Preliminary, please do not cite.

THE IMPACT OF EARLY-CAREER UNEMPLOYMENT ON LONG-TERM LABOR MARKET OUTCOMES — A DISTRIBUTIONAL MICRO-DATA ANALYSIS OF STATE DEPENDENCE

Achim Schmillen[1,2] and Matthias Umkehrer[1,3]

We formally test whether early-career unemployment has a causal impact on long-term labor market outcomes with German administrative matched employer-employee data that allow us to follow more than 300,000 individuals over 25 years. Using an innovative censored quantile instrumental variable estimator and instrumenting early-career unemployment with local labor market conditions at labor market entry, we show that youth unemployment has significant and long-term scarring effects. These effects are especially pronounced in the right tail of the unemployment distribution where an additional day of youth unemployment leads to an increase of prime-age unemployment by up to five days.

Keywords: Scarring, state dependence, censored quantile instrumental variable regressions.

JEL-Classification: J64, J62, C20.

1. INTRODUCTION

Over the past five years youth unemployment has risen considerably in the United States and most European countries. In 2010 the OECD-wide unemployment rate for 15- to 24-year-olds stood at 16.7 percent — that is the highest level within the last 25 years. Figures for some of the larger OECD member countries were even more elevated, with the youth unemployment rate reaching 18.4 percent in the US, 27.8 percent in Italy, 32.9 percent in Greece and a stunning 41.6 percent in Spain [source: OECD (2012)]. These worryingly high rates have stoked fears that "[t]he harm today's youth unemployment is doing will be felt for decades, both by those affected and by society at large" [The Economist (2011, p. 60)].

In order to decide if such fears are justified, we test whether the incidence and/or duration of early-career unemployment leads to a higher amount of unemployment later in life, other things being equal. We find that such *true state dependence* [as defined by Heckman and Borjas (1980)] exists; early-career unemployment has a causal impact on prime-age unemployment.[1] What is more,

[1]Institute for Employment Research (IAB)

[2]Institute for East and Southeast European Studies (IOS)

[3]Correspondence to: Matthias Umkehrer, Institute for Employment Research (IAB), Regensburger Strasse 104, D-90478 Nürnberg, Germany, e-mail: matthias.umkehrer@iab.de, phone: +49 (911) 179-6211, fax: +49 (911) 179-3296.

[1]According to Heckman and Borjas (1980, p. 247), the observation that "the greater the

this "scarring" effect of youth unemployment varies considerably across the (conditional) distribution of prime-age unemployment. In fact, scarring is strongest in the right tail of this distribution. While at the median an additional day of youth unemployment leads to an increase of prime-age unemployment by 2.40 days, for individuals at the 95^{th} percentile another day of early-career unemployment induces 5.37 days of prime-age unemployment *ceteris paribus*. These high numbers imply that the long-term scarring effect of youth unemployment is not only statistically significant but also economically important.

In contrast to most previous work on scarring, we rely on an administrative matched employer-employee data set that contains detailed longitudinal information. With this data set we undertake one of the first truly long-run investigations of scarring. Our analysis is based on the complete employment biographies of all 340,000 men who graduated from Germany's dual education system in 1980. Our data make it possible to identify the exact time and place of labor market entry for these 340,000 individuals and to track them for every day of the first 24 years of their professional career. Compared to a traditional analysis of distinct unemployment spells focusing on durations or Markov transition rates our approach is better able to capture lagged or structural effects of youth unemployment. It also provides more suitable measures of long-term labor market "success" or "failure" than a period-to-period approach.

To the best of our knowledge, ours is the first study of scarring which allows marginal effects to vary across the dependent variable's conditional distribution. A large proportion of individuals in our sample experience no or only short phases of unemployment during their professional career while others suffer from repeated and prolonged periods of joblessness. That is why the explanatory variable's coefficient at the mean may not help in understanding the mechanisms that cause unemployment to have long-term scarring effects. And it may be of little relevance for those individuals with the highest amount of prime-age unemployment.

