricted to larger training firms, the sample is slightly older, slightly better educated and has a higher fraction men than the full sample of apprentices.[43] A high fraction of the sample is concentrated among very large firms (with more than 500 employees) as these have larger training

[41] These sample restrictions are also imposed by Dustmann and Schoenberg (2002) for the same reasons.
[42] For the same reason, Acemoglu and Pischke (1998) also focus on firms with at least 50 employees. The comparison of training programs and the fate of apprentices graduating from larger vs. very small training firms is interesting in its own right and is a question for future research. Very large training firms will be discussed separately below.
[43] See Appendix Table D1 for tabulations from the raw sample. The raw tabulations are similar to samples from other data sets (e.g., see Winkelmann (1996) for the GSOEP, Acemoglu and Pischke (1998) for the Qualification- and Career Survey of the Bundesinstitut fuer Berufsbildung (BIBB), and Euwals and Winkelmann (1996) for the IAB Employment Subsample), with the main exceptions noted in the text.

programs. Most training lasts between two to three years, but longer training spells are not uncommon. The fraction moving from the training firm is 35%, which is slightly lower than in the raw sample and slightly higher than some tabulations from the GSOEP or the IAB employment sub-sample suggest. This is could be due to the fact that the current sample counts even brief separations from the training firm as moving and that the sample is more recent.[44] The mean of the fraction of 'other' movers is slightly higher than the full mean since it is calculated from the basic sample of graduating apprentices without restrictions on labor force attachment.[45] The standard deviations for fraction 'other' movers and for training firm's annual growth in overall employment show that there is indeed a great deal of variation within the sample. Due to the focus on large firms and apprenticeship programs, the training wage is only 90% lower than the first wage, which is higher than apprenticeship standards. Out of the main sample, 89834 apprentices graduated in 1992, 68283 graduated in 1993 and 60763 graduated in 1994.[46] Most apprentices are trained in the manufacturing sector, followed by banking and trade.

The last two columns of Table 3 shows the characteristics of the initial sample by years of potential labor market experience for those with valid wage observations in those years. As mentioned, there are some variations in observable characteristics over time, and the sensitivity analysis will control for these changes directly by restricting the data to a balanced panel. Part of

---

[44] The fraction of apprentices who at the end of training had a job offer from their training firm in 1999 was 65%, in between the estimates of the German Socio-Economic Panel (GSOEP), the IAB Sub-Sample, and the fraction in the current sample (Berufsbildungsbericht 2001, Chapter 4.5, Table 81). More puzzling is the difference with respect to the BIBB's Qualification and Career Survey, which suggests that 84% of workers stayed at their training firm. In light of the other estimates this seems very high, and may be due to the fact that the BIBB is a retrospective. A thorough comparison of various data-sources on the German apprenticeship system stands out.

[45] The fraction movers is calculated on the full basic sample of apprentices in the "Dual System" since it supposed to measure a characteristic of the training firms. Moreover, this minimizes measurement error by keeping the sample as large as possible. The fact that the average fraction 'other' movers is higher than the full average implies that taking firm fixed-effects may not completely take into account permanent differences in firm turnover rates. This is further discussed below.

[46] The number of apprenticeships has been declining in large-scale manufacturing (which traditionally trained half of all apprentices) since 1989 (Grund- und Strukturdaten 2000/2001, Chapter 3, First Table). However, the decline cannot account for the differences in the sample sizes across cohorts. Instead, it appears that the 1992 cohort includes some workers trained in East Germany. Apprentices graduating in 1992 have on average lower training duration. Along most other characteristics (including the training wage, training firm size, fraction mover) the samples do not vary systematically across cohorts in a way to suppose that these differences could influence the analysis. This is shown in Table D2 in Appendix D. The cohorts differ reflecting changing economic conditions in Germany during the sample period (1992 was still influenced by the economic boom brought by reunification, while 1993 and 1994 were recession years).

these changes are likely to be due to compulsory military service, as the fraction of men rises from year 1 to year 3, while the fraction of foreigners and the fraction of those with a high-school degree declines.[47] Other characteristics, such as the firm-size distribution, average training wages, training firms' employment growth or the fraction movers are stable over time. This is reassuring, since it means that the change in sample decomposition among cohorts shown in the last rows has little effect on the properties on the sample (the decomposition across cohorts is the same in year 1 and 3, but the fifth experience year of cohort 1994 is missing). Real daily wages grow at about 3% per year, a lower rate than for American high-school graduates (for which wages grow at about 5% per year during the first ten years of the labor market).

3.3. Empirical Implementation

The data set consists of an unbalanced panel of apprentices each observed at their first, third, and fifth year of potential labor market experience after the end of their apprenticeship. For the purpose of estimation, the observations are stacked into a panel, and all estimates are obtained from the stacked model. Since error terms will be correlated across individuals and potentially also within training firms, this system is akin to a system of seemingly unrelated regressions (SUR). The system of equations will be estimated with OLS, yielding consistent estimates. To take into account correlations between time periods and within training firms, all standard errors are clustered at the level of the training firm.[48] In case of OLSFE, the training firm fixed effects are restricted to be the

---

[47] Drawing from information off the web from the German defense ministry and the German institute for social service, in 2002 80000 men joined military service and in 2000 125000 joined social service. In 1998, there were 350000 men aged 18-22. A back of the envelope calculation assuming that all men are drafted in that age range suggests that the risk of exiting the labor force at these ages is not small: 205000/350000=58.6%. If the chance is equally spaced across the five ages, then the risk at each age is 12%. Unfortunately, the sample does not allow for an unambiguous identification of those leaving their training firm to join the military. In some cases if the worker plans to return to the training firm, firms can notify the social security administration of an 'interrupted' employment relationship. Dustmann and Schoenberg (2002) have tried to identify 'military-movers' by looking for young men who spent a continuous stretch of 12 to 18 months out of the labor force (in the mid-90s, military service was 10 months and social service was 13 months, but some may commit to serve in the military longer). In the current data set the fraction of movers who are out of the labor force continuously for 12 to 18 months is 11.8% for men and 4.9% for women.

[48] The cluster is *not* interacted with period. It therefore includes all observations on an individual and takes care of cross-individual correlation as well. Given that the regressors are the same across the three periods, in case of OLS it would be equivalent to estimate the model separately for each period. However, if there are cross-equation restrictions as in the case of OLSFE (or IVFE), estimating the stacked model is necessary to obtain correct estimates of the standard errors (Ruud 2000, p.703).

same across periods, and thus the model amounts to a restricted SUR. Again, the standard errors of the estimates are clustered at the training firm level. As further discussed below, changes in the sample composition occur due to military service or exit into other forms of employment. To ascertain that these changes over time do not affect my results, I run all my specification on both balanced and unbalanced panels.

If the panel were balanced, to obtain IV estimates I would estimate a single first stage for the endogenous variable (mover status at the end of training) and three reduced form equations. Since the same individuals are observed at different points in time, the latter again form a SUR (or restricted SUR in the case of IVFE). The IV estimate would then be obtained by dividing the reduced form by the first stage coefficients. Since the sample changes between periods, this approach would be akin to a two-sample IV estimator in which the samples used to obtain the first stage and reduced form coefficients are not independent (Angrist and Krueger 1992). Therefore, particular care had to be taken in estimating the standard errors (Murphy and Topel 1985).[49] Instead, here I estimate three separate first stages for the three periods in a SUR model, and use the resulting coefficients on the instrument to obtain my IV estimator. This ensures that the samples from which I estimate the first and second stage coefficients are exactly equal and avoids the problems of two-sample instrumental variables. All standard errors will again be clustered at the training firm level, taking into account correlations of individuals over time and within training firms.

The empirical analysis then proceeds in four basic steps. First, it compares those moving and staying at the training firm at the end of the apprenticeship using OLS and controlling for observable characteristics such as training firm size, training wage, or workers' education. Second, it considers stepwise more complicated IV and fixed effect estimates to obtain the true effect of moving out of the training firm. Third, it will check the basic results in a sensitivity analysis; among others, it considers the alternative instrument, a balanced panel, alternative restrictions on labor

[49] The assumption of independence is not needed to obtain consistent estimates. However, the covariance matrix of the estimator has to be adjusted accordingly. Murphy and Topel (1985) discuss this in the framework of two-step estimation, derive estimators for the covariance (Theorem 2), and draw analogies to 2SLS (p.376).

force participation, and very large training firms. Last, the analysis compares the results from the various estimates to the theory to learn about potential mechanism of job mobility.

4. The Long-Run Effects of Leaving the Training Firm

A high fraction of workers leaves the training firm right after training. Panel A of T able 4 shows that stayers have about 9-10% higher wages than movers, and that this difference is basically unchanged after 5 years in the labor market (the differences in the first, third, and fifth year are 9.3%, 9.7%, and 9.7%, respectively). Panel B shows that movers are different than stayers. Movers tend to be less educated, more likely to be male, and more likely to be trained at smaller firms. They are concentrated in the service sector and in transport and communications, and are more likely to be blue-collar workers. Movers also receive lower training wages. Moreover, they work at firms that pay slightly lower training wages, that have lower employment growth, and that have a much higher average fraction of their apprentices moving.[50] This is a first indication of the strong correlation between the probability of moving and the firm's retention rate of apprentices. Note also that movers are more likely to graduate in 1994 – a recession year.

Table 5 shows the differences in real wages between movers and stayers after controlling for the characteristics of the worker, the firm, and of the apprenticeship. All regressions also include interactions between cohort and experience dummies. Column 1 shows raw estimates only including experience-cohort effects, which are essentially the same as the raw average differences shown in Table 4. Including individual characteristics (Column 2) does little to change the effects (results from the full regression are in Appendix Table E1). Adding firm size and firm growth rates reduces the differences only by about 1% (Column 3). The gap is significantly reduced to slightly more than 5% when the log real training wage and dummies for training duration are included (Column 4).[51]

[50] Two thirds of the difference in training wage among stayers and movers is explained by differences among their firms. Note that on average workers with lower training wages have higher wages in the labor market, i.e., there appears to be mean reversion. Thus, without controlling for mean reversion movers will appear to have higher wage growth than stayers relative to training wages.

[51] Since I cannot include individual worker's fixed effects, including training wages partially controls for productive ability (in a limited fashion as training wages are not necessarily reflective of workers' skills). Individual fixed effects cannot be included since the outcome of interest is the difference in wage levels, not wage changes.

Conditional on training and firm variables occupation and industry controls result in only small decline in the differences (Columns 5 and 6). The most important conclusion of Table 5 is that while the overall difference in wages among movers and stayers declines significantly when additional characteristics are controlled for, it remains stable over time at all specifications (if at all, differences seem to increase slightly with experience).

However, these estimates may not reflect true 'causal' effects. If they are due to initial sorting of less able workers into establishments with lower retention rates (and lower wages), then controlling for firm fixed effects should solve the problem, as they force the comparison to be done relative to workers trained at the same firm. The wage differences among movers and stayers after controlling for firm fixed effects alone are shown in the third column of Table 6 (the first two columns report the estimates of the first and fifth specifications in Table 5). The results suggest that firm fixed effects alone can go a considerable way in explaining the raw difference in wages. Thus, part of the individual and training characteristics in the previous regressions may simply pick up sorting among firms. However, adding firm fixed effects to the full OLS specification reduces the differences shown in Column 2 by only half a percentage point (regression not shown). Thus, sorting alone does not appear to be able to explain the remaining difference between movers and stayers.

Another suggested explanation for the remaining wage differences is that movers are negatively selected with respect to stayers. To control for the possibility of negative selection, I use the fraction of movers among other apprentices graduating at the same firm in the same year as an instrument for the probability of moving. As discussed above, if adverse selection is the main source of wage losses, then workers leaving firms with higher turnover rates should be less negatively selected. Table 7 shows the first stage regressions of a dummy for leaving the training firm at the end of apprenticeship on the instrument and observable characteristics. The specification adopted throughout the main analysis includes a rich set of covariates on individual, firm, and training characteristics – the exact specification is that of Column 5 in Table 5 (and reproduced in Column 1 of Table 6).[52] As described above, to ensure that the distribution of characteristics in the first stage and the reduced form is the same, all variables in the first stage are fully interacted with experience.[53] Column 1 of Table 7 shows that the fraction 'other' movers is a strong determinant of the probability of moving. The coefficient on the instrument changes somewhat across experience years in response to changes in the sample composition, but overall is quite similar (however, given the sample sizes, most of the differences are statistically significant). The IV estimates, obtained by dividing the coefficients from the reduced form with those of the first stage, and the correct IV standard errors, clustered at the establishment level, are shown in the fourth column of Table 6. Compared to the basic OLS estimate, the estimated wage differences among stayers and movers become *more* negative than OLS and are on the order of magnitude of the 'raw' wage differences. In addition, using the level of firms' fraction 'other' movers as instrument one still obtains persistent differences over time – from the first to the fifth year of potential labor market experience, the difference declines only by 1%. Thus, adverse selection alone cannot seem to explain the observed wage differences. On the other hand, the increase in the gap among movers and stayers is consistent with the hypothesis that less able workers sort themselves into firms with lower retention rates, or that these firms provide training of lower quality.

Both the results on OLSFE and IV suggest that workers sort themselves into firms according to turnover rates. However, the fact that OLSFE is still negative suggests that firms selectively displace their worse workers.[54] In this case, we have to control for permanent differences *between* firms as well for differences among movers and stayers *within* firms. To do so, I add firm fixed effects to the basic IV model of the previous paragraph. This means the instrument for the

---

[52] Note that industry effects are not included in the main specification. Industry effects represent an alternative specification to control for permanent differences in turnover rates among groups of firms. Unfortunately, the distinction between shocks relative to the firm average or the industry average is not so well defined. This is discussed separately in the sensitivity analysis.

[53] The coefficients on the instrument in the reduced form are shown in Table E2 in the Appendix. Other coefficients in the first stage and the reduced form are shown in Table E3.

[54] An alternative explanation of such a pattern would be if the sorting process involves symmetric learning. Then it might be reasonable to suppose that workers of high ability that initially were at a high turnover firm will move voluntarily. Less able workers who were at a low turnover firm would then likely to be displaced. If this is the case, then again there are both differences in the average type of movers across firms *and* movers might be of lower ability than stayers at the same firm. This is taken up again in Section 7.

probability of moving is now the deviation of the fraction of movers in a cohort from the firm specific average. As discussed above, by focusing on movers who under normal business conditions would have stayed at the firm, this should mimic the event of a random displacement and should be free of a bias due to adverse selection. Moreover, controlling for firm fixed effects the comparison is made with workers graduating from the same firm and this addresses differences caused by sorting. The first stage is again shown in Table 7 (Column 2). The coefficients on fraction 'other' movers is now smaller than in Column 1, but of more reasonable size and still highly significant.[55] The IV estimates are shown in the last column Table 6.

Including firm-fixed effects, the estimated wage difference among stayers and movers in the first year after entry into the labor market is –11.3%. This is the largest estimate among all specifications shown so far. However, using firm fixed effects the difference is not persistent and decreases to –6.8% and –2.9% in the third and fifth year of the labor market, respectively. These are clear signs of a 'catch-up' of movers towards wage levels of stayers. While the estimate after three years is still significantly different from zero, the estimate after five is not. Unfortunately, the standard error on that estimate is large (partly because one cohort is missing for that year) so we cannot exclude bigger effects. However, the estimates after three and five years are significantly different from the initial difference at a 1% level (see the lower panel of Table 6). Even the decline in the last two years is significant at the 10% level.

Figure 2 plots the basic estimates of the effect of moving in a single graph. The main patterns shown in the figure, which will repeat themselves throughout the empirical analysis, can be summarized as follows:

- OLS differences are smaller than raw differences, but persistent and significant
- OLS with only firm fixed effects reduces difference as much as other covariates, but not more
- IV without fixed effects tends to imply larger differences than OLS

[55] Consider a firm with 20 apprentices. Suppose the average fraction of workers who move after apprenticeship training from the firm is 40%. If instead of 12 workers the firm only retains 4, that fraction rises to 80% (=16/20). Since the coefficient in the first stage is roughly 0.15, a temporary increase in the fraction 'other' movers of 40% implies an increase in the probability of moving of 6% (an increase of 15% relative to the baseline). If the coefficient is 0.75 as in the case of IV without fixed effects, the implied increase is 30% (an increase of 75% relative to the baseline).

- IV with firm fixed effects leads to larger initial differences and a significant catch-up pattern.

These results suggest that controlling for both permanent differences across firms and firm-specific shocks is crucial in the analysis of the effects of worker mobility. There seems to be both sorting of workers between firms and sorting of workers within firms, such that neither simple IV estimates nor firm-fixed effects alone can sufficiently account for the underlying selection mechanism. Once this is done, the wage differences between movers and stayers are no longer permanent but show strong signs of mean reversion. Moreover, consistent with separate results on the role of mobility in the German labor market, the fact that the difference estimated by IVFE is initially bigger than the OLS estimate suggests that a large portion of mobility is voluntary. Simple estimates miss this distinction as they group several types of workers together. By averaging over positively and negatively selected workers they tend to underestimate the initial effect of moving.

To see that these results are robust and not driven by some peculiarities of the data or by choice of particular specifications, it is useful to consider some simple graphs of the reduced form. Controlling only for experience-cohort-effects, Panel A of Figure 3 shows the simple average of real wages by intervals of the fraction 'other' movers (the instrument in levels) for the three experience years. The three panels show that the linearity assumption underlying the results in Table 6 is justified.[56] Moreover, the relationship is negative and does not change over time. Panel B of Figure 3 shows the same graph but now controlling for firm fixed effects. This is the average *deviation* of wages from firm means by twelve brackets of the demeaned instrument. Again, linearity cannot be rejected, and for the first experience year there is a clear significantly negative relationship – larger shocks to the firms' average retention rate induce higher wage gains or losses. However, now the overall pattern between years is one of a successive weakening of the relationship: the estimated

[56] The brackets are defined as {[0-.1),[.1-.2),[.2-.3), …, [.9,1), '=1'}, i.e., the 11th 'bracket' is for fraction 'other' mover equal to one. The graphs also show regression lines of regressions of the averages weighted by their standard errors on a constant and a linear trend. The slopes are all significantly different from zero but not significantly different from each other. Moreover, the linearity assumption cannot be rejected by a simple Chi-Squared test. The test-statistic is the sum of squared residuals of the weighted trend-regressions with the number of brackets as degrees of freedom. The test-statistics are .0034, .0017, and .0026 for year 1, 3, and 5, respectively. The 95th percentile of the Chi-squared with 11 degrees of freedom is 23.03.

slopes rotate around zero from being significantly negative to almost a flat line.[57] These changes are significant and indicate that bad or good luck is only transitory. Those apprentices whose firms had a lower than average retention rate as they graduated from training have initially higher wages than usually paid by the firm, but these differences fade over time. On the other hand, those who graduated from firms shedding more apprentices than usual have only temporarily lower wages then they could have expected based on average firm outcomes. These patterns reinforce the finding that for graduating German apprentices positive and negative firm specific shocks fade over time.

4.1. Sensitivity – Measurement and Sample Decomposition

The data set does not distinguish between voluntary and involuntary movers, and so I treat both types the same when calculating the fraction 'other' movers. Given that the results in the previous section suggest a high degree of voluntary mobility, it might be suspected this procedure introduces some measurement error in the instrument for smaller firms, where job offers and mobility of single workers may influence each other. This does not affect the IV estimates per se since predictive power of the first stage is good, but a closer look will enhance the credibility of the results. To do so, the first panel of Figure 4 shows the fraction of movers by the same intervals of the fraction 'other' movers as before (this is the analogue to Panel A of Figure 3). As the results from the first stage regression in Table 7 (Column1) suggest, there is a strong positive relationship. The second panel of Figure 4 shows the same figure for deviations from firm means (the analogue to Panel B of Figure 3). On average, a higher 'shock' to the fraction 'other' movers leads to higher than average probability of moving, and this is what the first stage in Table 7, Column 2 picks up. However, the assumption of linearity works less well in this setting – for small deviations of the retention rate around the mean (between –10% and 10%) the relationship seems to be negative.

---

[57] The brackets are defined as {[-1,-.5),[-.5,-.4),[-.4,-.3), [-.3,-.2), [-.2,-.1), [-.1,.0), [0,.1), [.1,.2), [.2,.3), [.3,.4), [.4,.5), [.5,1)}. The distribution is concentrated among small deviations around the mean, but there are still a sizable number of large positive and negative deviations. The estimated slope coefficients from a regression of the cell-averages on a trend and a constant weighted by the inverse of their standard errors are -.0043 (.0008), -.0019 (.0007), and -.0005 (.0009) for experience years 1, 3, and 5, respectively. Intercepts are .026 (.0008), .01 (.005), and .0015 (.006). The changes from year 1 to year 3 and 5 in both slopes and levels are statistically significant at least at the 5% level. The test-statistics for the chi-squared of linearity in this case are for each year .0004, .0009, and .0006; the 95th percentile of the Chi-squared with 12 degrees of freedom is 21.03.

While this might be intuitive for smaller firms, one would like to exclude this variation from our analysis.

To go to the core of the problem, I use the second version of the instrument based only on those movers who spent at least 30 days out of covered employment (but not less than 10 months to exclude military leavers). Albeit this is too crude an approximation to capture all involuntary movers, it is likely to exclude most voluntary movers from the sample. Using this definition, the relationship between the fraction 'other' movers and the likelihood of moving is again close to linear even controlling for firm fixed effects (Figure 4, Panel C). This suggests that the non-linearities seen in Panel B of the same Figure are indeed due to voluntary mobility. Using the more narrowly defined instrument, one obtains very similar results as with the main instrument. Columns 3 and 4 of Table 7 displays the first stage coefficients and Figure 5 (Panel A) and Table 8 show the main estimates. The basic patterns shown in Table 6 and seen in Figure 2 are clearly confirmed. As might be expected, those movers coming from firms from which a high fraction of workers move and spend some time in non-employment now have much more negative wages then stayers (i.e., IV is more negative). However, once we control for permanent differences in wages and fraction movers across firms these estimates are reduced and the wage losses of movers decline even more strongly than before. Column 5 of Table 8 shows that the decline in the effect over time is again highly significant.

Another concern expressed above is that changes in the sample decomposition might affect the results – e.g., if among movers the best workers leave for the military, or if the worst workers sequentially drop out over time. Note that since the OLS estimates are stable, this had to occur only for those induced to move by the deviation of firms' retention rates of trainees from average. To address this problem, I first restrict the sample to include only workers who have valid wage observations for each period in the sample.[58] Panel B of Figure 5 shows the main results for this sample, while Tables 9 contains the relevant estimates. Note that for this sample of workers with high labor force attachment the OLS estimates of the wage difference among movers and stayers is

[58] This means we will have three observations for all workers graduating in 1992 and 1993, and two observations for those graduating in 1994.

smaller than for the full sample. Moreover, the initial decline estimated by IV with fixed effects is larger, suggesting that among these workers there is a bigger fraction of voluntary movers. Again, while all other estimates suggest stable or slightly declining differences, instrumenting mobility with firm-shocks to the fraction 'other' movers leads to clear decline in the initial disadvantage from moving out of the training firm. Catch-up is strong, but appears not as complete as before, largely due to the fact that OLS estimates are now smaller. If we repeat this analysis with our narrowly defined instrument, the results instead point to stronger convergence with slight overshooting in the fifth period. This is shown in Panel C of Figure 5 (the estimates are in Appendix Table E4). Given the high standard errors on the estimate in the last experience year, we would expect to observe some positive estimates if the true parameter is zero.

One might argue that the panel-sample is overly restrictive in that it is likely to exclude many men leaving for military service. As an alternative restriction on labor force attachment that does not automatically exclude those in military or social service we therefore restricted total time spent out of covered labor force to be at most 6 months per year (thus no more than 560 total days in experience year 3 and 900 total days in experience year 5). The corresponding estimates are shown in Panel D of Figure 5 and Appendix Table E5. The results are again striking – while all other coefficients predict only slightly changing wage differences among movers and stayers, the IV estimate shows a larger initial difference with a strong following decline towards zero. In this case, the estimate is almost exactly zero, although again the standard errors do not exclude some positive or negative effects. Overall, the sensitivity analysis suggests that the key results of the main analysis summarized in Figure 2 are a robust feature of the data.

4.2. Sensitivity – Large Firms

That large firms provide a different environment for the career development of young workers than smaller firms has long been argued in the literature. The question whether mobility has different effects for those leaving large firms is thus of particular interest. I therefore restrict the sample of apprentices to those 57% who graduate from establishments that employ at least 500 workers. While this is not representative of the German apprenticeship system as a whole, it is likely to be representative of large training programs within the system. Appendix Table D3 displays basic statistics for the sample of graduates from large firms. With respect to the full sample, one finds fewer women, slightly longer training durations, higher training wages, and a smaller fraction of movers.

The full analysis is repeated as before and Panel A of Figure 6 and Table 10 summarize the results. The first stages are again strong both with and without fixed effects (not shown).[59] However, the main results imply some important differences with respect to the full sample. First, the effect of moving is now more negative in all possible specifications. As will be further discussed below, this is likely to be due to the loss of firm-size wage-premia since movers from large firms often find jobs at smaller establishments. Second, the simple IV estimate is now more negative than the raw differences, suggesting that initial selection among larger firms with different average retention rates is stronger than in the full sample. Third, while the estimated wage differences from IV with firm fixed effects show a clear sign of decline over time it remains significantly negative even after five years in the labor market.[60]

These basic patterns recur if I restrict the instrument to be calculated with movers initially exiting the main labor force (the 'narrow' definition of the instrument – IV2), if I restrict workers to be in the sample all periods or if I impose stronger restrictions on labor force attachment (not shown). This suggests that for workers graduating from large establishments initial luck seems to matter even after five years in the labor market. To see this more directly, Figure E1 in the Appendix again shows the graph of the reduced form discussed above. While a rotation in the slopes is apparent, Figure E1 demonstrates how even five years after the effect of below and above average fraction 'other' movers in the year of graduation have significant effects on wages. To address this issue further Panel B of Figure 6 shows the persistence of luck for this group of workers directly. It

[59] The first stage is again shown at brackets of the demeaned instrument in Figure 4, Panel D. As suspected, for large firms the graphs show no significant non-linearity for small deviations of the instrument in the mean. Again, this suggests that the non-linearities observed for the full sample of firms was due to measurement problems arising from voluntary mobility.
[60] Note that as in other specifications, the differences estimated by IV without fixed effects also slightly decline over time. This may be an indication that workers leaving large firms with low retention rates are less negatively selected. However, the decline is much weaker (1.9% for IV vs. 4.5% for IVFE).

graphs the average wage deviation from training firm means for workers in the upper quartile, lower quartile and inter-quartile range of deviations of the fraction 'other' movers from firm-specific means. The result shows that those with the highest unexpected decreases in firms' retention rates of young trainees have initial wage losses but then recover towards the firm average. However, this recovery is only partial and the average difference in the fifth period is still significant. Similarly, those exposed to higher than usual retention rates experience declining but persistent wage gains with respect to the firm average.[61]

Large firms may provide exceptional career chances to young workers or they may simply pay higher wages to all their employees (e.g., Brown and Medoff 1989, Oi and Idson 1999). In either case, the wage loss of movers with respect to stayers should be driven by a change in the establishment size at which movers are currently employed. To gauge this possibility, Table 11 shows the transition matrix among employment size of the training establishment and size of the establishment *of the first job* (these are not the employing establishments in the first experience year). Movers leaving larger establishments are very likely to switch to smaller firms, suggesting that the persistent losses seen in Figure 6 for all specifications are driven by losses in firm size wage premia. This would be consistent with the results found by Krashinsky (2002). Using a data set similar to mine, Dustmann and Schoenberg (2002) find that the difference in wage losses of movers from firms with 50-500 and firms with 500+ employees is 5.5%. Consistent with the hypothesis of differential firm size premia, controls for establishment size reduce this difference to an insignificant percentage point.[62]

4.3. Sensitivity – Additional Specifications

The main results so far have been obtained by using firm fixed effects. However, it might be reasonable to ask whether there are permanent differences in average retention rates of young trainees along other dimensions, most notably sectors and geographical regions. Instead of taking

[61] The size of these estimates appears small, but to get a final effect they have to be divided by the coefficients from the first stage.
[62] See their Table 4. A formal regression analysis of these effects using my data stands out until the information on employment firm sizes at the three experience years becomes available.

firm fixed effects, I also conditioned alternatively on industry effects, region effects, and their interaction. The basic result is unchanged: estimated wage differences in these models are again permanent, and only deviations of retention rates from firm means are able to isolate unselected involuntary movers. However, including industry fixed effects in the basic IV estimates reduces the estimated wage differences among movers and stayers, such that these are now more similar to OLS than to the raw differences (compare Figure E2 in the appendix to Figure 2).[63] Moreover, industry fixed effects explain as much as firm fixed effects when included in OLS. These results suggest that part of the differences in retention rates and assignment between firms reflect differences at the industry level. This would be expected by the literature on industry wage differentials that has extensively documented systematic differences in wages across firms. Gibbons et al. (2002) suggest that these differences are partly driven by selection of workers with different levels of ability into firms. The literature so far has concentrated on differences in average wages, but the results in this paper suggest similar differences and sorting patterns also occur along industries' average retention rates of young employees. However, the fact that deviations of retention rates from industry averages do not seem to be able to isolate 'random' displacements suggests that there are permanent differences between firms even within industries.[64]

5. Interpretation – Sorting and Job Search

The discussion so far has interpreted the empirical results as suggestive of an environment in which establishments differ along key features that lead young apprentices to sort into firms at the start of training; moreover, the findings support an important role for voluntary job mobility. This interpretation of the results can be made more precise using the basic model of wage determination introduced in Section 2. As discussed in Section 2.2, each estimator is affected by a different mechanism underlying job and wage mobility. Thus, I can use the different estimates to learn about which theories appear to be necessary or particularly relevant for explaining the empirical patterns in

[63] Region fixed effects had some explanatory power for the OLS estimates, but very little effect on the IV estimates.
[64] This is consistent with recent findings in the literature that shows how even within industry, region or size groups there is considerable variation in job creation and destruction rates.

my sample of young apprentices.[65] The implications of the various models discussed in Section 2 are summarized in Table 12 and Figure 1. The following paragraphs briefly assess each of them in turn.

First, models of adverse selection in the labor market predict that wage losses of young apprentices leaving the training firm estimated by IV should be significantly smaller than those estimated by OLS. In case of initial sorting along firms' retention rates, the same prediction holds for IV with firm fixed effects. This is summarized in the first row of Table 12 and Panel A of Figure 1. Neither of the main predictions appears to hold: IV tends to be more negative than OLS at all experience years, and IVFE is initially more negative as well, only gradually reverting to zero.

Second, only models of initial assignment of young workers into training firms predict a significant role for firm fixed effects (Row 2, Table 12 and Panel B, Figure 1). Firm fixed effects turn out to be important in explaining wage losses, suggesting that initial sorting matters in this sample. Moreover, the fact that IV using firms' retention rates as instrument is more negative than OLS suggests that sorting occurs at least in part along firms' average turnover rates – firms with high retention rates attract less able workers. On the other hand, initial sorting cannot explain the catch-up pattern shown by IVFE. Moreover, the fact that OLSFE remains significantly negative suggests some degree of negative selection.

Third, the pattern of the preferred estimate, IVFE, can most readily be explained by a simple search model. A model of job search predicts an immediate loss of rents from the initial search for a better employer, and a gradual reversion as workers again search for better jobs over time. Moreover, the fact that the effect of a random displacement is initially *more* negative than what is estimated by OLS indicates a there is voluntary job mobility among young apprentices – again, consistent with job search.[66] However, search models also predict a decline in wage losses estimated by OLS, OLSFE, or IV, and this does not appear to be the case in the full sample. This is because

[65] Note that this approach can be understood as a special case of omitted variables – each theory can be thought of predicting the role of a different omitted variable, and each estimate as addressing one or more of these. The omitted variables would be those in equation (1): workers' unobserved abilities, average ability within firms, and detailed mobility status.

[66] The fact that the estimated wage differences from IV are more negative than OLS also suggests that the bias from voluntary mobility is greater than the bias from negative selection. OLS is biased by sorting along firms' retention rates plus the net bias from negative selection and voluntary mobility. IV, on the other hand, is only biased from initial sorting. Since IV is more negative than OLS, this suggests the net bias from within firm selection on OLS is positive.

workers displaced from a 'random' displacement are a minority among all workers leaving their training firm, and therefore receive a low weight among estimates pooling all groups of workers.[67]

Fourth, a model predicting negative selection in the absence of asymmetric information is one where initial assignment is imperfect and occurs gradually over time. Since both able and less able workers should switch firms as employers learn about them, OLS might be either positively or negatively biased. Since both OLS and OLSFE are permanently negative, this suggests that negative selection outweighs positive sorting based on ability.[68] Positive sorting plays also less of a role relative to initial assignment; as discussed in Section 2, sequential sorting suggests that better workers move out of firms with high turnover towards more desirable jobs. However, IV tends to be more negative than OLS, suggesting that those leaving firms with high turnover are less able, consistent with initial assignment. Overall, some gradual sorting with strong initial assignment would be consistent with data. However, since workers are always paid according to their marginal product, this model again predicts no true effects of a displacement, in contrast to what is shown by IVFE.[69]

Last, models of defined career paths, either within firms (internal labor markets) or between firms ("stepping stone" human capital accumulation) predict that leaving the training firm on wages may have permanent negative effects. While this is not the case in the full sample, permanent effects arise for workers leaving from larger firms. This is consistent with the existence of internal labor markets – they would also explain presence of some negative selection if wages of less able workers cannot be lowered. However, it would also be consistent with losses of more narrowly defined firm-size wage premia.

Not surprisingly, neither theory can explain all aspects of the data, nor does any theory completely fail. In this particular application, initial assignment of workers to firms along firms'

[67] One can see that some convergence arises in the panel sample or among those graduating from larger firms. Since these are both cases in which workers are likely to be generally of higher ability, this is consistent with the hypothesis that in the main sample random movers are outnumbered by selected movers.

[68] Note that positive sorting is not the same as voluntary mobility – the latter has a direct effect on wages since workers find better paying jobs, while the former occurs with all workers receiving their expected marginal product.

[69] Note that using as instrument the fraction movers who spent time in unemployment (IV2) confirms the general pattern found: firms with a high fraction of unemployed leavers appear to attract worse workers (IV2 is more negative). Moreover, changes in IV2 are not affected by voluntary mobility, and thus initial wage losses should be greater for the unselected sample (IVFE more negative) – and this is borne out by the data.

retention rates of young trainees appears to be a strong feature of the data. The result that sorting occurs along firms' turnover rates re-enforces and adds to the findings from the recent literature on the interaction between firm and worker heterogeneity using matched employer-employee data sets (Abowd et al. 1999, Margolis 1995). It is also consistent with the recent literature on heterogeneity in firms' growth and turnover rates in the cross-section (Davis and Haltiwanger 1990, Abowd et al. 1998). The next step will be to obtain further evidence on *permanent* differences in growth trends and turnover rates between firms. A second important feature of the data is the role of voluntary and beneficial job mobility among young German apprentices. This is consistent with the role of job search found for young American workers by Topel and Ward (1992), and matches related evidence on young apprentices in Germany (Euwals and Winkelmann 2001) and on young German workers in general (Section 3 and Appendix C of this paper).

Consistent with what others have found before, there also appears to be some of negative selection in the data. While this could be explained by models of adverse selection, these have a difficulty explaining the large amount of voluntary job change observed in the labor market. An alternative model explaining negative selection that is also consistent with the role of initial sorting, and is thus favored by the results of this paper, is that of sequential job search. However, another relevant explanation of negative selection for the German case is the presence of wage rigidities (Dustmann and Schoenberg 2002). If wages are rigid, for example due to the presence of collective bargaining or unions, then firms would again have an incentive to displace their worse workers.

6. Heterogeneity Among Firms

The main results of the paper highlight the importance of differences in firms' retention rates of young trainees and their potential effects for the study of the effects of displacements. To see this differences most clearly, Panel A of Figure 6 shows the distribution of the fraction 'other' movers among movers and stayers for all three cohorts. The distribution for stayers is heavily skewed to the left with a peak at firms who retain almost all other apprentices. Movers on the other hand are more equally distributed with a peak at firms with fraction 'other' movers equal to one. Panel B of Figure 6 shows the same distributions for deviations of fraction movers from firm means. The difference in the two figures is striking, as in the lower panel the differences among movers and stayers are considerably smaller. Movers still tend to come out of firms with above average fraction 'other' movers and this is the main variation exploited in the paper. But this simple comparison suggests that it is important to address permanent differences in the retention rates of young trainees when comparing movers and stayers in the German apprenticeship system.[70]

Table 13 suggests that establishments training apprentices in Germany differ among several dimensions, only one of which is the retention rate of the young workers they train. The table shows various characteristics of training firms at high, medium, and low average retention rates and makes three basic points. First, firms with lower average retention rates also differ in basic observable characteristics such as size or average training wages. Since these could in part explain lower wages of movers, the paper controls for them in the regression analysis. This also suggests that high- and low-turnover firms may differ along other characteristics as well, especially in the quality of training they provide. As suggested at the outset, workers graduating from high turnover firms may have not been worse workers initially but may have received worse training. Since it may have persistent consequences for earnings of younger workers, this will be an important question for future research. Second, firms with higher average retention rates seem to be more attractive employers after apprenticeship training as well. The lower panel of Table 13 shows that among these average job attachment for young apprentices after training is longer.[71] Last, it can be shown that there is a continuous and linear negative relationship between most of the characteristics shown in Table 13 and the fraction of workers leaving a training establishment at graduation. Average turnover rates appear to be an important of property of establishments that is systematically related to other basic

---

[70] The distributions in Panel B also show that the analysis is identified off of larger shocks around the mean. While the bulk of the distribution is concentrated around smaller deviations, the sample size is large enough to guarantee a sufficient number of people in the cells with high shocks. The graphs also pool the three cohorts of graduating apprentices, which provides additional identifying variation.

[71] As would be expected given initial assignment of more able workers into 'keeper' firms, wages paid after training are higher as well. Note that among firms with 50 or more employees unionization rates and coverage by regional or national wage-agreements are very high (90% and above); differences in industrial relations cannot be used to explain differences in turnover rates.

characteristics of firms.[72] The question why some firms appear to be 'career-firms,' when careers evolve within or between firms, and how this is related to training and other firm characteristics is another fruitful are of research.

7. Conclusion

Economists and policy makers have long been concerned that displacements and wage losses early in a career might lead to permanent disadvantages. A recent literature indeed suggests that young workers suffer persistent wage losses from an early job displacement. However, this paper has argued that simple estimates might overstate earnings losses if less able workers tend to be hired by firms with higher turnover rates. Such a process of initial assignment of workers to firms adds to any bias arising from negative selection because employers selectively displace their least able workers; or from positive selection because young workers leave voluntarily to take better jobs. Using longitudinal data on German apprentices and their training firms the paper addresses these complex selection and sorting mechanisms directly at the firm level. To measure the true long-run effect of leaving the training firm at the end of an apprenticeship, it uses changes in firms' retention rates of young apprentices as an instrument for displacement. This exploits variation in firms' hiring rates over time to best approximate a 'random' displacement. Moreover, by introducing firm fixed effects it uses workers at the same training firm as comparison group for displaced workers and thereby controls for initial assignment of workers to firms.

Simple comparisons of wages of those leaving the training firm and those staying suggest long-term effects of a displacement in a similar order of magnitude as found among American workers. However, controlling for selection *within* firms and sorting *between* firms the estimates show no permanent effects of an initial displacement. Instead, wage losses of leaving the training firm are initially large and then gradually revert to zero within the first five years of labor market experience.

[72] That turnover rates and wages are related has been suggested by the literature on efficiency wages, and Krueger and Summers (1988) provide some supporting evidence. Neal (1998) shows how sorting could explain the same phenomena.