Moreover, we use the innovative censored quantile instrumental variable estimator introduced by Chernozhukov, Fernández-Val and Kowalski (2011). This estimator not only allows marginal effects to vary over the distribution of prime-age unemployment but also takes account of the fact that more than 60 percent of the individuals in our sample experience no prime-age unemployment at all. Even more importantly, the estimator makes it possible to instrument youth unemployment with local labor market conditions at labor market entry. Therefore, we do more than "just" to show that unemployment is highly persistent amongst a group of individuals: we argue that we really capture a causal relationship.

---

number of previous spells of unemployment and the longer their duration, the more likely is the event that an individual will be unemployed at a point in time" can be explained either because "individuals differ in certain unmeasured variables that influence their probability of experiencing unemployment but that are not influenced by the experience of unemployment" or because "past unemployment (...) alters preferences, prices or constraints that determine, in part, future unemployment". They call the second mechanism true state dependence.

Loosely following Gregg (2001), the local unemployment rate right before graduation from the dual education system is used as an instrument. We argue that the conditions that prevail just before labor market entry influence the quality of initial matching of apprentices to firms and are thus relevant for early-career adjustment processes. At the same time, individuals are on average 17 years old at the beginning of training and we assume choice of location to be exogenous at that age. Therefore, we argue that identification is achieved because — conditional on local labor market conditions at the transition from youth to prime-age and other control variables — the instrument can influence prime-age unemployment only indirectly through scarring effects caused by early-career unemployment.

In terms of related theoretical literature, our research is connected to those models that provide explanations for true state dependence. While Mortensen (1986) stresses the disincentive effects of unemployment insurance, Vishwanath (1989) and Lockwood (1991) quote stigma effects and Pissarides (1992) mentions the decay of human capital.

This study also draws on the empirical literature on scarring. Apart from Gregg (2001), examples include Arulampalam, Booth and Taylor (2000) and Burgess, Propper, Rees and Shearer (2003) for Great Britain, Mroz and Savage (2006) for the United States, Nilsen and Reiso (2011) and Nordström Skans (2011) for Norway and Mühleisen and Zimmermann (1994), Schmelzer (2010) and Niedergesäss (2011) for Germany. The relationship is especially strong with Nilsen and Reiso (2011), Nordström Skans (2011) and Niedergesäss (2011) who also use longitudinal administrative micro data but differ with regard to the identification strategy, the empirical methodology, the scope and the exact focus of the analysis.

Besides, this study is also related to the broader empirical literature on the long-term effects of labor market events or decisions early in the professional career. Prominent studies include von Wachter and Bender (2006) — who demonstrate that displacement leads to persistent wage losses for some groups of young workers while for others losses are substantial but drop to zero within five years — as well as Raaum and Røed (2006) and Oreopoulos, von Wachter and Heisz (2008), who show that business cycle conditions at time of labor market entry have economically significant and long-lasting wage and employment effects.

The remainder of this paper is structured as follows: The next section introduces our data set, the Integrated Employment Biographies. In Section 3, descriptive evidence is presented that shows — amongst other things — that unemployment tends to be very persistent. Sections 4 and 5 contain methods and results of multivariate analyses that strongly suggest that this persistence in unemployment is not (only) due to observable or unobservable differences across individuals but that there really is a causal link between early-career and prime-age unemployment. Section 6 concludes.

2. DATA

Our study relies on the Integrated Employment Biographies (IEB) of the Institute for Employment Research, Nuremberg (IAB). The IEB contain the universe of all individuals who received unemployment benefits and/or were employed subject to social security contributions in Germany at least once between 1975 and 2008. Only employees not covered by social security, like civil servants or family workers, and self-employed persons are not in the data. All in all, the IEB cover about 80% of Germany's total workforce and encompass detailed longitudinal information on employment status, wages, socioeconomic and firm characteristics exact to the day. Because Germany's social security agencies use the underlying administrative data to compute both social security contributions and unemployment benefits, they are highly reliable [cf. Oberschachtsiek, Scioch, Seysen and Heining (2009)].