Some persistence arises for workers leaving large training firms, consistent with the presence of firm-size wage premia or internal labor markets. Since each estimate is affected by different confounding factors, the results can also be used to learn about the importance of various mechanisms underlying job and wage mobility. In particular, the importance of firm fixed effects suggests a strong role for initial sorting of workers into firms. Less able workers appear to sort into firms with low retention rates and possibly lower quality of training. Second, the estimates suggest that voluntary and beneficial job mobility is very common among young apprentices, consistent with what has been found for young workers in the United States. Last, there appears to be negative selection as well, which could be due to ongoing sorting in the market as employers learn about workers' abilities. Alternatively, it could arise from adverse selection or the presence of wage rigidities in the German labor market.

These estimates suggest that particular care has to be taken to address the role of heterogeneity among firms and sorting of workers into firms when analyzing job changes of young workers. The fact that at least part of sorting occurs along firms' turnover rates is particularly relevant for the study of young displaced workers, since it implies displacement is not a random event even controlling for selection within firms. The results also confirm the role of heterogeneity among firms emphasized in the recent literature. However, much is as of yet unknown regarding permanent differences in firms' average turnover rates and their interaction with wages or training quality.

As suggested above, part of the effect attributed to initial assignment in this paper could also be due to differences in training quality. Initial access to high quality training could make an important impact on long-run career developments. For example, as suggested by Okun (1973), local labor market conditions could affect access to training, leading to permanent impacts on earnings. Preliminary results from the United States indeed suggest that entering the labor market in a recession permanently reduces earnings (von Wachter 2001b). In future work, the data of this paper can be used to obtain variation in the supply of apprenticeships and in cohort sizes of those finishing school at the local level to learn more about the role of initial labor market conditions.

References

Abowd, John, Patrick Corbel, and Francis Kramarz (1999). The Entry and Exit of Workers and the Growth of Employment: An Analysis of French Establishments'. *Review of Economics and Statistics* 81(2): 170-87.

Abowd, John, Francis Kramarz, and David Margolis (1999). 'High Wage Workers and High Wage Firms.' *Econometrica* 67(2): 411-45.

Acemoglu, Daron and Jörn-Steffen Pischke (1996). 'Why do Firms Train? Theory and Evidence.' *Quarterly Journal of Economics* 113(1): 79-119.

Anderson, Patricia M. and Bruce D. Meyer (1994). 'The Extent and Consequences of Job Turnover.' *Brookings Papers on Economic Activity. Microeconomics*: 177-236.

Angrist, Joshua D. and Alan B. Krueger (1992). 'The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples.' *Journal of the American Statistical Association* 87 (418): 328-336.

Antel, John J. (1985). 'Costly Labor Contract Renegotiation and the Labor Mobility of Young Men.' *American Economic Review*. 75: 976-991.

Arulampalam, Wiji, Paul Gregg, and Mary Gregory (2001). 'Unemployment Scarring.' *Economic Journal* 111: F577-F584.

Baker, George, Michael Gibbs, and Bengt Holmstrom (1994). 'The Internal Economics of the Firm: Evidence from Personnel Data.' *Quarterly Journal of Economics* 109: 881-919.

Bartel, Ann and George J. Borjas (1981). 'Wage Growth and Job Turnover: An Empirical Analysis.' In: S. Rosen (ed.) *Studies in Labor Markets*. Chicago, IL: University of Chicago Press.

Bauer, Thomas K. und Stefan Bender (2002). 'Technological Change, Organizational Change, and Job Turnover: A Descriptive Analysis for Germany.' IZA. Mimeo.

Beaudry, Paul and John DiNardo (1992). 'The Effect of Implicit Contracts on the Movements of Wages over the Business Cycle' *Journal of Political Economy:* 99(4).

Bender, Stefan, Anette Haas, and Christoph Klose (2000). 'IAB Employment Subsample 1975-1995: Opportunities for Analysis Provided by the Anonymised Subsample.' IZA Discussion Paper No. 117.

Bender, Stefan, Christian Dustmann, David Margolis, and Costas Meghir (1995). 'Worker Displacement in France and Germany.' The Institute for Fiscal Studies Working Paper Series No. W99/14.

Bender, Stefan, Jurgen Hilzendegen, Gotz Rohwer, and Helmut Rudolph (1996). 'Die IAB-Beschaftigtenstichprobe 1975-1990.' *Beitraege zur Arbeitsmarkt- und Berufsforschung* 197.

Berufsbildungsbericht (2001). Bonn, Germany: Bundesministerium fur Bildung und Forschung.

Blanchard, Olivier and Peter Diamond (1994). 'Ranking, Unemployment Duration, and Wages.' *Review of Economic Studies* 61: 417-34.

Blanchflower, David G. and Andrew J. Oswald (1994). *The Wage Curve*. Cambridge, MA and London: MIT Press.

Blau, Francine D. and Lawrence M. Kahn (1997). 'Gender and Youth Employment Outcomes: The U.S. and West Germany, 1984-1991.' NBER Working Paper No. 6078.

Buechtermann, Cristoph (1989). 'Befristete Arbeitsvertraege nach dem Beschaeftigtenfoerderungs-gesetz.' Forschungsvorhaben im Auftrag des Bundesministerium fuer Arbeit und Soziales.

Brown, Charles and James Medoff (1989). 'The Employer Size-Wage Effect.' *Journal of Political Economy* 97(5): 1027-59.

Burdett, Kenneth (1978). 'A Theory of Employee Job Search and Quit Rates.' *American Economic Review* 68(1): 212-20.

Couch, Kenneth A. (2001). 'Earnings Losses and Unemployment of Displaced Workers in Germany.' *Industrial and Labor Relations Review* 54(3): 559-72.

Davis, Steven J. and John Haltiwanger (1990). 'Gross Job Creation, Gross Job Destruction, and Employment Reallocation.' *Quarterly Journal of Economics* 107(3): 819-863.

Doeringer, Peter and Michael Piore (1971). *Internal Labor Markets and Manpower Analysis.* Lexington, MA: D. C. Heath and Company.

Dustmann, Christian and Uta Schoenberg (2002). 'Why do German Firms Offer Apprenticeship Training?' University College London. Mimeo.

Ellwood, David (1982). 'Teenage Unemployment: Permanent Scars or Temporary Blemishes?' In: Freeman, Richard and David Wise (eds.) *The Youth Labor Market Problem: its Nature, Causes and Consequences*. Chicago: U. Chicago Press.

Euwals, Rob and Rainer Winkelmann (2001). 'Why do Firms Train? Empirical Evidence of the First Labour market Outcomes of Graduated Apprentices.' IZA Discussion Paper No. 319.

Fallick, Bruce C. (1996). 'A Review of the Recent Empirical Literature on Displaced Workers.' *Industrial and Labor Relations Review* 50(1): 5-16.

Farber, Henry (1993). 'The Incidence and Cost of Job Loss: 1982-1991.' *Brookings Papers of Economic Activity: Microeconomics.* 1993: 73-132.

Farber, Henry (1994). 'The Analysis of Interfirm Worker Mobility.' *Journal of Labor Economics* 12: 554-93.

Farber, Henry and Robert Gibbons (1996). 'Learning and Wage Dynamics.' *Quarterly Journal of Economics* 111(4).

Franz, Wolfgang, Joachim Inkmann, Winfried Pohlmeier, and Volker Zimmermann (2000). 'Young and Out in Germany: On the Youth's Chances of Labor Market Entry in Germany.' In: D. Blanchflower and R. Freeman (eds.) *Youth Employment and Joblessness in Advanced Countries.* Chicago: Chicago University Press.

Gardecki, Rosella and David Neumark (1998). 'Order from Chaos? Effects of Early Labor Market Experiences on Adult Labor Market Outcomes.' *Industrial and Labor Relations Review* 51(2): 299-322.

Gibbons, Robert and Lawrence Katz (1991). 'Layoffs and Lemons.' *Journal of Labor Economics* 9(4): 351-380.

Gibbons, Robert and Lawrence Katz (1992). 'Does Unmeasured Ability Explain Inter-Industry Wage Differences?' *Review of Economic Studies* 59: 515-35.

Gibbons, Robert and Michael Waldman (1999). 'A Theory of Wage and Promotion Dynamics Inside Firms.' *Quarterly Journal of Economics* 114: 1321-58.

Gibbons, Robert, Lawrence Katz, Thomas Lemieux and Daniel Parent (2002). 'Comparative Advantage, Learning, and Sectoral Wage Determination.' NBER Working Paper No. 8889.

Greenwald, Bruce (1986). 'Adverse Selection in the Labor Market.' *Review of Economic Studies* 53: 325-47.

Groshen, Erica L. (1991). 'Sources of Inter-Industry Wage Dispersion: How Much Do Employers Matter?' *Quarterly Journal of Economics* 106(3): 869-84.

Grund- and Strukturdaten 2000/2001 (2001). Bonn, Germany: Bundesministerium fur Bildung und Forschung.

Gustafson, Cynthia K. (1998). 'Job Displacement and Mobility of Younger Workers.' Center for Labor Economics Working Paper No. 8.

Harhoff, Dietmar and Thomas J. Kane (1997). 'Is the Apprenticeship System a Panacea for the U.S. Labor Market?' *Journal of Population Economics* 10: 171-96.

Hashimoto, Masanori and Ross A. Miller (1997). 'How do Training and Early Labor Market Experience Affect the Economic Well Being of Youths?' Mimeo.

Heckman, James J. and George J. Borjas (1982). 'Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers From a Continuous Model of Heterogeneity and State Dependence.' *Economica*, New Series 42(187): 247-83.

Heckman, James J., Rebecca Roselius and Jeffrey Smith (1994). 'US Education and Training Policy: A Re-Evaluation of the Underlying Assumptions of the 'New Consensus." In: Solomon, Lewis and Alec Levenson (eds). *Labour Markets, Employment Policy and Job Creation*. Boulder: Westview Press.

Jacobson, Louis, Robert LaLonde and Daniel Sullivan (1992). 'Earnings Losses of Displaced Workers.' *American Economic Review* 83(4): 685-709.

Jovanovic, Boyan and B. Nyarko (1997). 'Stepping Stone Mobility.' *Carnegie-Rochester Conference Series in Public Policy* 46: 289-337.

Kletzer, Lori G. (1998). 'Job Displacement.' *Journal of Economic Perspectives* 12(1): 115-36.

Kletzer, Lori G. and Robert W. Fairlie (2001). 'The Long-Term Costs of Job Displacement for Young Adult Workers.' University of California Santa Cruz. Mimeo.

Krashinsky, Harry (2002). 'Evidence on Adverse Selection and Establishment Size In the Labor Market.' *Industrial and Labor Relations Review 56(1): 84-96.*

Krueger, Alan B. and Lawrence H. Summers (1987). 'Efficiency Wages and the Inter-Industry Wage Structure.' *Econometrica* 56: 259-94.

Machin, Stephen and Alan Manning (1999). 'The Causes and Consequences in Long-Term Unemployment in Europe.' In: O. Ashenfelter and D. Card (eds.) *Handbook of labor economics*. Volume 3A. Amsterdam, New York, and Oxford: Elsevier Science.

Manning, Alan (2001). *Monopsony in Motion.* (forthcoming).

Margolis, David (1995). 'Firm Heterogeneity and Worker Self-Selection Bias Estimated Returns to Seniority.' CIRANO Working Paper No. 95s-4.

Margolis, David, Veronique Simonnet and Lars Vilhuber (2000). 'Using Early Career Experiences and Later Career Outcomes to Distinguish Between Models of Labor Market Behavior Under Institutional Constraints.' CNRS Working Paper. Universite Paris 1.

McLaughlin, Kenneth J. (1991). 'A Theory of Quits and Layoffs with Efficient Turnover.' *Journal of Political Economy* 99: 1-29.

Mincer, Jacob (1986). 'Wage Changes in Job Changes.' In: R.G. Ehrenberg (ed.) *Research in Labor Economics*. Vol. 8.Greenwich, CT: JAI Press.

Mincer, Jacob and Boyan Jovanovic (1981). 'Labor Mobility and Wages.' In: Rosen, Sherwin (ed), *Studies in Labor Markets*. Chicago: University of Chicago Press.

Murphy, Kevin M. and Finis Welch (1992). 'The Structure of Wages.' *Quarterly Journal of Economics* 107(1): 285-326.

Murphy, Kevin M. and Robert H. Topel (1985). 'Estimation and Inference in Two-Step Econometric Models.' J*ournal of Business and Economics Statistics* 3(4): 370-79.

Neal, Derek (1998). 'The Link Between Ability and Specialization: An Explanation for Observed Correlations Between Wages and Mobility Rates.' *Journal of Human Resources* 33(1): 173-200.

Neumark, David (1998). 'Youth Labor Market in the U.S.: Shopping Around vs. Staying Put.' NBER Working Paper No. 6581.

Oi, Walter Y. and Todd L. Idson (1999). 'Firm Size and Wages.' In: O. Ashenfelter and D. Card (eds.) *Handbook of labor economics*. Volume 3A. Amsterdam, New York, and Oxford: Elsevier Science.

Okun, Arthur M. (1973). 'Upward Mobility in a High-Pressure Economy.' *Brookings Papers of Economic Activity* 1: 207-52.

Rosen, Sherwin (1972). 'Learning and Experience in the Labor Market.' *Journal of Human Resources* 7(3).

Ruud, Paul A. (2000). *An Introduction to Classical Econometric Theory*. New York, Oxford: Oxford University Press.

Ryan, Paul (2001). 'The School-To-Work Transition: A Cross-National Perspective.' *Journal of Economic Literature* 39 (March): 34-92.

Salop, Joanne and Steven Salop (1976). 'Self-Selection and Turnover in the Labor Market.' *Quarterly Journal of Economics* 90.

Stern, David, N. Finkelstein, J. R. Stone III, J. Latting, and C. Dornsife (1995) *School to Work: Research on Programs in the United States*. London and Washington: Falmer Press.)

Topel, Robert and Michael Ward (1992). 'Job Mobility and the Careers of Young Men.' *Quarterly Journal of Economics* 107(2): 439-479.

von Wachter, Till (2001a). 'A Model of Learning and Adverse Selection.' University of California Berkeley. Mimeo.

von Wachter, Till (2001b). 'The Persistence of Bad Luck – Long Term Effects on Earnings of Beginning to Work in a Recession.' University of California Berkeley. Mimeo.

Waldman, Michael (1984). 'Job Assignment, Signaling, and Efficiency.' *Rand Journal of Economics* 15: 255-67.

Weiss, Andrew and Ruqu Wang (1998). 'A Sorting Model of Labor Contracts: Implications for Layoffs and Wage-Tenure Profiles.' *Labour Economics* 5(3): 359-83.

Winkelmann, Rainer (1996). 'Employment Prospects and Skill Acquisition of Apprenticeship-Trained Workers in Germany.' *Industrial and Labor Relations Review* 49(4): 658-672.

Appendix A

The following appendix briefly summarizes the basic estimators and relates them to the theoretical models presented in Section 2. Suppose the goal is to estimate the effect of leaving the training firm on wages for $t = 1,3,5$ periods of potential labor market experience of young workers after the end of the apprenticeship. For notational simplicity, dependence on observable characteristics other than a constant is suppressed. The estimation procedure itself is discussed in Appendix B. The equations estimated for OLS and OLSFE are

$$w_{it} = \mu + \delta_t D_{i0} + \varepsilon_{it}$$
$$w_{it} = \mu + \delta_t D_{i0} + \phi_j + \varepsilon_{it}$$

The resulting estimates are calculated as

$$\Rightarrow \hat{\delta}_t^{OLS} = \frac{\operatorname{cov}(w_{it}, D_{i0})}{\operatorname{var}(D_{i0})}$$
$$\Rightarrow \hat{\delta}_t^{OLSFE} = \frac{\sum(w_{it} - \overline{w}_{j(i)})(D_{i0} - \overline{D}_{j(i)})}{\sum(D_{i0} - \overline{D}_{j(i)})^2}.$$

If D_{i0} is instrumented by the fraction of other apprentices moving out of the training firm, the corresponding IV estimates are calculated from

$$\Rightarrow \hat{\delta}_t^{IV} = \frac{\operatorname{cov}(w_{it}, z_{ij(i)})}{\operatorname{cov}(D_{i0}, z_{ij(i)})}$$
$$\Rightarrow \hat{\delta}_t^{IVFE} = \frac{\sum(w_{it} - \overline{w}_{j(i)})(z_{ij(i)} - \overline{z}_{j(i)})}{\sum(D_{i0} - \overline{D}_{j(i)})(z_{ij(i)} - \overline{z}_{j(i)})},$$

where the necessary assumptions are $\operatorname{cov}(a_i - \overline{a}_{j(i)}, z_{ij(i)}) = 0$ and $\operatorname{cov}(V_{i0}, z_{ij(i)}) = 0$ for IV and $\operatorname{cov}(a_i - \overline{a}_{j(i)}, z_{ij(i)} - \overline{z}_{j(i)}) = 0$ and $\operatorname{cov}(V_{i0} - \overline{V}_{j(i)}, z_{ij(i)} - \overline{z}_{j(i)}) = 0$ for IVFE.

The core mechanisms of the five theoretical approaches to job and wage mobility discussed in the text imply the following stylized wage equations:

Adverse Selection

$$\Rightarrow w_{it} = \gamma + a_i + \varepsilon_{it}$$

Initial Assignment

$$\Rightarrow w_{it} = \gamma + (a_i - \overline{a}_{j(i)}) + \overline{a}_{j(i)} + \varepsilon_{it}$$

Job Search

$$\Rightarrow w_{it} = \gamma + \delta_{It} D_{i0} + (\delta_{Vt} - \delta_{It}) V_{i0} + \varepsilon_{it}$$

Sequential Sorting

$$\Rightarrow w_{it} = \gamma + (a_i - \overline{a}_{j(i)}) + \overline{a}_{j(i)} + \varepsilon_{it}$$

Institutions/Rosen

$$\Rightarrow w_{it} = \gamma + (a_i - \overline{a}_{j(i)}) + \overline{a}_{j(i)} + \delta_{It} D_{i0} + \varepsilon_{it}.$$

Combining these data generating processes with the various statistical procedures, for each theory in turn one obtains the following estimates.

Adverse Selection

$$\Rightarrow \hat{\delta}_{It}^{OLS} = \frac{\text{cov}(a_i, D_{i0})}{\text{var}(D_{i0})} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{OLSFE} = \frac{\text{cov}(a_i - \overline{a}_{j(i)}, D_{i0} - \overline{D}_{j(i)})}{\text{var}(D_{i0} - \overline{D}_{j(i)})} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{IV} = 0$$

$$\Rightarrow \hat{\delta}_{It}^{IVFE} = 0$$

As mentioned in the text, if there is some initial sorting based on firms' retention rates, one would expect that IV is still negative (potentially more than OLS), but that IVFE is zero.

Initial Assignment

$$\Rightarrow \hat{\delta}_{It}^{OLS} = \frac{\text{cov}(\overline{a}_{j(i)}, D_{i0})}{\text{var}(D_{i0})} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{OLSFE} = 0$$

$$\Rightarrow \hat{\delta}_{It}^{IV} = \frac{\text{cov}(\overline{a}_{j(i)}, z_{ij(i)})}{\text{cov}(D_{i0}, z_{ij(i)})} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{IVFE} = 0$$

Under initial assignment, IV would be more negative than OLS since $\text{cov}(\overline{a}_{j(i)}, z_{ij(i)}) = \text{cov}(\overline{a}_{j(i)}, D_{i0})$ but $\text{cov}(D_{i0}, z_{ij(i)}) < \text{var}(D_{i0})$. The weaker the instrument, the larger will the difference be.

Job Search

$$\Rightarrow \hat{\delta}_{It}^{OLS} = \delta_{It} + (\delta_{Vt} - \delta_{It})\frac{\text{cov}(V_{i0}, D_{i0})}{\text{var}(D_{i0})}$$

$$\Rightarrow \hat{\delta}_{It}^{OLSFE} = \delta_{It} + (\delta_{Vt} - \delta_{It})\frac{\text{cov}(V_{i0} - \overline{V}_{j(i)}, D_{i0} - \overline{D}_{j(i)})}{\text{var}(D_{i0} - \overline{D}_{j(i)})}$$

$$\Rightarrow \hat{\delta}_{It}^{IV} = \delta_{It} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{IVFE} = \delta_{It} < 0$$

The basic search model assumes individuals are equal but jobs differ. Thus, neither selection based on unobserved ability nor initial sorting matters. As discussed in the text, the model predicts that losses from an involuntary displacement are temporary.