This study focuses on those individuals that start their employment career after graduating from Germany's dual education system. This system combines apprenticeships in a company and vocational education at a school in one course and is the way through which around 60 percent of young people enter the labor market. Because apprentices have to pay social security contributions, periods in the dual education system are listed in the IEB. Limiting our sample to individuals going through the system allows us to identify socio-demographic and employment-related variables at the time of graduation, i.e. right before the actual labor-market entry.

Our two key variables are *early-career unemployment* — defined as the total length in days of all unemployment spells of an individual in the eight years after finishing the first apprenticeship — and *prime-age unemployment*, the overall length of unemployment spells in the subsequent 16 years. While the latter is our dependent variable, the former is the key regressor. Section 3 will explain the rationale behind dividing the employment career into exactly these two phases.[2]

About 90% of individuals registered as unemployed are eligible for unemployment relief or related benefits. The IEB only contain information on individuals officially registered as job-seeking who do not receive any unemployment benefits from the year 2000 onwards; individuals who for some reason are not registered as unemployed but still willing to take up a job are not covered at all. That is why our benchmark definition of unemployment encompasses exactly those spells of unemployment that are associated with the receipt of benefits.[3]

[2] According to the IEB, 58 percent of the sample entered the labor market on December 31 1980. This seems unlikely and may be an artifact caused by employers that reported changes in employment status only at the end of the calendar year (which was legal in 1980). The actual time of graduation might therefore lie before the one we use. However, our main explanatory variable — the duration of early-career unemployment — is not affected by this issue because unemployment always induces a report by the social security agencies.

[3] This definition might somewhat distort the unemployment pattern of women, a comparatively large number of whom do not qualify for unemployment benefits. This is the main reason why women are not covered by this study.

Using the receipt of unemployment benefits to define unemployment episodes has one important consequence: because regulations concerning unemployment benefits have somewhat varied during the last decades, it is difficult to compare the length of unemployment periods from different points in time. To circumvent this issue and to be sure that results are not driven by cohort effects, we restrict our analysis to one labor market entry cohort. More precisely, we focus on those individuals that finished their first apprenticeship in 1980.[4]

The following variables are included in the multivariate analysis of Section 5 as controls and also because assessing their effects on prime-age unemployment might be interesting in themselves (unless otherwise noted all variables are extracted from the last spell inside the dual education system):

- *Education level.* Because education is known to be closely related to the occurrence of unemployment, we include a dummy variable that measures whether an individual holds a high school diploma.
- *Graduation age.* Because we control for whether an individual finished high school, a positive relationship between graduation age and prime-age unemployment might exist.
- *German nationality.* In particular because of discrimination by employers, Germans might face comparably low prime-age unemployment.
- *Weekly wages.* Elevated wages in the beginning of the professional career could be a sign of high ability and thus be associated with lower prime-age unemployment. At the same time, they could lead to higher reservation wages and ultimately to higher unemployment.
- *Occupation.* Schmillen (2012) and Schmillen and Möller (2012) document long-term unemployment effects of the occupation pursued early in the professional career. We control for the initial occupation with dummy variables for Blossfeld's (1987) twelve occupation categories: agricultural occupations, unskilled manual occupations, skilled manual occupations, technicians, engineers, unskilled services, skilled services, semiprofessions, professions, unskilled commercial occupations, skilled commercial occupations and managers.
- *Region.* Regions are captured by dummy variables for the ten West German federal states (with the state of Schleswig-Holstein as the reference region and omitting Berlin). *A priori* one would expect that working in a state with a rather favorable economic development at the start of the first apprenticeship should mean a comparatively small amount of prime-age unemployment *ceteris paribus.*

[4]In order to ensure valid and undistorted results and to limit the impact of non-standard employment careers, East Germans or those who finished their first apprenticeship at age 27 or later are excluded as are individuals with no IEB records in the eight years after they finished their first apprenticeship and/or the subsequent 16 years. If an apprenticeship lasts for less than a year and is followed by another apprenticeship within three months, we consider the latter to be the individual's first apprenticeship. After all this data cleansing our sample consists of 342,020 men.