Sequential Sorting

$$\Rightarrow \hat{\delta}_{It}^{OLS} = \frac{\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})} + \frac{\operatorname{cov}(\bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})}$$

$$\Rightarrow \hat{\delta}_{It}^{OLSFE} = \frac{\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})}$$

$$\Rightarrow \hat{\delta}_{It}^{IV} = \frac{\operatorname{cov}(a_i - \bar{a}_{j(i)}, z_{ij(i)})}{\operatorname{cov}(D_{i0}, z_{ij(i)})} + \frac{\operatorname{cov}(\bar{a}_{j(i)}, z_{ij(i)})}{\operatorname{cov}(D_{i0}, z_{ij(i)})}$$

$$\Rightarrow \hat{\delta}_{It}^{IVFE} = 0$$

In the case of sequential sorting, the bias of OLS coming from within firm selection is ambiguous; on the one hand, good firms will separate from their least able workers, such that $\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{0i}) < 0$. On the other hand, less desirable firms will separate from their better workers, implying $\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{0i}) > 0$. The sign of the bias is indicated by OLSFE, which is net of the effect of initial assignment (and less negative than OLS). Since good workers are sorted out of firms with high average turnover rates (low average retention rates), $\operatorname{cov}(a_i - \bar{a}_{j(i)}, z_{ij(i)}) > 0$. This implies that IV is less negative than OLS unless initial assignment is very strong (i.e., unless $\operatorname{cov}(\bar{a}_{j(i)}, z_{ij(i)}) / \operatorname{cov}(D_{i0}, z_{ij(i)})$ is very negative).

Institutions/Rosen

$$\Rightarrow \hat{\delta}_{It}^{OLS} = \delta_{It} + \frac{\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})} + \frac{\operatorname{cov}(\bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{OLSFE} = \delta_{It} + \frac{\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0})}{\operatorname{var}(D_{i0})} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{IV} = \delta_{It} + \frac{\operatorname{cov}(\bar{a}_{j(i)}, z_{ij(i)})}{\operatorname{cov}(D_{i0}, z_{ij(i)})} < 0$$

$$\Rightarrow \hat{\delta}_{It}^{IVFE} = \delta_{It} < 0$$

In the case that some larger firms have internal labor markets and provide sheltered and well-defined career paths, the loss of such an opportunity could imply permanent negative effects. However, if wages are rigid in the internal labor market these firms might fire the worst workers instead of lowering their wages, thus $\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) < 0$. Similarly, these firms are more desirable, and they might attract the best workers, suggesting that $\operatorname{cov}(\bar{a}_{j(i)}, D_{i0}) < 0$ and $\operatorname{cov}(\bar{a}_{j(i)}, z_{ij(i)}) < 0$. OLSFE is predicted to be less negative than OLS, while IV could be more negative if $\operatorname{cov}(D_{i0}, z_{ij(i)})$ is very small. Under the simple sequential human capital accumulation, wages are assumed to be fully flexible and thus $\operatorname{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) = 0$. Similarly, the model has no predictions regarding the supply of career jobs between different firms, and thus $\operatorname{cov}(\bar{a}_{j(i)}, D_{i0}) = 0$.

Appendix B

The objective is to estimate the effect of moving from the training firm on wages three years after the end of training. The data available is from three graduating cohorts of apprentices who each are observed three years. Omitting the cohort subscript let D_{ij0} be the mover status at the end of training of worker i leaving training firm j; let w_{ijt} be log real wages; let x_{ij0} denote observable characteristics of workers, training, and training firm; and let ϕ_{i0}^{j} and ϕ_{i0}^{c} be training firm and cohort dummies, respectively. The aim is to estimate the effect of moving on wages in the regressions

$$y_{ijt} = D_{ij0}\delta_t + x_{ij0}\beta_t + \sum_{j=1}^{J} \gamma_j \phi_{i0}^{j} + \sum_{c=1}^{3} \gamma_{ct}\phi_{i0}^{c} + \varepsilon_{ijt}, \quad t = 1,3,5,$$

with or without the training firm fixed effects (where J denotes the number of firms). The main parameters of interest are $\delta_1, \delta_3, \delta_5$. To estimate these, the observations are stacked into a panel

$$\begin{bmatrix} w_1 \\ w_3 \\ w_5 \end{bmatrix} = \begin{bmatrix} D_1 \\ 0 \\ 0 \end{bmatrix}\delta_1 + \begin{bmatrix} 0 \\ D_3 \\ 0 \end{bmatrix}\delta_3 + \begin{bmatrix} 0 \\ 0 \\ D_5 \end{bmatrix}\delta_5 + \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}\beta_1 + \begin{bmatrix} 0 \\ x_3 \\ 0 \end{bmatrix}\beta_3 + \begin{bmatrix} 0 \\ 0 \\ x_5 \end{bmatrix}\beta_5 + \begin{bmatrix} \Phi_1^j \\ \Phi_3^j \\ \Phi_5^j \end{bmatrix}\gamma + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_3 \\ \varepsilon_5 \end{bmatrix},$$

where x now all control variables with time varying effects. In matrix format this equation can be rewritten as

$$W = \Lambda_1 D\delta_1 + \Lambda_2 D\delta_2 + \Lambda_3 D\delta_3 + \Lambda_1 X\beta_1 + \Lambda_2 X\beta_2 + \Lambda_3 X\beta_3 + \Phi^j\gamma + \bar{\varepsilon}, \qquad \text{(B1)}$$

where $W = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$, $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, $D = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix}$, $\Phi^j = \begin{bmatrix} \Phi_1^j \\ \Phi_2^j \\ \Phi_3^j \end{bmatrix}$, $\bar{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$, and for ease of notation I have defined $\Lambda_1 = \begin{bmatrix} I_{N_1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\Lambda_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_{N_2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\Lambda_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_{N_3} \end{bmatrix}$. The dimension of each vector ξ_t is $N_t \times 1$, such that the full vectors $\Xi = [\xi_1', \xi_2', \xi_3']'$ are $N \times 1$ with $N = N_1 + N_2 + N_3$ being the total number of observations.

System $(B1)$ is basically a SUR model with cross-equation restrictions. The equations are related because $\mathrm{cov}(\varepsilon_{ijt}, \varepsilon_{ijs}) \neq 0$ and potentially $\mathrm{cov}(\varepsilon_{ijt}, \varepsilon_{ljs}) \neq 0$ for $s \neq t$; the latter correlation could for example arise because of common shocks to the training firm. OLS and OLSFE are not efficient but consistent under these assumptions on the error structure.

To take into account the potential correlations of individuals and trainees within the same training firm, both models are calculated using STATA's cluster procedure to cluster standard errors at the training firm level.

Equation $(B1)$ is also supposed to be estimated with instrumental variables since it is suspected that $E\{D_{ij0}\varepsilon_{ijt}\} \neq 0,\ \forall\ t = 1,3,5,\ i = 1,...,N$. Consider the vector of instruments $Z = [z'_1, z'_2, z'_3]'$ for D, for which it is assumed that $E\{z_{ij(i)}\varepsilon_{ijt}\} = 0,$ $\forall\ t = 1,3,5,\ i = 1,...,N$. Then IV can be implemented by first regressing D on Z and the other variables of the model in a first stage of the form

$$D = \Lambda_1 Z\pi_1 + \Lambda_2 Z\pi_2 + \Lambda_3 Z\pi_3 + \Lambda_1 X\tilde{\beta}_1 + \Lambda_2 X\tilde{\beta}_2 + \Lambda_3 X\tilde{\beta}_3 + \Phi^j \tilde{\gamma} + \bar{u} \qquad \text{(B2)}$$

and regressing W on Z in the reduced form

$$W = \Lambda_1 Z\phi_1 + \Lambda_2 Z\phi_2 + \Lambda_3 Z\phi_3 + \Lambda_1 X\breve{\beta}_1 + \Lambda_2 X\breve{\beta}_2 + \Lambda_3 X\breve{\beta}_3 + \tilde{\Phi}^j + \tilde{\bar{\varepsilon}}\,. \qquad \text{(B3)}$$

The IV estimator for $\delta = (\delta_1, \delta_3, \delta_5)$ is then obtained by dividing the first and second stage coefficients, i.e., $\hat{\delta}_t = \dfrac{\hat{\phi}_t}{\hat{\pi}_t},\ t = 1,3,5$.[52]

To see that this is a consistent estimator for δ, first substitute stage $(B2)$ into the structural equation $(B1)$ to obtain

$$W = \Lambda_1 \left[\Lambda_1 Z\pi_1 + \Lambda_2 Z\pi_2 + \Lambda_3 Z\pi_3 + \Lambda_1 X\tilde{\beta}_1 + \Lambda_2 X\tilde{\beta}_2 + \Lambda_3 X\tilde{\beta}_3 + \Phi^j\tilde{\gamma} + \bar{u}\right]\delta_1 + ... + \tilde{\bar{\varepsilon}}$$
$$\Leftrightarrow W = \Lambda_1\left[Z\pi_1 + X\tilde{\beta}_1 + \Phi^j\tilde{\gamma} + \bar{u}\right]\delta_1 + \Lambda_2\left[Z\pi_2 + X\tilde{\beta}_2 + \Phi^j\tilde{\gamma} + \bar{u}\right]\delta_2 + \Lambda_3\left[Z\pi_3 + ... + \bar{u}\right]\delta_3 + ... + \tilde{\bar{\varepsilon}},$$

where I have used the fact that $\Lambda_1\Lambda_2 = \Lambda_1\Lambda_3 = \Lambda_2\Lambda_3 = 0$. Rearranging terms one obtains that $Y = \Lambda_1 Z\pi_1\delta_1 + \Lambda_2 Z\pi_2\delta_2 + \Lambda_3 Z\pi_3\delta_3 + \Lambda_1 F\breve{\beta}_1 + \Lambda_2 F\breve{\beta}_2 + \Lambda_3 F\breve{\beta}_3 + \tilde{\Phi}^j + \tilde{\bar{\varepsilon}}$, where $\breve{\beta}_t \equiv \beta_t + \tilde{\beta}_t\delta_t$, $\tilde{\Phi}^j \equiv \Phi^j\gamma + \Lambda_1\Phi^j\tilde{\gamma}\delta_1 + \Lambda_2\Phi^j\tilde{\gamma}\delta_2 + \Lambda_3\Phi^j\tilde{\gamma}\delta_3$ and $\tilde{\bar{\varepsilon}} \equiv \bar{\varepsilon} + \Lambda_1\bar{u}\delta_1 + \Lambda_2\bar{u}\delta_2 + \Lambda_3\bar{u}\delta_3$. Since we have $E\{z_t u_t\} = 0$ and $E\{z_t\varepsilon_t\} = 0$, it follows that $E\{(\Lambda_t Z)'\tilde{\bar{\varepsilon}}\} = E\{z_t(\varepsilon_t + u_t\delta_t)\} = E\{z_t\varepsilon_t\} + \delta_t E\{z_t u_t\} = 0\ \ \forall\ t = 1,3,5$. Thus, OLS estimation yields consistent estimators in the first stage and the reduced form, i.e.,

[52] The IV estimator can also be obtained by estimating the fitted value $\hat{X}$ from the first stage in equation (B2) and substituting it into equation (B1) for X to obtain the second stage. The coefficients on $[\Lambda_1\hat{X}, \Lambda_2\hat{X}, \Lambda_3\hat{X}]$ are then the IV-estimates. This amounts to using $[\Lambda_1\hat{X}, \Lambda_2\hat{X}, \Lambda_3\hat{X}]$ as instruments for $[\Lambda_1 X, \Lambda_2 X, \Lambda_3 X]$ and thus the same proof of consistency applies. Note that this is not the same as treating $[\Lambda_1 X, \Lambda_2 X, \Lambda_3 X]$ as three separate endogenous variables and using $[\Lambda_1 Z, \Lambda_2 Z, \Lambda_3 Z]$ as instruments in three separate first stages.

$p\lim\hat{\phi}_t = \pi_t \delta_t$ and $p\lim\hat{\pi}_t = \pi_t$. Then the Slutsky-Theorem implies

$\hat{\delta}_t^{IV} = \dfrac{\hat{\phi}_t}{\hat{\pi}_t} \rightarrow \delta_t$ as $n \rightarrow \infty$.

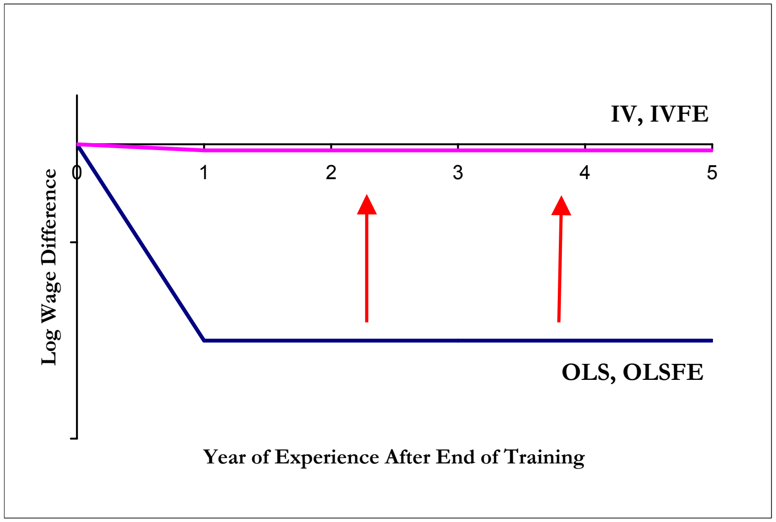
If we assume that the variance matrix of the error terms is scalar (this is only explained for clarity and <u>not</u> what is done in the paper), by the partitioned inverse formula the correct variance matrix for the IV estimator is $\text{var}\{\hat{\delta}\} = \sigma_{IV}^2 \left(\widehat{D}_{[3]}' M_B \widehat{D}_{[3]}\right)^{-1}$, where $\widehat{D}_{[3]} = \left[\Lambda_1 \widehat{D}, \Lambda_2 \widehat{D}, \Lambda_3 \widehat{D}\right]$ and $B \equiv [X, \Phi]$. The variance term is estimated consistently by $\hat{\sigma}_{IV}^2 = \dfrac{1}{N-K}\left(\widehat{Y} - \widehat{D}\hat{\delta}^{IV}\right)'\left(\widehat{Y} - \widehat{D}\hat{\delta}^{IV}\right)$, where $\widehat{D} \equiv M_{[X,\Phi]}\left[\Lambda_1 D, \Lambda_2 D, \Lambda_3 D\right]$ and $\widehat{Y} \equiv M_{[X,\Phi]} Y$. In the paper, $\text{var}\{\vec{\varepsilon}\}$ will be assumed to be block diagonal, such that

$$\text{var}\{\hat{\delta}\} = \left(\widehat{X}' M_B \widehat{X}\right)^{-1} \widehat{X}' M_B \,\text{var}\{\vec{\varepsilon}\} M_B \widehat{X} \left(\widehat{X}' M_B \widehat{X}\right)^{-1}.$$

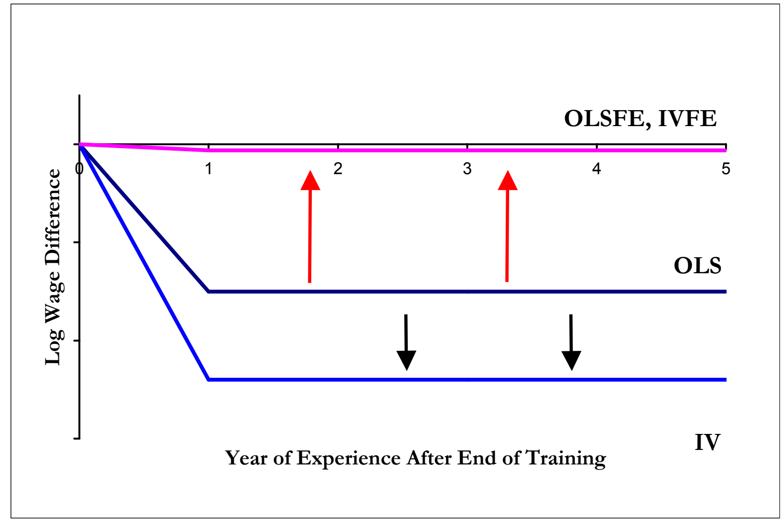
The fitted residuals implied by the IV estimates are again used to calculate the blocks of the variance-covariance matrix. This procedure can be implemented using STATA’s cluster sub-routine.

Figure 1: Implications from Basic Models of Job and Wage Mobility for Different Estimates of Wage Losses of Apprentices Who Leaving Training Firm At Graduation

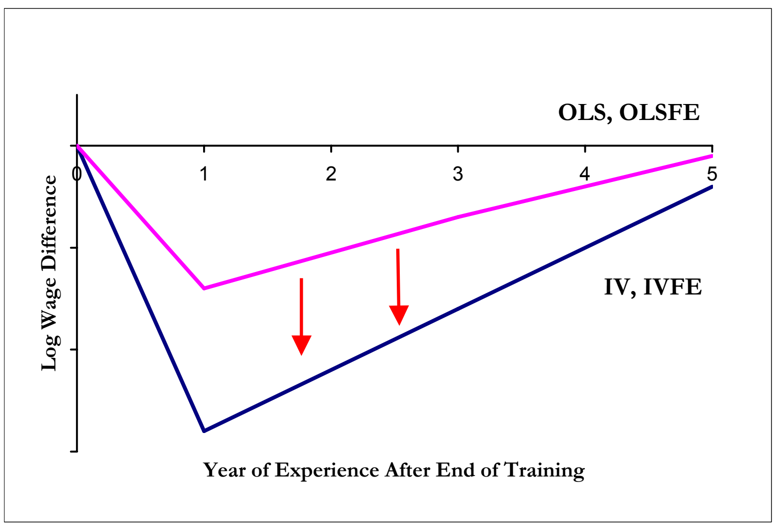
Panel A: IV, FE and Adverse Selection



Panel B: IV, FE and Initial Assignment



Panel C: IV, FE and Job Search



Panel D: IV, FE and Sequential Sorting

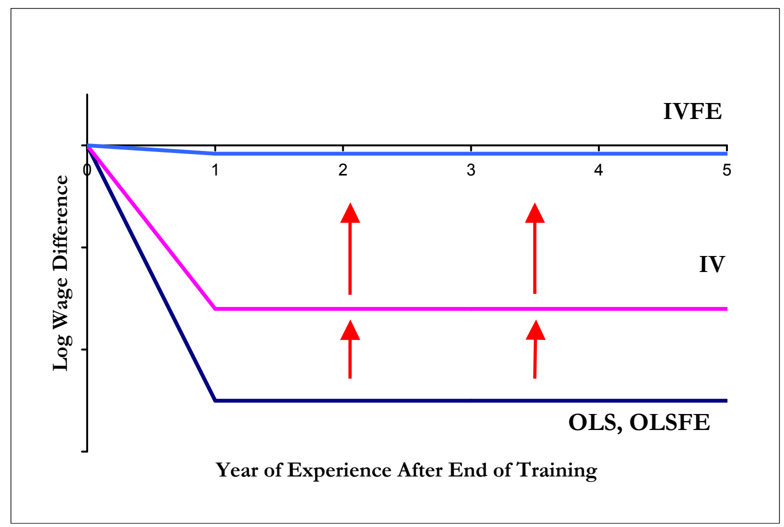


Figure 2: OLS and IV Estimates of Wage Losses from Leaving Training Firm At Graduation, With and Without Firm Fixed Effects - Main Sample

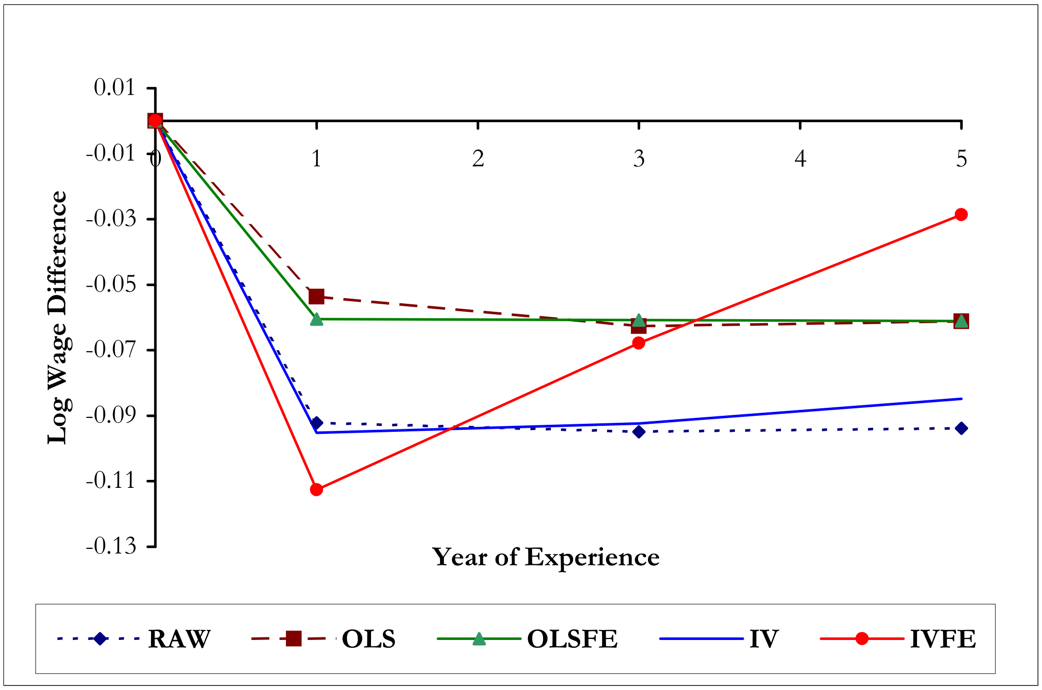


Figure 3: Average Real Wages By Intervals of Fraction 'Other' Apprentices Moving from Training Firm at End of Training

Panel A: Without Firm Fixed Effects | **Panel B: With Firm Fixed Effects**

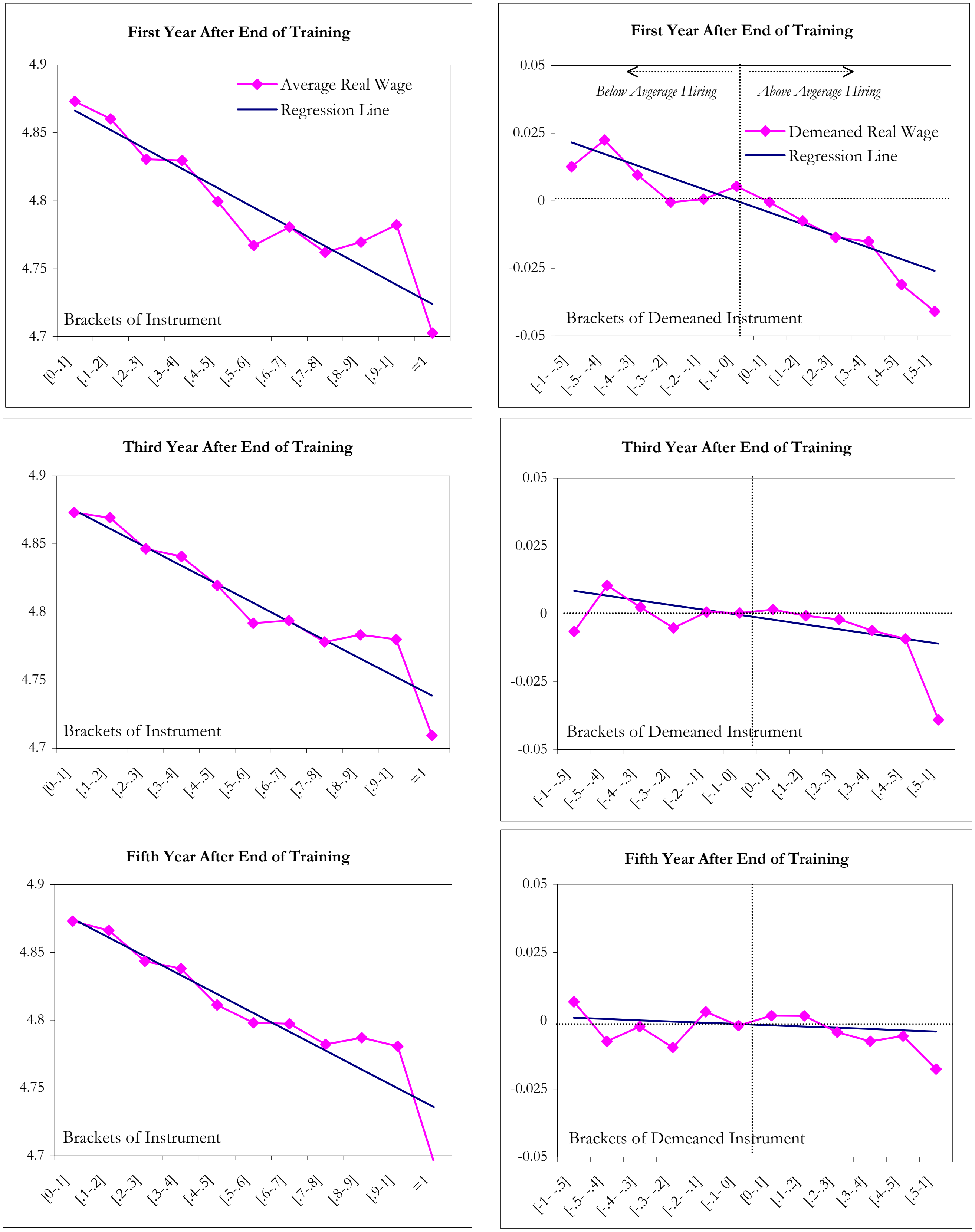
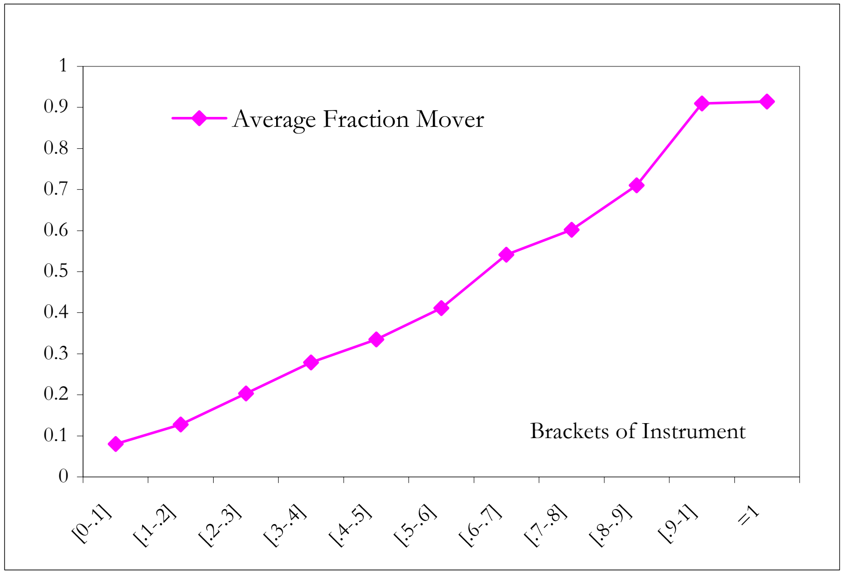
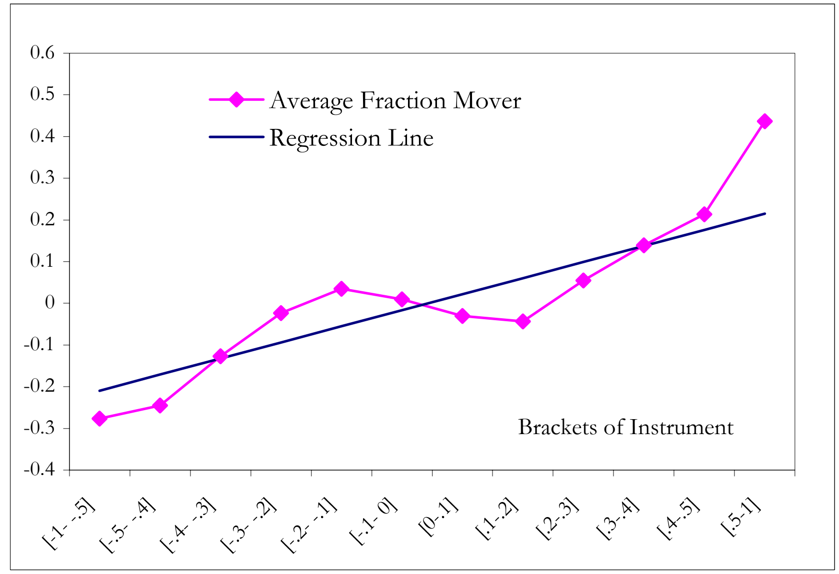


Figure 4: Fraction Movers by Intervals of Instrument, With or Without Firm Fixed Effects

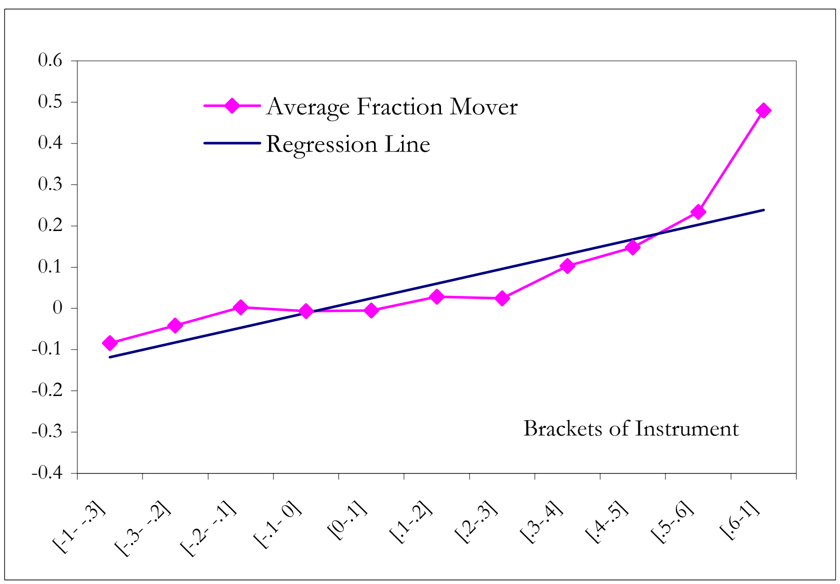
Panel A: Without Fixed Effects, Main Sample



Panel B: With Firm Fixed Effects, Main Sample



Panel C: With Firm Fixed Effects, Unemployed Movers (IV2)



Panel D: With Firm Fixed Effects, Large Firms

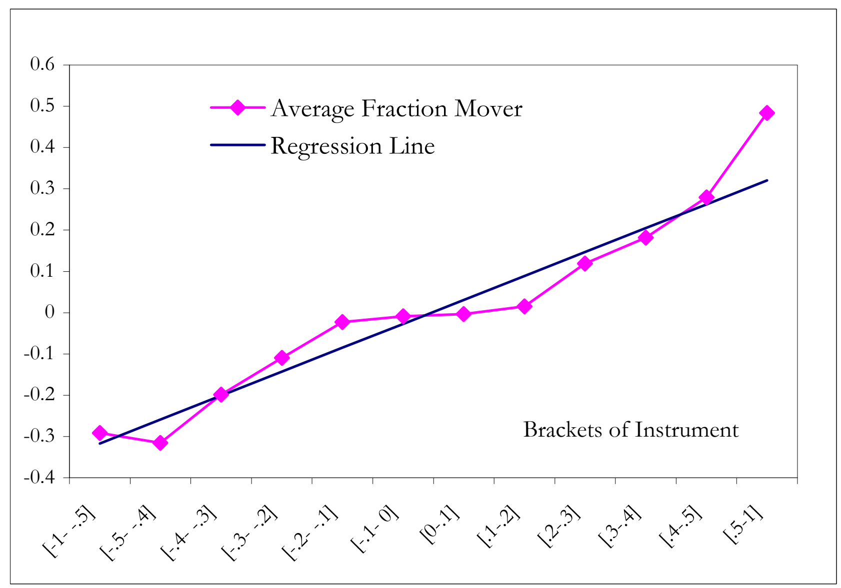
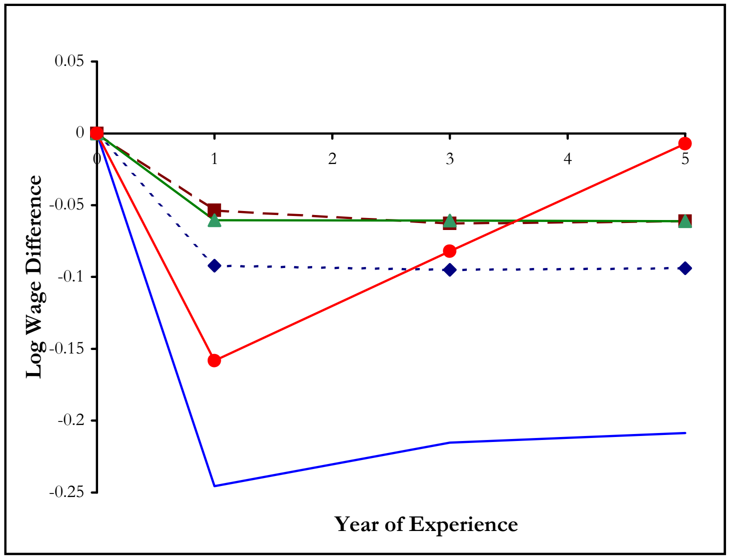
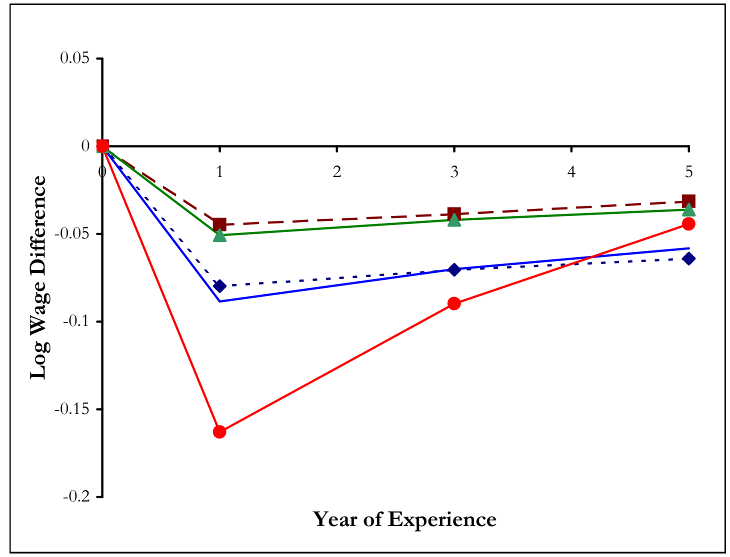


Figure 5: Sensitivity - Effects on Wages of Leaving Training Firm at Graduation, Various Specifications

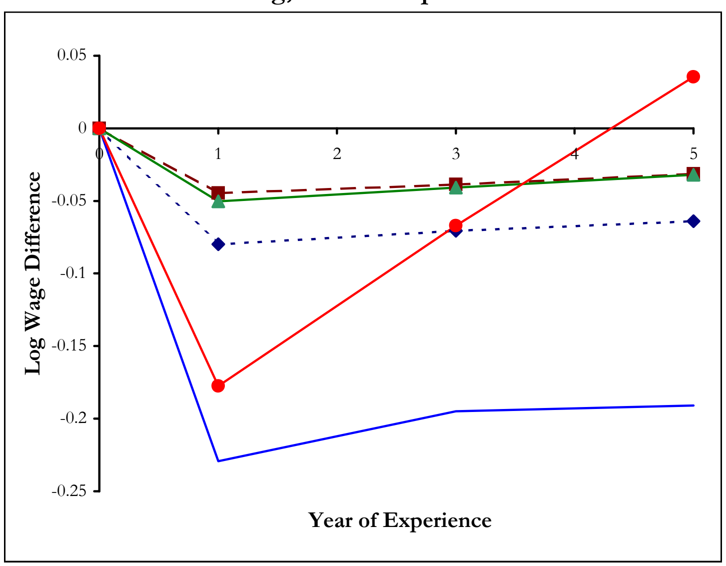
Panel A: Effect of Moving, Fraction Unemployed (IV2)



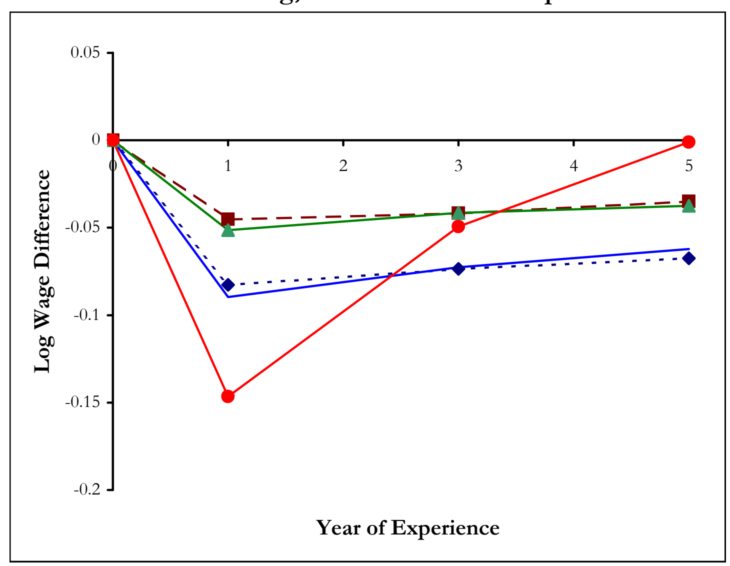
Panel B: Effect of Moving, Panel Sample



Panel C: Effect of Moving, Panel Sample with IV2



Panel D: Effect of Moving, Restricted OLF Sample



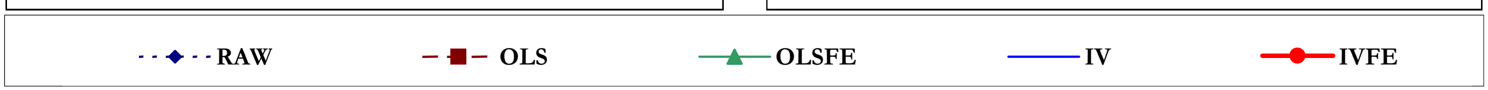
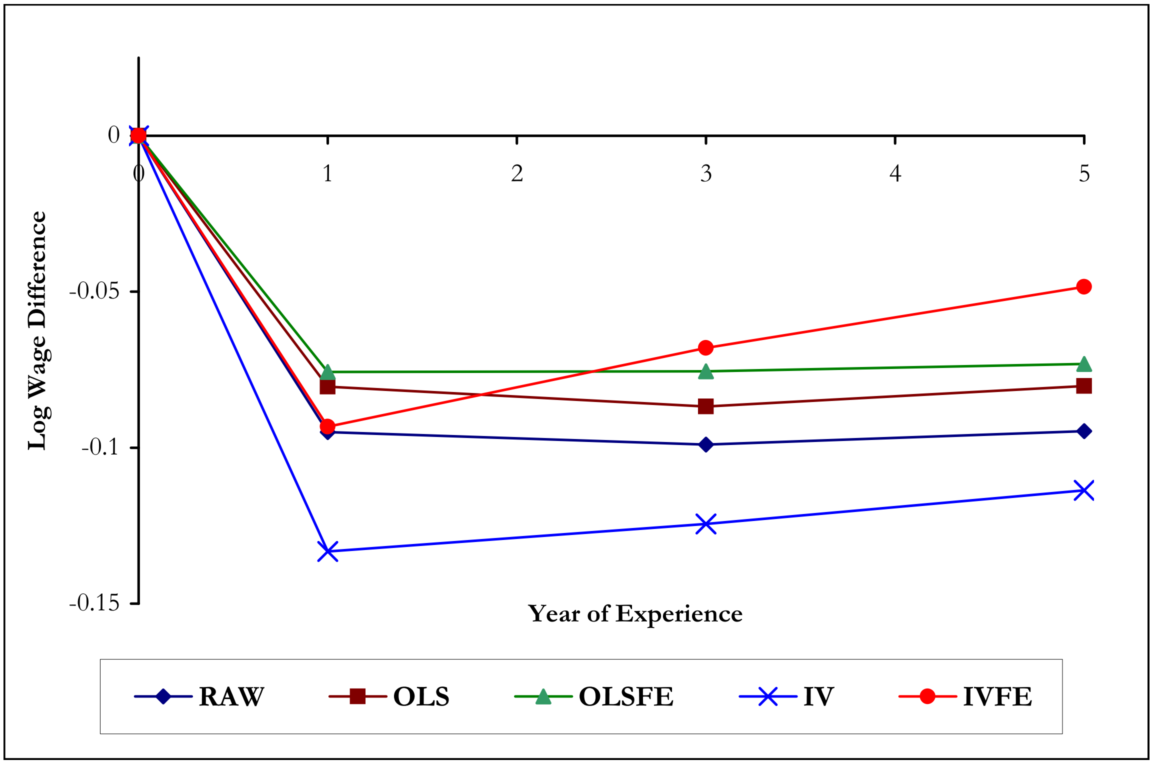