- *Sector of the employer.* Dummy variables for ten aggregated sectors are included: energy and mining, manufacturing, construction, trade, transport and communication, financial intermediation, other services, non-profits and households and public administration. The agricultural sector serves as the reference category.
- *Number of direct and indirect changes of employer during the eight years after finishing the first apprenticeship.* Direct changes of employer are defined as changes with an interruption of employment of less than three weeks. If the interruption lasts longer and the worker is not recalled by his/her former employer, then it is counted as an indirect change. We conjecture that individuals with many indirect changes of employer might face higher prime-age unemployment. at the same time, direct changes of employer are likely to reflect voluntary early-career job mobility and should lead to little depreciation of human capital. Thus they could even be associated with less prime-age unemployment [cf. Schmelzer (2010)].
- *Local unemployment at the transition from youth to prime-age (i.e. in 1988).* County-specific unemployment rates are used to capture local labor demand at the transition from youth to prime-age with the appropriate county being determined by the location of the last pre-transition employment spell.

3. DESCRIPTIVE EVIDENCE

3.1. *Labor Market Entry*

At the time of labor market entry, the individuals in our sample are on average 19 years old. The initial apprenticeship lasts on average 770 days while its median duration is 853 days. About 19 percent of the sample population experience at least one more apprenticeship spell until 2004. These subsequent periods in the dual education system tend to be markedly shorter than the initial ones.

After graduation, 60 percent of graduates stay with their training firm. For those who do not stay there, the first employment subject to social security contributions is on average recorded 649 days after graduation. The time between graduation and the first job might not only encompass periods of unemployment and job search but also self-employment, military service or tertiary education. Also, half of those individuals that do not stay with their training firm after graduation enter an employment relationship subject to social security contributions within 73 days.

3.2. *Unemployment and Income over the Professional Career*

Already in the first years of the professional career, unemployment is very unevenly distributed among the individuals in our sample: as the fourth column of Table I shows, the Gini coefficient of total annual unemployment in our sample is 0.90 in 1981. In 1982 it drops to 0.85 and arrives at its minimum value of 0.83 in 1983. Afterwards, the Gini coefficient rises again and reaches 0.92 in 1989 before staying more or less constant.

TABLE I

INEQUALITY AND IMMOBILITY IN THE DISTRIBUTIONS OF ANNUAL UNEMPLOYMENT AND INCOME

		unemployment			income		
year	obs.	total sum (m days)	inequality (Gini coef.)	immobility (Spearman's ρ)	total sum (bn EUR)	inequality (Gini coef.)	immobility (Spearman's ρ)
1981	294,775	4,06	0.9048	0.5182	3,68	0.3197	0.5032
1982	268,229	8,08	0.8526	0.6077	3,91	0.3446	0.6262
1983	264,241	10,07	0.8323	0.5994	4,42	0.2961	0.6923
1984	268,605	9,61	0.8476	0.6103	5,21	0.2565	0.7420
1985	271,451	9,34	0.8580	0.6269	5,73	0.2414	0.7721
1986	280,756	8,23	0.8796	0.6295	6,60	0.2278	0.7868
1987	284,444	7,83	0.8908	0.6437	7,30	0.2230	0.7986
1988	286,323	7,02	0.9046	0.6258	7,88	0.2211	0.8048
1989	286,719	5,48	0.9248	0.6053	8,38	0.2192	0.8020
1990	286,292	4,28	0.9406	0.5977	8,98	0.2163	0.8271
1991	283,912	3,87	0.9457	0.6057	9,18	0.1998	0.8609
1992	281,390	4,18	0.9433	0.6247	9,44	0.2015	0.8671
1993	278,689	5,43	0.9310	0.6861	9,52	0.2124	0.8707
1994	274,922	5,97	0.9259	0.6760	9,25	0.2160	0.8765
1995	273,025	5,61	0.9298	0.6832	9,54	0.2236	0.8841
1996	269,120	6,24	0.9218	0.7190	9,26	0.2171	0.9238
1997	267,045	6,49	0.9190	0.7073	9,11	0.2234	0.9185
1998	266,436	6,03	0.9245	0.7115	9,27	0.2305	0.9145
1999	262,609	5,27	0.9308	0.6912	9,35	0.2257	0.9202
2000	263,220	4,68	0.9385	0.6845	9,54	0.2223	0.9202
2001	261,439	4,86	0.9370	0.6790	9,49	0.2229	0.9220
2002	258,783	5,83	0.9253	0.7273	9,34	0.2275	0.9333
2003	257,825	6,69	0.9151	0.7366	9,49	0.2478	0.9432
2004	255,536	6,69	0.9156	—	9,29	0.2517	—