Figure 6: Effect on Wages of Leaving Very Large Training Firms

Panel A: OLS and IV Estimates of Wage Differences Between Movers and Stayers With and Without Firm Fixed Effects



Panel B: Wage Deviations from Firm Means by Intervals of Demeaned Instrument

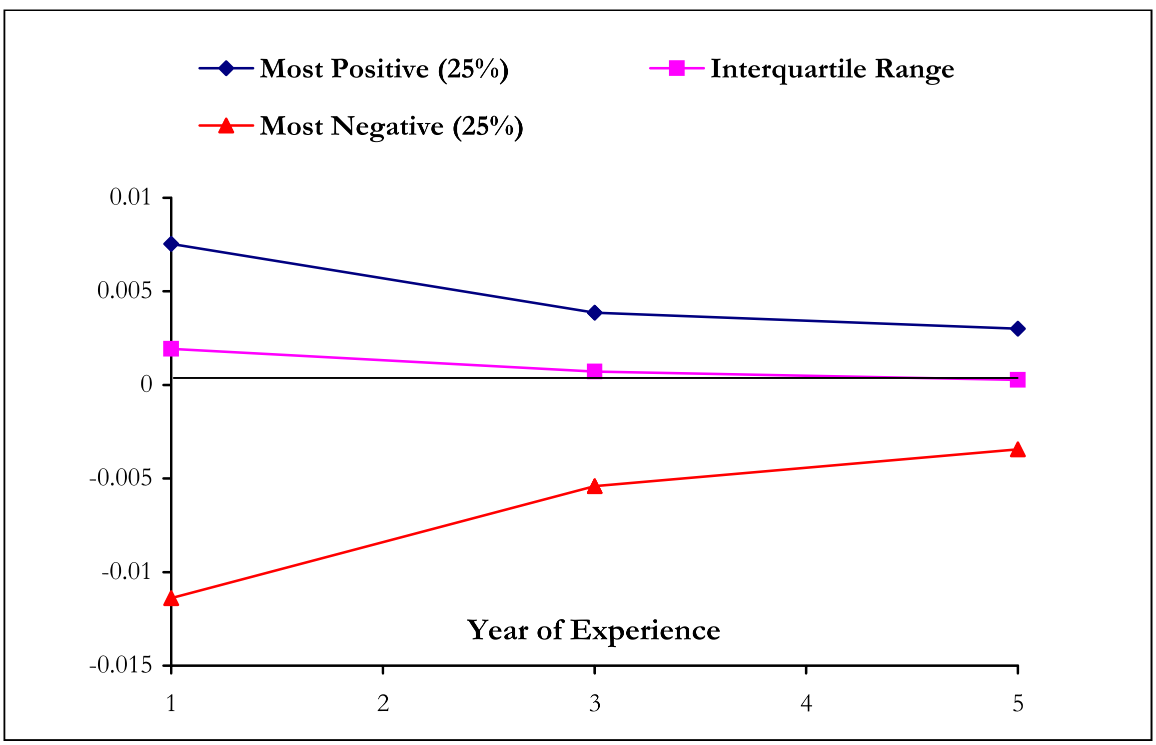
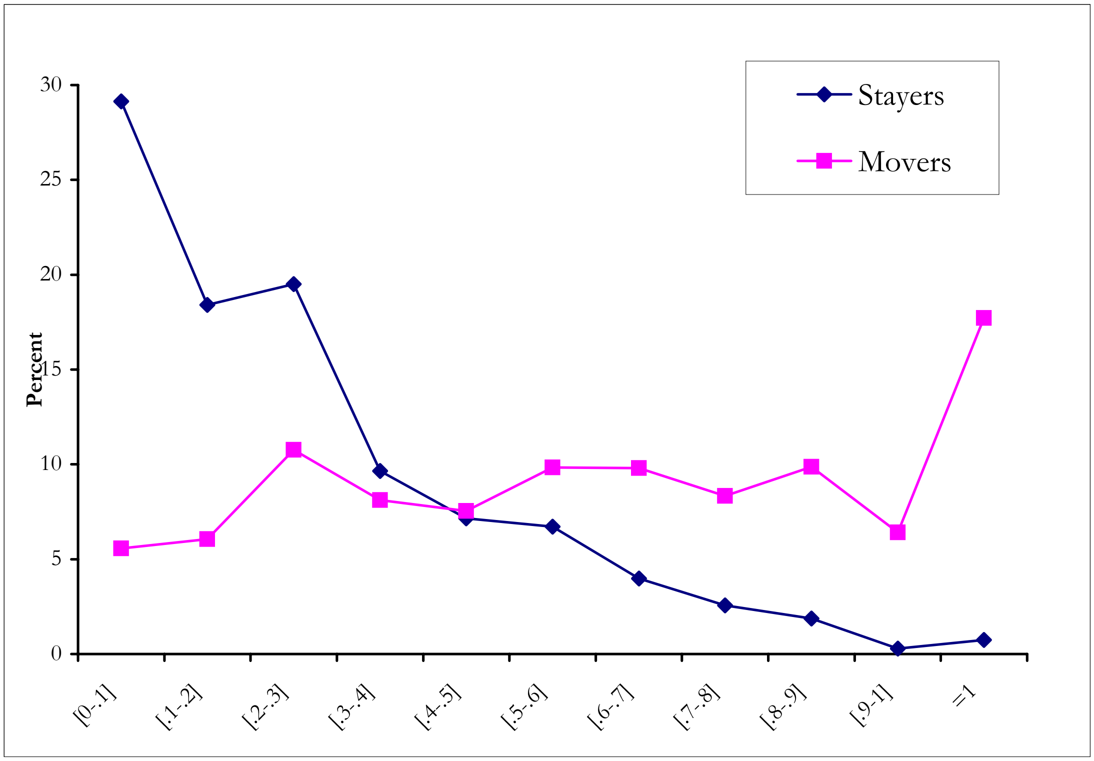


Figure 7: Distribution of Firms' Retention Rates of Young Apprentices At the End of Training

Panel A: Distribution of Fraction 'Other' Movers



Panel B: Distribution of Deviations of Fraction 'Other Movers' From Firm Means

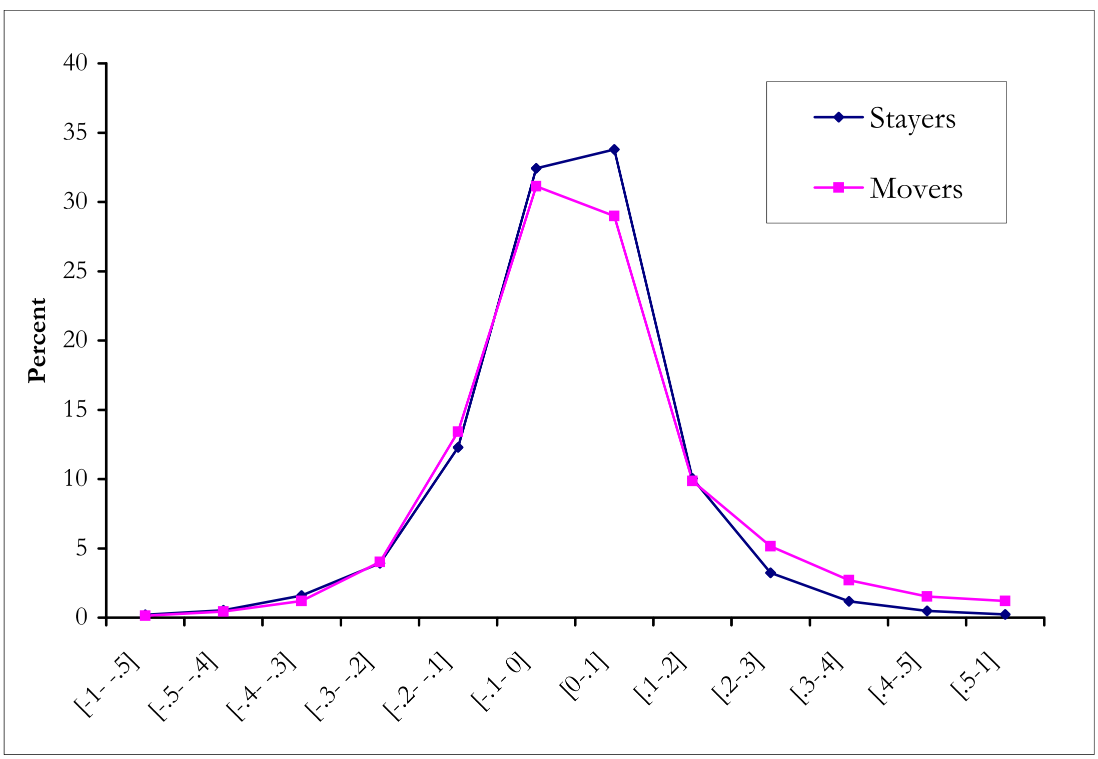


Table 1: Models of Job Mobility and Wage Changes and Their Implications for OLS Estimates of Displacement Effects

Model	Basic Assumptions and Results	Main Empirical Implications	Basic Papers	Implications in Terms of Basic Wage Equation
Adverse Selection	- workers differ in ability, but ex-ante identical to employers - only current employer learns - displacement is signal - random quits sustain market	- OLS biased by negative selection - more exogenous quits, less bias	Greenwald (1986), Gibbons&Katz (1991), Acemoglu&Pischke (1998)	$\mathrm{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) < 0$, $\delta_{It} = \delta_{Vt} = 0$
Initial Assignment	- workers differ in ability - perfect information - firms differ in turnover rates - low ability workers sort into high turnover firms	- OLS biased for levels by initial sorting, not for changes - exogenous displacement has no effect with right comparison group	Salop&Salop (1976), Wang&Weiss (1998)	$\mathrm{cov}(\bar{a}_{j(i)}, D_{i0}) < 0$, $\delta_{It} = \delta_{Vt} = 0$
Job Search	- fixed wage distribution - fixed offer arrival rate, search on the job - wages rise with experience - workers move voluntarily	- OLS upward biased: voluntary movers gain - true temporary negative effect from loss of "search capital"	Burdett (1978), Topel&Ward (1992)	$\delta_{Vt} > 0$, $\mathrm{cov}(V_{i0}, D_{i0}) > 0$, $\delta_{It} < 0$, $\frac{\partial \delta_{It}}{\partial t} < 0$
Sequential Sorting	- as in "Initial Assignment", but information imperfect - employers learn over time - job changers have wage gains and wage losses	- OLS biased in levels and changes - bias from initial sorting - exogenous job loss no effect	Gibbons&Katz (1992), Gibbons&Katz& Lemieux&Parent (2002)	$\mathrm{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) \neq 0$, $\mathrm{cov}(\bar{a}_{j(i)}, D_{i0}) < 0$, $\delta_{It} = \delta_{Vt} = 0$
Institutions/ Rosen	- "Ports of Entry": restricted internal labor market careers - "Stepping Stones": sequential skill growth paths	- OLS shows true negative effect - possibly sorting into and negative selection out of 'career'-firms	Rosen (1972), Doeringer&Piore (1977)	$\delta_{It} < 0 \; \forall t$, $\mathrm{cov}(\bar{a}_{j(i)}, D_{i0}) < 0$, $\mathrm{cov}(a_i - \bar{a}_{j(i)}, D_{i0}) < 0$

Notes: For a discussion of the theories and the notation of the model see Section 2 of text.

Table 2: Young Men In Early Years of Labor Market in US and Germany

		Germany	United States
Panel A: Men, Age 18-34			
Actual Years of Experience at 5 (10) Years of Potential Experience	5 Years	4.30	3.73
	10 Years	8.91	8.19
Average Number of Jobs at 5 (10) Years of Potential Experience	5 Years	2.27	4.56
	10 Years	3.65	6.96
Fraction Leaving Job at 2 (6) Quarters of Job Tenure	2 Quarters	0.15	0.22
	6 Quarters	0.10	0.11
Panel B: Men, Age 18-34, Jobs Lasting 6+ Quarters			
Average Completed Job Duration		7.29	7.00
(Standard Deviation)		(4.13)	(3.67)
Fraction Changing Job		0.25	0.28
(Standard Deviation)		(0.43)	(0.45)
Average Wage Growth on Job		0.09	0.07
(Standard Deviation)		(0.22)	-
Panel C: Controlling for Experience and Tenure			
Average Wage Growth Within Jobs		0.20	0.14
(Standard Error)		(0.00)	(0.01)
Average Wage Growth Between Jobs		0.21	0.20
(Standard Error)		(0.03)	(0.02)
Fraction Growth After 10 Years		80%	60%

Notes: Left Column- own calculations from IAB employment subsample including apprentices (see notes of appendix Table C1). Right Column - Topel & Ward (1992)

Table 3: Sample Characteristics of Main Sample of Apprentices

	All Graduates	Subsample With Valid Wage By Potential Experience	
		1 Year	5 Years
Age at End of Training	20.9	20.9	21.0
Fraction High School	0.16	0.18	0.16
Fraction Male	0.62	0.50	0.62
Fraction German	0.92	0.90	0.93
Training Duration 2-3 Years	0.62	0.64	0.72
Training Duration >3 Years	0.22	0.18	0.13
Training Firm Size 500+	0.57	0.56	0.56
Training Firm Annual Employment Growth	-0.02	-0.02	-0.02
	(0.13)	(0.13)	(0.12)
White Collar Worker	0.46	0.55	0.47
Manufacturing	0.47	0.44	0.48
Services and Trade	0.18	0.19	0.19
Banking, Insurance	0.15	0.17	0.14
Transport, Communications	0.08	0.07	0.07
Fraction Moving at Graduation	0.35	0.30	0.32
Average Fraction Movers Among Other Apprentices	0.37	0.35	0.36
	(0.30)	(0.29)	(0.30)
Log Training Wage	3.91	3.92	3.91
	(0.33)	(0.33)	(0.35)
Log Real Wage	-	4.82	4.94
	-	(0.23)	(0.28)
Cohort 1992	0.41	0.41	0.66
Cohort 1993	0.31	0.31	0.34
Cohort 1994	0.28	0.29	0.00
Number of Observations	218880	142958	106648

Notes: Sample of apprentices who graduated in 1992 to 1994 from establishments with at least 50 employees and at least 5 graduating apprentices. For additional sample restrictions, see text. The first column shows sample statistics for the entire sample of graduating apprentices. The last column shows the same characteristics for apprentices with valid wage observations at exactly 1 and 5 years of potential labor force experience. The only characteristic changing over time is the wage, all other variables pertain to the training period.

Table 4 Characteristics of Workers Leaving (Movers) and Staying (Stayers) at Training Firm At Graduation

	Year of Experience	Workers Staying at Training Firm	Workers Leaving Training Firm
Panel A: Wages In First 5 Years After End of Training			
Average Log Real Daily Wages By Years of Potential Labor Force Experience	**1**	4.85	4.76
		(0.0009)	(0.0015)
	3	4.93	4.83
		(0.0011)	(0.0018)
	5	4.97	4.87
		(0.0013)	(0.0021)
Panel B: Characteristics of Workers and Training Firms			
Age at End of Training		20.9	21.0
Fraction High School		0.18	0.14
Fraction Male		0.61	0.65
Fraction German		0.93	0.89
Training Duration 2-3 Years		0.64	0.57
Training Duration >3 Years		0.21	0.26
Training Firm Size 500+		0.60	0.51
Training Firm Annual Employment Growth		-0.02	-0.03
		(0.11)	(0.16)
White Collar Worker		0.49	0.41
Manufacturing		0.51	0.40
Services and Trade		0.15	0.23
Banking, Insurance		0.17	0.10
Transport, Communications		0.05	0.13
Log Real Training Wage		3.97	3.80
		(0.30)	(0.36)
Average Fraction Movers Among Other Apprentices		0.25	0.60
		(0.22)	(0.31)
Cohort 1992		0.42	0.40
Cohort 1993		0.31	0.31
Cohort 1994		0.27	0.29
Number of Observations		143248	75632

Notes: Standard errors in parentheses in Panel A, standard deviations in parentheses in Panel B. All differences are significant at 5% level.

Table 5: OLS-Estimates of the Effect of Leaving Training Firm at Graduation on Log Real Wages, Various Specifications

	Exp.	(1)	(2)	(3)	(4)	(5)	(6)
Effect of Leaving Training Firm on Wages By Year of Potential Experience	**1**	-0.092	-0.092	-0.079	-0.052	-0.054	-0.050
		(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
	3	-0.095	-0.095	-0.085	-0.062	-0.063	-0.054
		(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
	5	-0.094	-0.094	-0.086	-0.061	-0.061	-0.054
		(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Demographics		-	Yes	Yes	Yes	Yes	Yes
Firm Controls		-	-	Yes	Yes	Yes	Yes
Training Controls		-	-	-	Yes	Yes	Yes
Occupation		-	-	-	-	Yes	Yes
Industry		-	-	-	-	-	Yes
R2		0.06	0.16	0.20	0.25	0.26	0.30
MSE		0.246	0.232	0.227	0.220	0.218	0.213

Notes: Dependent variable is the log real daily wage. All Regression have interaction of experience dummies cohort-dummies. Demographics include age and dummies for German, male, and high school graduate. Firm variables include employment growth and three firm size dummies. Training variables include log real training wage and three dummies for training duration. Each regression has 427,546 observations. Coefficients on the other regressors are shown in Appendix Table E1. All standard errors are clustered at the firm level. Standard errors in parentheses.

Table 6: Different Estimates of Wage Losses of Apprentices Who Leave Training Firm at Graduation - Main Sample

	Year of Exp.	Raw Differences	OLS with Controls	OLS only Firm Fixed Effects	IV without Firm Fixed Effects	IV with Firm Fixed Effects
		(1)	(2)	(3)	(4)	(5)
Effect of Leaving Training Firm on Wages By Year of Potential Experience	**1**	-0.092	-0.054	-0.060	-0.095	-0.113
		(0.0042)	(0.0031)	(0.0028)	(0.0080)	(0.0264)
	3	-0.095	-0.063	-0.061	-0.092	-0.068
		(0.0042)	(0.0033)	(0.0031)	(0.0073)	(0.0219)
	5	-0.094	-0.061	-0.061	-0.085	-0.029
		(0.0046)	(0.0038)	(0.0035)	(0.0090)	(0.0350)
H0: 1=3						
F-Statistic		2.62	28.89	0.001	124.23	7.19
df		11609	11609	11609	11609	11609
p-value		0.105	0.000	0.973	0.000	0.007
H0: 1=5						
F-Statistic		0.39	8.1	0.002	35.67	8.27
df		11609	11609	11609	11609	11609
p-value		0.532	0.004	0.965	0.000	0.004

Note: The dependent variable is the log real daily wage. The first rows report estimates of a dummy for moving out of training firm after end of training interacted with experience-dummies. The last rows report F-test statistics for equality of these coefficients. All specifications include cohort-experience-dummies. The regression models of columns 2, 4, and 5 also include age and dummies for German, male, and high school graduate; training firm employment growth rate, three firm size dummies; log real training wage and three dummies for training duration. Each regression has 427,546 observations. All standard errors are clustered at the establishment level (11609 establishments). Standard errors in parentheses.