Notes: All income figures are at constant 2005 prices. The reported distributional statistics on annual income are based on individuals with at least one IEB spell in the respective year.

The Gini coefficient's trajectory can probably be explained by the interplay of two mechanisms: First, at every point in time a high amount of unemployment will tend to be distributed more evenly than a low volume. Second, for any given amount of unemployment, the distribution appears to become more and more uneven over the course of the professional career. The first mechanism does not dominate the second because this would imply that Gini coefficients for years with an equal amount of overall unemployment should be identical. To see that this is not the case, one may compare the Gini coefficients for 1981 and 1991, two years with a roughly equal amount of overall unemployment.

Table I also shows that — at least in the short run — mobility in the distribution of annual unemployment is very low. Its fifth column displays the values from Spearman's rank correlation coefficients between the unemployment distri-

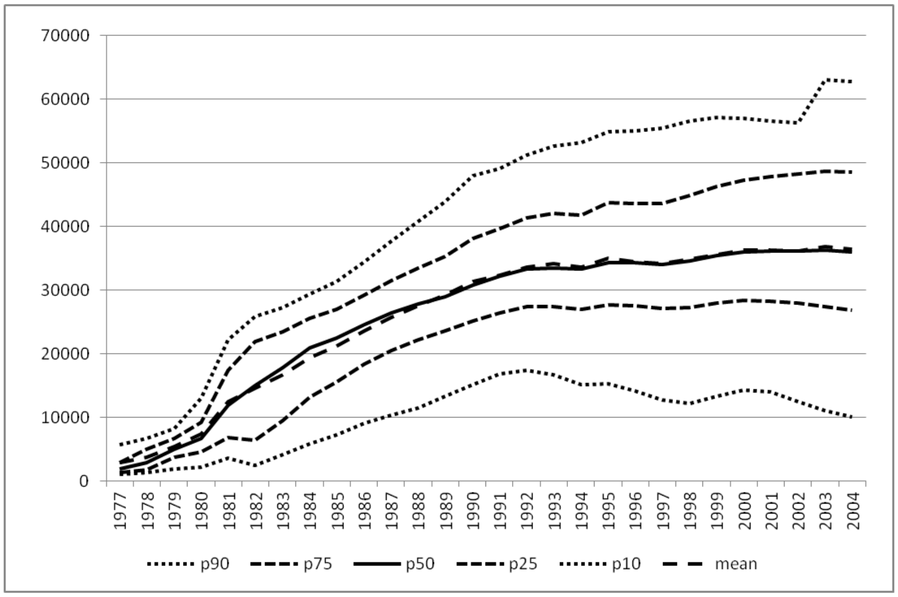


FIGURE 1.— Development of annual income (1980 = employment year 0)

butions of subsequent years (where a higher value indicates a higher immobility in the distribution). As the table makes clear, the correlation coefficients almost monotonously increase from an already high value of 0.52 in 1981 to 0.74 in 2004.

Figure 1 shifts the attention from unemployment to income (defined as labor earnings plus payments form unemployment insurance) by plotting the trajectory of the average total annual income and selected quantiles of the income distribution. It demonstrates that in 1979, the last year before labor market entry, the individuals in our sample on average have an income of 5,373 euros. In the first years on the labor market, their average income increases pretty fast. Already in 1982 it reaches 14,606 euros and then continues to rise to its maximum value of 36,812 euros in 2003, before falling slightly in the last year of the observation period.