Table 7: First Stage Regressions - Linear Models of Probability of Leaving Training Firm at Graduation

		<u>Instrument 1</u>: Fraction 'Other' Graduates Leaving Firm at the End of Training		**<u>Instrument 2</u>: Fraction 'Other' Graduates Leaving Firm with Non-Employment Spells**	
	Year of Exp.	**No Firm Fixed Effects**	**Firm Fixed Effects**	**No Firm Fixed Effects**	**Firm Fixed Effects**
Instrument Interacted With Year of Potential Experience	**1**	0.761	0.159	0.733	0.219
		(0.0060)	(0.0200)	(0.0152)	(0.0183)
	3	0.792	0.187	0.814	0.260
		(0.0060)	(0.0200)	(0.0155)	(0.0179)
	5	0.787	0.148	0.848	0.209
		(0.0070)	(0.0210)	(0.0204)	(0.0209)
R2		0.29	0.34	0.14	0.34
MSE		0.391	0.376	0.431	0.376
Number of Firms		-	11610	-	11610

Notes: Dependent variable is dummy for moving out of training firm at end of training. All Regression as well as experience*cohort and occupation dummies as well age and dummies for German, male, and high school graduate; training firm employment growth rate, three firm size dummies; log real training wage and three dummies for training duration. Each regression has 427,546 observations. Coefficients on other regressors are shown in Appendix Table E2. The reduced form coefficients are shown in Appendix Table E3. All standard errors are clustered at the firm level. Standard errors in parentheses.

Table 8: Different Estimates of Wage Losses of Apprentices Who Leave Training Firm at Graduation - Alternative Instrument Using Workers Who Spent Time in Unemployment

	Year of Exp.	Raw Effect	OLS with Controls	OLS with only Fixed Effects	IV without FE	IV with FE
		(1)	(2)	(3)	(4)	(5)
Effect of Leaving Training Firm on Wages By Year of Potential Experience	1	-0.092	-0.054	-0.060	-0.245	-0.158
		(0.0042)	(0.0031)	(0.0028)	(0.0117)	(0.0275)
	3	-0.095	-0.063	-0.061	-0.215	-0.082
		(0.0042)	(0.0033)	(0.0031)	(0.0095)	(0.0216)
	5	-0.094	-0.061	-0.061	-0.209	-0.007
		(0.0046)	(0.0038)	(0.0035)	(0.0129)	(0.0400)
H0: 1=3						
F-Statistic		2.62	28.89	0.001	22.93	13.18
df		11609	11609	11609	11609	11609
p-value		0.105	0.000	0.973	0.000	0.000
H0: 1=5						
F-Statistic		0.39	8.1	0.002	11.64	13.81
df		11609	11609	11609	11609	11609
p-value		0.532	0.004	0.965	0.001	0.000

Note: The dependent variable is the log real daily wage. The first rows report estimates of a dummy for moving out of training firm after end of training interacted with experience-dummies. The last rows report F-test statistics for equality of these coefficients. All specifications include cohort-experience-dummies. The regression models of columns 2, 4, and 5 also include age and dummies for German, male, and high school graduate; training firm employment growth rate, three firm size dummies; log real training wage and three dummies for training duration. Each regression has 427,546 observations. All standard errors are clustered at the establishment level (11609 establishments). Standard errors in parentheses.

Table 9: Different Estimates of Wage Losses of Apprentices Who Leave Training Firm at Graduation - Panel Sample

	Year of Exp.	Raw Differences	OLS with Controls	OLS only Firm Fixed Effects	IV without Firm Fixed Effects	IV with Firm Fixed Effects
		(1)	(2)	(3)	(4)	(5)
Effect of Leaving Training Firm on Wages By Year of Potential Experience	1	-0.080	-0.045	-0.051	-0.088	-0.163
		(0.0043)	(0.0031)	(0.0028)	(0.0080)	(0.0287)
	3	-0.071	-0.039	-0.042	-0.070	-0.090
		(0.0043)	(0.0031)	(0.0030)	(0.0077)	(0.0276)
	5	-0.064	-0.031	-0.036	-0.058	-0.044
		(0.0047)	(0.0036)	(0.0034)	(0.0093)	(0.0403)
H0: 1=3						
F-Statistic		37.64	16.15	1.322	30.67	21.87
df		11100	11100	11100	11100	11100
p-value		0.000	0.000	1.000	0.000	1.000
H0: 1=5						
F-Statistic		36.89	30.41	1.513	25.78	16.05
df		11100	11100	11100	11100	11100
p-value		0.000	0.000	0.219	0.000	0.000

Note: The dependent variable is the log real daily wage. The first rows report estimates of a dummy for moving out of training firm after end of training interacted with experience-dummies. The last rows report F-test statistics for equality of these coefficients. All specifications include cohort-experience-dummies. The regression models of columns 2, 4, and 5 also include age and dummies for German, male, and high school graduate; training firm employment growth rate, three firm size dummies; log real training wage and three dummies for training duration. Each regression has 274,247 observations. All standard errors are clustered at the establishment level (11100 establishments). Standard errors in parentheses.

Table 10: Different Estimates of Wage Losses of Apprentices Who Leave Training Firm at Graduation - Large Firms

	Year of Exp.	Raw Differences	OLS with Controls	OLS only Firm Fixed Effects	IV without Firm Fixed Effects	IV with Firm Fixed Effects
		(1)	(2)	(3)	(4)	(5)
Effect of Leaving Training Firm on Wages By Year of Potential Experience	**1**	-0.095	-0.080	-0.076	-0.133	-0.093
		(0.0059)	(0.0050)	(0.0044)	(0.0126)	(0.0143)
	3	-0.099	-0.087	-0.076	-0.124	-0.068
		(0.0059)	(0.0054)	(0.0050)	(0.0113)	(0.0132)
	5	-0.095	-0.080	-0.073	-0.114	-0.048
		(0.0069)	(0.0064)	(0.0056)	(0.0142)	(0.0174)
H0: 1=3						
F-Statistic		2.79	6.25	0.000	3.14	8.47
df		3280	3280	3280	3280	3280
p-value		0.095	0.013	0.991	0.077	0.004
H0: 1=5						
F-Statistic		0	0	0.007	4.81	10.43
df		3280	3280	3280	3280	3280
p-value		0.945	0.976	0.936	0.028	0.001

Note: The dependent variable is the log real daily wage. The first rows report estimates of a dummy for moving out of training firm after end of training interacted with experience-dummies. The last rows report F-test statistics for equality of these coefficients. All specifications include cohort-experience-dummies. The regression models of columns 2, 4, and 5 also include age and dummies for German, male, and high school graduate; training firm employment growth rate, three firm size dummies; log real training wage and three dummies for training duration. Each regression has 240,441 observations. All standard errors are clustered at the establishment level (3280 establishments). Standard errors in parentheses.

Table 11: Transition Between Firm Sizes Immediately After Graduation from Apprenticeship Training

		Number or Employees at Employment Firm *of First Job*			
		1-50	**50-100**	**100-500**	**500+**
Number of Employees at Training Firm	**50-100**	33.1	**31.41**	22.37	13.12
	100-500	28.52	29.82	**25.29**	16.37
	500+	22.99	24.95	23.57	**28.49**
	Total	25.71	27.19	24.14	22.96

Notes: The table shows the fraction of movers at each originating bracket of training firm size who transit into different brackets of employment firm sizes. Main sample of apprentices.

Table 12: Summary of Implications from Basic Models of Job and Wage Mobility for Different Estimators of Wage Losses from Leaving Training Firm at Graduation

	OLS	OLSFE	IV1	IV2	IV1FE
(1) Adverse Selection	Negative, Permanent	Negative, Permanent	No Effect	No Effect	No Effect
(2) Initial Assignment	Negative, Permanent	No Effect	More Negative, Permanent	Even More Negative, Permanent	No Effect
(3) Job Search	Negative, Temporary	Negative, Temporary	More Negative, Temporary	Even More Negative, Temporary	More Negative, Temporary
(4) Sequential Sorting	Negative, Permanent	Less Negative, Permanent	Less Negative, Permanent	Even Less Negative, Permanent	No Effect
(5) Institutions/ Rosen	Negative, Permanent	Less Negative, Permanent	Less Negative, Permanent	Less Negative, Permanent	Even Less Negative, Permanent

Notes: Models of job mobility and their implications for different estimators are discussed in Section 2.2 and Section 5.

<u>Estimators:</u>

OLS - OLS estimates of wage loss from leaving training firm at graduation

OLSFE - OLS estimates with training firm fixed effects

IV1 - IV estimates with fraction 'other' apprentices leaving training firm at graduation as instrument

IV2 - IV estimates with 'other' apprentices who leave firm and spent time unemployed as instrument

IV1FE - IV1 with with fixed effects

Table 13: Characteristics of Training Firms with High, Medium and Low Average Retention Rates of Trainees at Graduation

	Low Fraction Movers	Medium Fraction Movers	High Fraction Movers
Range of Fraction Moving at Graduation	**0-20%**	**20-60%**	**60-100%**
Average Fraction Moving at Graduation	0.10	0.37	0.77
Average Training Firm Size	622	584	425
(Median Training Firm Size)	(297)	(263)	(179)
Average Number of Graduates	9	9	9
(Median Number of Graduates)	(6)	(6)	(6)
Average Number of Trainees	33	35	40
(Median Number of Trainees)	(20)	(21)	(22)
Average Log Real Training Wage	3.95	3.88	3.73
Average Training Duration in Years	2.6	2.6	2.6
Average Training Firm Growth	0.001	-0.023	-0.045
Fraction Manufacturing	0.42	0.46	0.37
Fraction Services and Trade	0.19	0.28	0.41
Fraction FIRE	0.21	0.13	0.08
Average Fraction White Collar Workers	0.61	0.57	0.56
Average Job Duration of Hired Trainees	2.9	2.3	1.8
Average Log Real Wage of Hired Trainees			
One Year of Labor Market Experience	4.76	4.73	4.68
Five Years of Labor Market Experience	4.96	4.93	4.89

Notes: Characteristics of firms training apprentices by lowest 25%, inter-quartile range, and highest 25% of average fraction of apprentices who leave the firm at graduation. The number of firms is 11618. The relationships of training wages, firm size, firm employment growth, job duration, regular wages and fraction services is linear in the average fraction moving.

Figure C1: Change in Actual Labor Market Experience by Year of Potential Experience - US vs. Germany

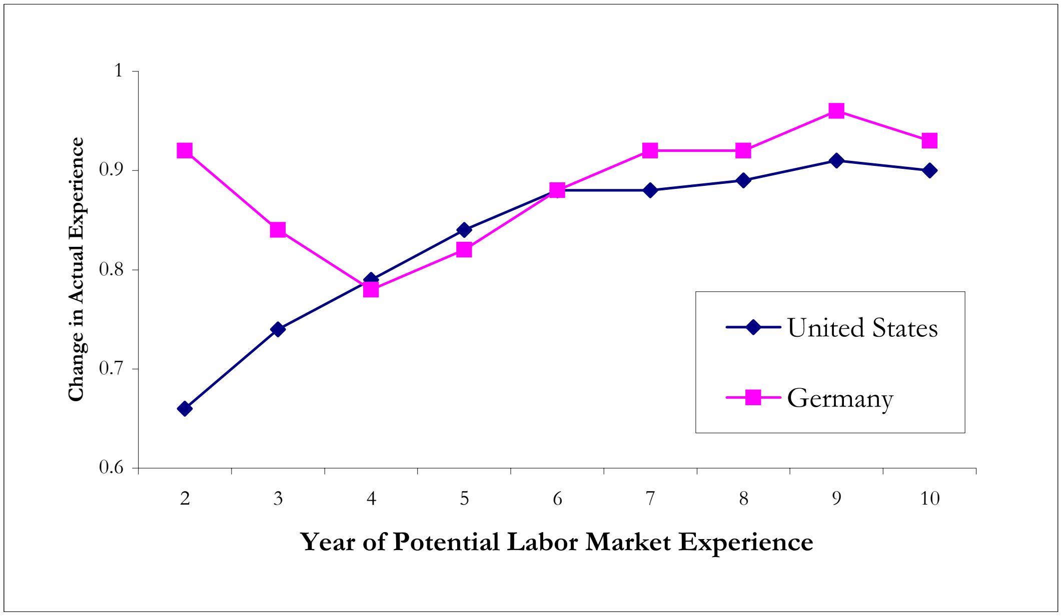
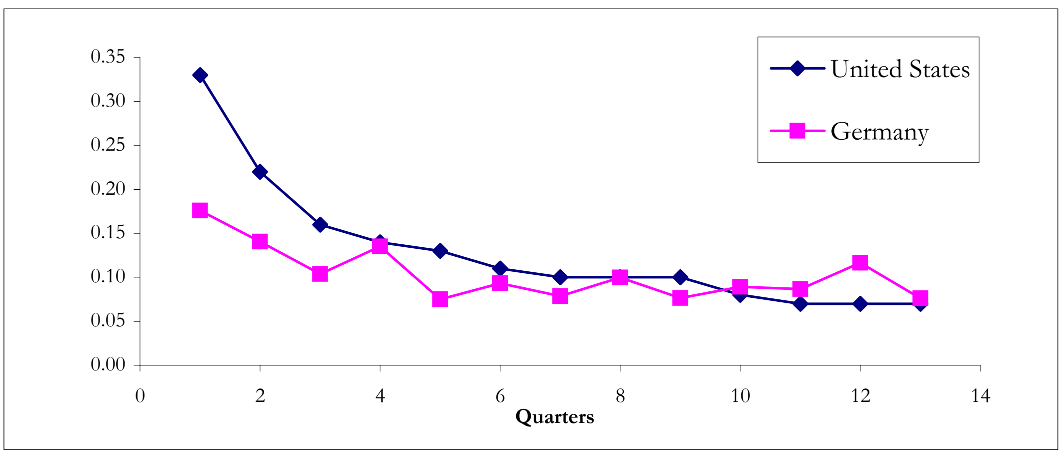


Figure C2: Job Mobility of Young Workers Germany vs. US

Panel A: Empirical Hazard of Job Exit



Panel B: Hazard of Job-To-Job Transitions

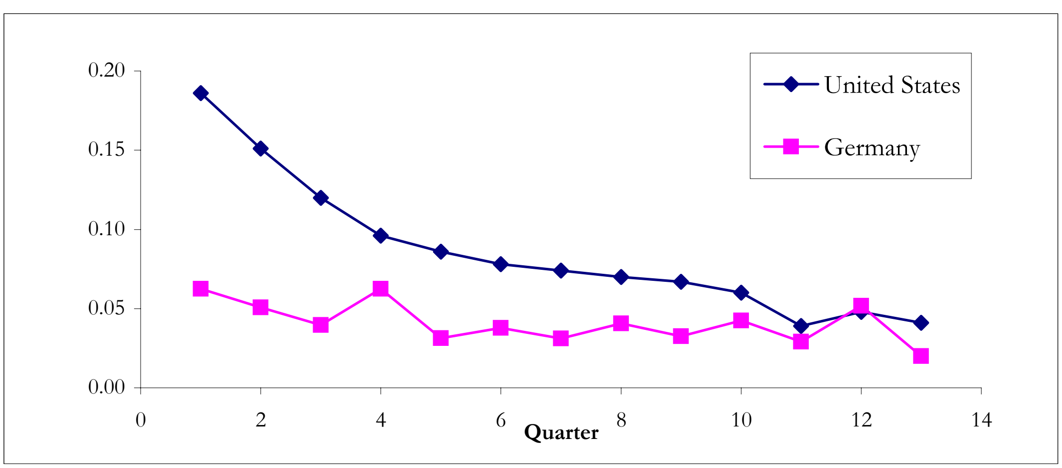
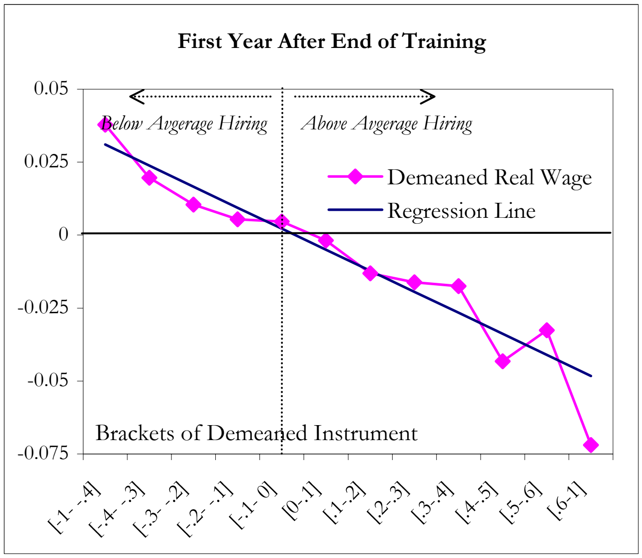
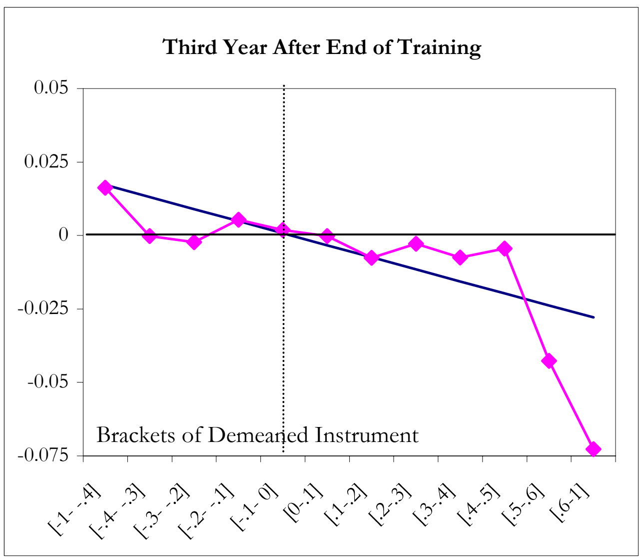


Figure E1: Average Deviation of Wages from Training Firm Means By Brackets of Instrument - Large Firms





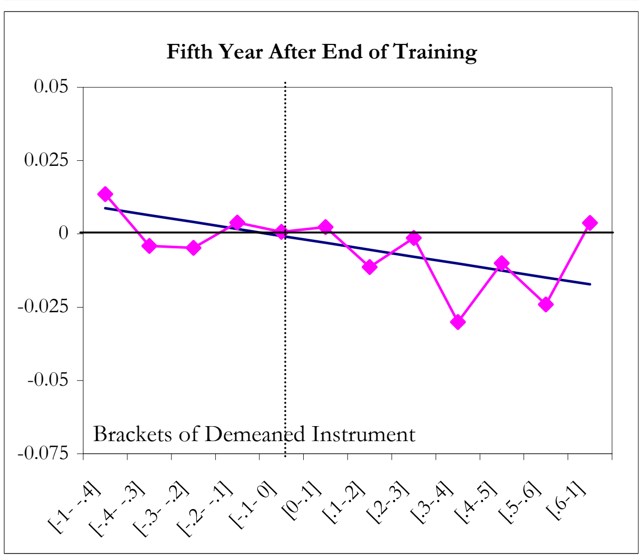


Figure E2: Estimates of Wage Losses from Leavin Training Firm at Graduation with 1- and 2-Digit Industry Fixed Effects

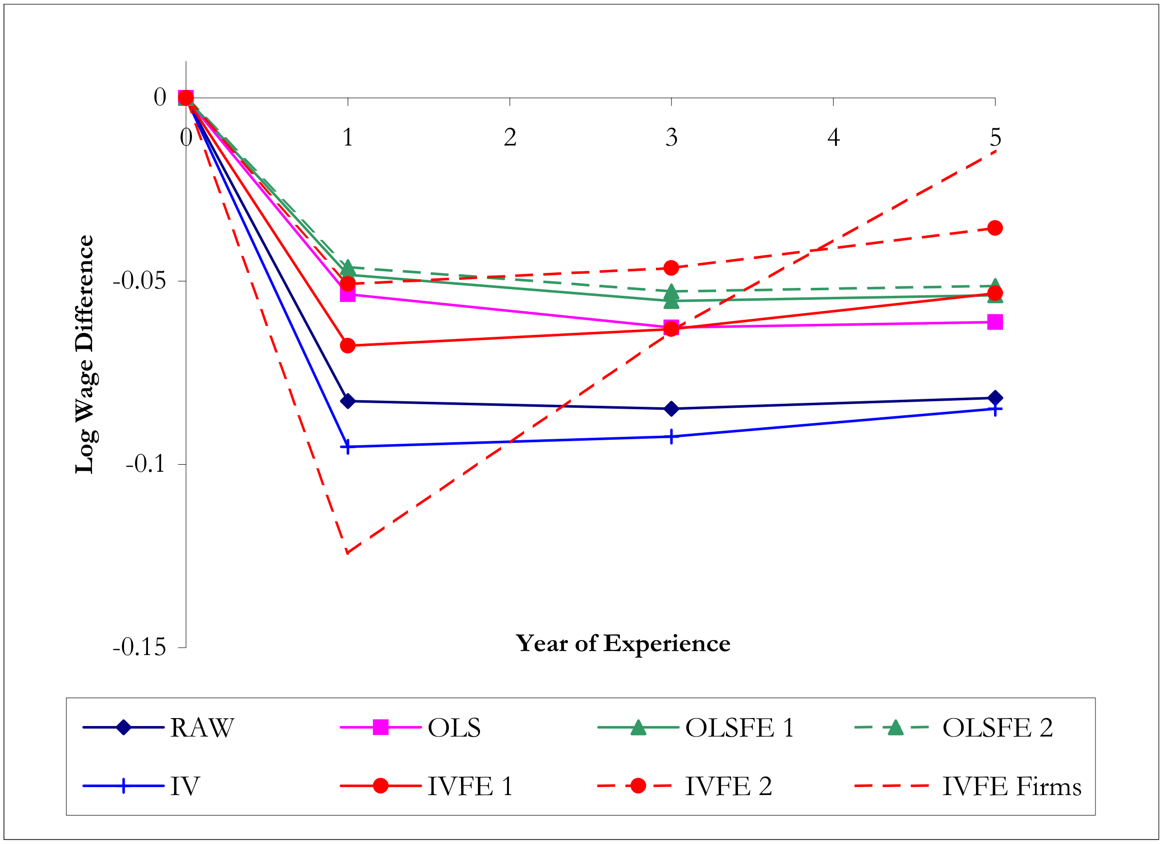


Table A1: Literature on Effects of Early Displacement and Early Job Mobility

Paper	Data	Approach	Results	Notes
Fairlie and Kletzer (2001)	NLSY	Event study of displacement. Fixed effects.	After 5 years, 15% lower real hourly wages for men.	Require 3 years of work prior to displacment.
Gustafson (1998)	NLSY	Event study of displacement. Fixed effects.	After 5 years, 6% lower real hourly wages.	Estimate for both genders.
Ellwood (1988)	NLS	Role of early labor market experience. Fixed effects.	Half a year lost experience: wages 10% lower.	Impacts 4 years later.
Gardecki and Neumark (1997)	NLSY	Effect of indicators of job mobility in first 5 years of career on wages. OLS. All indicators included jointly.	Early conditions not significant.	Impacts 10 years later.
Neumark (1998)	NLSY	Effect of indicators of job mobility in first 5 years of career on wages. IV.	For men, number of early jobs lowers wages -3% to -6%, significant.	Impacts 10 years later. Local labor market state as IV.
Margolis, Vilhuber and Simmonet (2001)	NLSY, DADS, GSOEP	Effect of indicators of job mobility in first 5 years of career on wages. Fixed Effects. All Indicators included jointly.	Early conditions not significant in pooled model.	Impacts 10 years later. USA, France, Germany.
Hashimoto and Miller (1997)	NLSY	Effect of number of early displacements on wages. Tobit 2SLS.	Effect of -6%, significant.	Impacts 10 years later. Estimate for both genders.
Franz, Inkmann, Pohlmeier, and Zimmermann (2000)	BIBB	Effect of early unemployment and failure to complete training on wages. OLS.	Failure -12% significant for men.	Retrospective Data Cross-Sectional Survey.

Notes: This table does claim to be exhaustive. For literature reviews on displaced workers, see Kletzer (1998) or Fallick (1996).

Table C1: Analysis of Within and Between Job Wage Changes - Germany vs. U.S.

Panel A: Regressions of Within Job Log Wage Changes					Panel B: Regressions of Wage Growth Between Jobs		
	Germany		United States			Germany	United States
	Estimate	Average	Estimate	Average		Estimate	
Intercept	0.198	-	0.139	-	**Intercept**	0.2059	0.1992
	(0.0041)	-	(0.0104)	-		(0.0329)	(0.0168)
DX^2	-0.009	14.05	-0.004	13.29	**DX^2**	-0.0057	-0.0013
	(0.0002)	(7.87)	(0.0005)	(6.67)		(0.0004)	(0.0003)
DT^2	0.0014	7.46	-0.0080	6.66	**T(j-1)**	0.0704	-0.0257
	(0.0007)	(6.16)	(0.0017)	(5.49)		(0.0062)	(0.0042)
DT^3	-0.00003	70.52	0.00040	56.15	**T(j-1)^2**	-0.0031	0.0016
	(0.00003)	(111.08)	(0.00010)	(85.93)		(0.0003)	(0.0003)
Completed Tenure	0.0014	7.29	0.0030	7.00	**T(j-1)^3**	0.00004	-0.00002
	(0.0005)	(4.13)	(0.0012)	(3.67)		(0.000005)	(0.000006)
Job Change	0.015	0.25	-0.013	0.28	**D(Comp. Tenure)**	0.0021	0.0028
	(0.0032)	(0.43)	(0.0070)	(0.45)		(0.0009)	(0.0007)
R2	0.1027	-	0.03	-	**R2**	0.121	0.037
Observations	28683	-	6698	-	**Observations**	2702	3367

Notes: Estimates for United States from Topel and Ward (1992), Tables VI and VII. Estimates for Germany from own calculations. Annual changes on quarterly log daily wages on jobs that last at least six quarters. Wages are expressed in 1996 DM and truncated at the maximal amount eligible for social security contributions. Experience and tenure is measured in completed years. To be consistent with Topel and Ward, the sample consists of West-German males. Only workers who were at least 18 years old at entry into the sample and who entered between 1975 and 1980 were kept. Similarly, workers who were older than 34 at the end of the sample period were dropped. Moreover, it was required that workers did not spend more than two years out of the labor force or in unemployment. In addition to the basic restrictions, Panel B also keeps only (1) jobs who lasted more than one quarter and (2) individuals with at least 13 years of potential labor market experience. Conditional on fulfilling these restritions, all estimates for Germany keep apprentices.

Table D1: Sample Statistics Without Restriction on Firm Size, Number of Apprentices, and Labor Force Participation

	All Graduates	Stayers	Movers
Age at End of Training	20.7	20.7	20.7
Fraction High School	0.11	0.12	0.10
Fraction Male	0.54	0.55	0.53
Fraction German	0.92	0.93	0.90
Training Duration 2-3 Years	0.66	0.67	0.64
Training Duration >3 Years	0.18	0.19	0.18
Training Firm Size 500+	0.22	0.23	0.21
Training Firm Annual Employment Growth	0.01	0.01	0.00
	(0.3)	(0.3)	(0.3)
White Collar Worker	0.52	0.53	0.51
Manufacturing	0.37	0.39	0.34
Services and Trade	0.36	0.32	0.40
Banking, Insurance	0.07	0.09	0.05
Transport, Communications	0.04	0.04	0.06
Fraction Moving at Graduation	0.45	-	-
Training Wage	3.76	3.82	3.68
	(0.4)	(0.4)	(0.4)
Cohort 1992	0.41	0.40	0.42
Cohort 1993	0.29	0.31	0.27
Cohort 1994	0.30	0.29	0.32
Number of Observations	815150	451661	363396

Notes: See text. Standard deviations in parentheses.

Table D2: Sample Statistics By Cohort of Graduating Apprentices Applying Main Sample Restrictions

	Year of Graduation		
	1992	1993	1994
Age at End of Training	21.0	20.9	20.8
Fraction High School	0.16	0.16	0.17
Fraction Male	0.62	0.64	0.62
Fraction German	0.93	0.91	0.90
Training Duration 2-3 Years	0.83	0.45	0.48
Training Duration >3 Years	0.00	0.39	0.36
Training Firm Size 500+	0.56	0.57	0.58
Training Firm Annual Employment Growth	0.001	-0.05	-0.03
	(0.11)	(0.14)	(0.14)
White Collar Worker	0.46	0.44	0.48
Manufacturing	0.47	0.48	0.46
Services and Trade	0.19	0.17	0.16
Banking, Insurance	0.13	0.15	0.17
Transport, Communications	0.08	0.08	0.09
Fraction Moving at Graduation	0.34	0.35	0.36
Average Fraction Movers Among Other Apprentices	0.35	0.39	0.37
	(0.30)	(0.30)	(0.32)
Log Training Wage	3.89	3.93	3.91
	(0.37)	(0.32)	(0.29)
Log Real Wage 1	4.83	4.81	4.82
	(0.23)	(0.23)	(0.23)
Log Real Wage 3	4.90	4.89	4.88
	(0.24)	(0.24)	(0.25)
Log Real Wage 5	4.95	4.91	0.00
	(0.28)	(0.29)	(0.00)
Number of Observations	89834	68283	60763

Notes: See text.

Table D3: Basic Sample Characteristics - Large Firms

	All Graduates	With Valid Wage By Potential Experience	
		1	5
Age at End of Training	20.9	20.9	21.0
Fraction High School	0.17	0.19	0.16
Fraction Male	0.67	0.55	0.67
Fraction German	0.91	0.90	0.93
Training Duration 2-3 Years	0.59	0.62	0.71
Training Duration >3 Years	0.26	0.21	0.14
Training Firm Annual Employment Growth	-0.03	-0.02	-0.02
	(0.10)	(0.09)	(0.08)
Fraction Moving at Graduation	0.31	0.26	0.29
Average Fraction Movers Among Other Apprentices	0.33	0.31	0.33
	(0.28)	(0.27)	(0.27)
Log Training Wage	3.96	3.97	3.97
	(0.29)	(0.29)	(0.31)
Log Real Wage	-	4.88	4.99
	-	(0.22)	(0.28)
Cohort 1992	0.40	0.40	0.67
Cohort 1993	0.31	0.31	0.33
Cohort 1994	0.28	0.29	0.00
Number of Observations	124956	80081	59370

Notes: See text.

Table E1: OLS-Coefficients from Wage Regressions, Other Variables, Various Specifications [All Regressors Interacted With Years of Potential Experience]

	Exp.	(1)	(2)	(3)	(4)	(5)	(6)
Constant		4.848	4.548	4.419	3.834	3.689	3.792
		(0.004)	(0.018)	(0.015)	(0.018)	(0.021)	(3.792)
Fraction High-School	**1**	-	0.100	0.091	0.084	0.103	0.063
		-	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
	3	-	0.117	0.111	0.107	0.119	0.086
		-	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	5	-	0.143	0.136	0.129	0.133	0.100
		-	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Fraction Male	**1**	-	0.145	0.131	0.126	0.078	0.072
		-	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
	3	-	0.117	0.105	0.097	0.068	0.061
		-	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
	5	-	0.114	0.103	0.099	0.090	0.082
		-	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Firm Size 100-500	**1**	-	-	0.071	0.047	0.043	0.033
		-	-	(0.005)	(0.004)	(0.004)	(0.004)
	3	-	-	0.063	0.041	0.038	0.027
		-	-	(0.004)	(0.004)	(0.004)	(0.004)
	5	-	-	0.067	0.045	0.044	0.033
		-	-	(0.005)	(0.005)	(0.005)	(0.005)
Firm Size 500+	**1**	-	-	0.170	0.130	0.111	0.090
		-	-	(0.006)	(0.005)	(0.005)	(0.005)
	3	-	-	0.149	0.112	0.099	0.078
		-	-	(0.005)	(0.005)	(0.005)	(0.004)
	5	-	-	0.141	0.106	0.100	0.079
		-	-	(0.006)	(0.006)	(0.006)	(0.006)
Employment Growth of Training Firm	**1**	-	-	0.036	0.021	0.036	0.021
		-	-	(0.012)	(0.009)	(0.009)	(0.008)
	3	-	-	0.000	-0.009	0.004	0.002
		-	-	(0.010)	(0.009)	(0.008)	(0.008)
	5	-	-	0.027	0.009	0.012	0.007
		-	-	(0.011)	(0.009)	(0.009)	(0.009)
Log Real Training Wage	**1**	-	-	-	0.184	0.199	0.161
		-	-	-	(0.004)	(0.004)	(0.004)
	3	-	-	-	0.162	0.171	0.139
		-	-	-	(0.003)	(0.003)	(0.004)
	5	-	-	-	0.150	0.153	0.127
		-	-	-	(0.004)	(0.004)	(0.004)

Notes: Main other regressors included in regressions shown in Table 5. For notes see Table 5.

Table E2: Coefficients on First Stages and Reduced Form - Main Sample

	Exp.	First Stage, No Firm FE	Reduced Form, No Firm FE	First Stage, with Firm FE	Reduced Form, with Firm FE
Constant		0.539	3.695	0.977	4.078
		(0.032)	(0.022)	(0.049)	(0.023)
Fraction High-School	**1**	0.006	0.101	-0.004	0.046
		(0.004)	(0.003)	(0.005)	(0.003)
	3	0.003	0.119	-0.010	0.067
		(0.004)	(0.003)	(0.005)	(0.003)
	5	0.006	0.132	-0.010	0.077
		(0.005)	(0.004)	(0.006)	(0.004)
Fraction Male	**1**	0.003	0.077	0.006	0.053
		(0.003)	(0.002)	(0.003)	(0.002)
	3	0.006	0.067	0.008	0.046
		(0.003)	(0.002)	(0.003)	(0.002)
	5	0.011	0.089	0.010	0.067
		(0.003)	(0.003)	(0.004)	(0.002)
Firm Size 100-500	**1**	-0.134	0.199	-0.187	0.109
		(0.005)	(0.004)	(0.009)	(0.003)
	3	-0.131	0.174	-0.183	0.087
		(0.005)	(0.004)	(0.009)	(0.003)
	5	-0.132	0.157	-0.185	0.077
		(0.005)	(0.004)	(0.010)	(0.004)
Firm Size 500+	**1**	-0.006	0.041	0.022	0.013
		(0.004)	(0.004)	(0.018)	(0.008)
	3	-0.002	0.037	0.024	0.009
		(0.004)	(0.004)	(0.018)	(0.008)
	5	0.008	0.042	0.029	0.012
		(0.005)	(0.005)	(0.019)	(0.009)
Employment Growth of Training Firm	**1**	-0.018	0.108	-0.004	0.019
		(0.005)	(0.005)	(0.025)	(0.010)
	3	-0.009	0.097	0.002	0.010
		(0.004)	(0.005)	(0.025)	(0.010)
	5	0.008	0.098	0.012	0.007
		(0.005)	(0.006)	(0.025)	(0.010)
Log Real Training Wage	**1**	-0.029	0.030	-0.088	0.026
		(0.011)	(0.009)	(0.018)	(0.007)
	3	-0.029	0.001	-0.077	-0.005
		(0.010)	(0.008)	(0.018)	(0.007)
	5	-0.023	0.010	-0.075	-0.012
		(0.011)	(0.009)	(0.019)	(0.009)
R2		0.29	0.26	0.34	0.41
MSE		0.391	0.219	0.376	0.196

Notes: The dependent variable is the log real daily wage in columns 1, 3, and 5, and a dummy for mover in columns 2 and 4. All regressors interacted with potential experience. For further notes see Tables 6 and 7.

Table E3: Reduced Form Regressions of Log Real Wage on Fraction 'Other' Mover and Covariates

	Year of Exp.	Instrument 1: Fraction 'Other' Graduates Leaving Firm at the End of Training		Instrument 2: Fraction 'Other' Graduates Leaving Firm with Non-Employment Spells	
		No Firm Fixed Effects	Firm Fixed Effects	No Firm Fixed Effects	Firm Fixed Effects
Instrument Interacted With Year of Potential Experience	**1**	-0.072	-0.018	-0.180	-0.035
		(0.006)	(0.004)	(0.008)	(0.006)
	3	-0.073	-0.013	-0.175	-0.021
		(0.006)	(0.004)	(0.008)	(0.006)
	5	-0.067	-0.004	-0.177	-0.002
		(0.007)	(0.005)	(0.011)	(0.008)
R2		0.26	0.41	0.26	0.41
MSE		0.219	0.196	0.218	0.196
Number of Firms		-	11610	-	11610

Notes: Dependent variable is the log real daily wage at 1, 3, and 5 years of potential labor market experienc. All Regression have worker, firm, and training variables as well as experience*cohort and occupation dummies. Each regression has 437,546 observations. Table E2 displays coefficients on other regressors. All standard errors are clustered at the firm level. Standard errors in parentheses.

Table E4: Different Estimates of Wage Losses of Apprentices Who Leave Training Firm at Graduation - IV2, Panel Sample

	Year of Exp.	Raw Differences	OLS with Controls	OLS only Firm Fixed Effects	IV without Firm Fixed Effects	IV with Firm Fixed Effects
		(1)	(2)	(3)	(4)	(5)
Effect of Leaving Training Firm on Wages By Year of Potential Experience	1	-0.080	-0.045	-0.050	-0.229	-0.177
		(0.0043)	(0.0031)	(0.0029)	(0.0118)	(0.0143)
	3	-0.071	-0.039	-0.041	-0.195	-0.067
		(0.0043)	(0.0031)	(0.0030)	(0.0109)	(0.0132)
	5	-0.064	-0.031	-0.036	-0.191	0.036
		(0.0047)	(0.0036)	(0.0034)	(0.0157)	(0.0174)
H0: 1=3						
F-Statistic		37.64	16.86	1.595	36.61	36.24
df		11100	11100	11100	11100	11100
p-value		0.000	0.000	0.006	0.000	0.000
H0: 1=5						
F-Statistic		36.89	30.41	1.546	9.74	8.13
df		11100	11100	11100	11100	11100
p-value		0.000	0.000	0.006	0.000	0.000

Note: The dependent variable is the log real daily wage. The first rows report estimates of a dummy for moving out of training firm after end of training interacted with experience-dummies. The last rows report F-test statistics for equality of these coefficients. All specifications include cohort-experience-dummies. The regression models of columns 2, 4, and 5 also include age and dummies for German, male, and high school graduate; training firm employment growth rate, three firm size dummies; log real training wage and three dummies for training duration. Each regression has 274,247 observations. All standard errors are clustered at the establishment level (11100 establishments). Standard errors in parentheses.

Table E5: Different Estimates of Wage Losses of Apprentices Who Leave Training Firm at Graduation - Sample with Restriction on Labor Force Participation

	Year of Exp.	Raw Differences	OLS with Controls	OLS only Firm Fixed Effects	IV without Firm Fixed Effects	IV with Firm Fixed Effects
		(1)	(2)	(3)	(4)	(5)
Effect of Leaving Training Firm on Wages By Year of Potential Experience	**1**	-0.083	-0.045	-0.051	-0.090	-0.147
		(0.0042)	(0.0030)	(0.0026)	(0.0079)	(0.0273)
	3	-0.074	-0.042	-0.042	-0.073	-0.049
		(0.0042)	(0.0031)	(0.0028)	(0.0074)	(0.0216)
	5	-0.068	-0.035	-0.037	-0.062	-0.001
		(0.0047)	(0.0036)	(0.0032)	(0.0091)	(0.0352)
H0: 1=3						
F-Statistic		31.67	5.04	1.234	24.72	25.09
df		11574	11574	11574	11574	11574
p-value		0.000	0.025	0.267	0.000	0.000
H0: 1=5						
F-Statistic		32.1	16.89	1.086	20.63	19.77
df		11574	11574	11574	11574	11574
p-value		0.000	0.000	0.297	0.000	0.000

Note: The dependent variable is the log real daily wage. The first rows report estimates of a dummy for moving out of training firm after end of training interacted with experience-dummies. The last rows report F-test statistics for equality of these coefficients. All specifications include cohort-experience-dummies. The regression models of columns 2, 4, and 5 also include age and dummies for German, male, and high school graduate; training firm employment growth rate, three firm size dummies; log real training wage and three dummies for training duration. Each regression has 392,878 observations. All standard errors are clustered at the establishment level (11574 establishments). Standard errors in parentheses.