As the figure also makes clear, the spread between the 10$^{\text{th}}$ and the 90$^{\text{th}}$ percentile of the income distribution continuously widens from 1977 to 2004. Still, until 1992 annual income rises quite fast over the whole distribution. From 1993 on, income rises much slower and even tends to shrink again for those in its distribution's lower quartile.

Gini coefficients indicate that inequality in annual income is comparably high at the beginning of the professional life (cf. the seventh column of Table I). From its maximum in 1982, this inequality measure markedly drops until 1991 and then starts to rise again. One explanation for this pattern could be that many of those individuals who initially fail to find a productive employer-employee match are able to catch up within the early years of their professional careers. This idea is also supported by the rank correlation between the income distributions of subsequent years: Spearman's rank correlation coefficients start with a value of 0.50 which quickly rises to 0.80 in 1988. From then on the increase is somewhat slower, but the correlation coefficient still almost reaches its maximum possible value of one for the years 2003/2004.

3.3. *Changes of Employer*

Topel and Ward (1992) and others suggest that the early years of the professional career are often characterized by high job mobility. This view is confirmed by Figure 2 which plots annual job mobility rates, defined as the ratio of individuals who experience at least one change of employer to the total number of individuals who are employed for at least one day in any particular year.

As described in Section 2, we distinguish between two forms of job mobility: direct and indirect changes of employer. Figure 2 shows that indirect changes of employer are especially pronounced in the early years of the professional career. The rate of such changes falls from roughly forty percent in 1982 to merely ten percent in 1990. From that year on it remains almost constant. All in all, 60 percent of all indirect changes of employer occur between labor market entry and eight years later. Apart from the strong economic recoveries around 1990 and 2000 the rate of direct changes of employer is generally lower and appears to fluctuate less over time. It does not appear to be particularly high in the early years of the professional career either. Maybe at least partly because the recession of the early 1980s made it hard for individuals to find better jobs, only about 41 percent of all direct changes of employer take place until 1988.[5]

Table II investigates whether unstable employment patterns early in the professional career are related to high amounts of prime-age unemployment. The table reports whether or not the five percent of individuals with the highest amount of prime-age unemployment have an elevated number of changes of employer during the early years of their career.

If the number of direct and indirect changes of employer was independent of the amount of prime-age unemployment one would expect all shares reported in the different columns of Table II (which sorts individuals according to the number of changes of employer during the early years of their professional career) to be close to five percent. However, this only seems to be the case for those with less than four direct changes of employer. Individuals exhibiting four or more direct changes within the first years of their professional career are overrepresented among the group with the highest amount of prime-age unemployment. This suggests that direct changes of employer are not harmful as long as they do not

[5] Over the entire observation period, 15 percent of individuals continually stay with their initial employer. About 75 percent experience at least one direct change of employer.

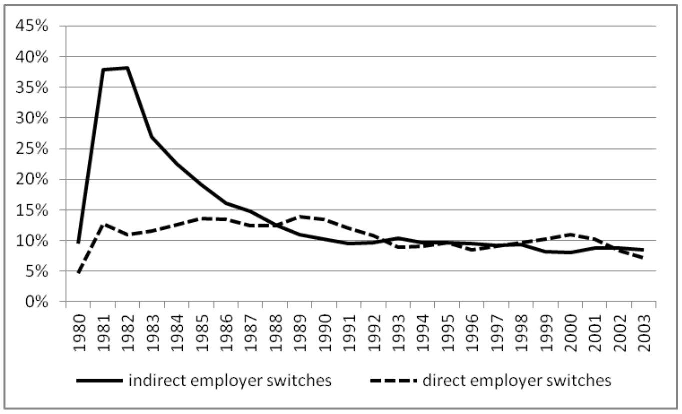


FIGURE 2.— Job mobility rates

happen too often. In contrast, many indirect changes of employer in the years after graduation are clearly an attribute of individuals with a high amount of unemployment later in life. For example, 29 percent of those with at least ten indirect changes of employer are among the five percent of the individuals with the highest amount of prime-age unemployment.

TABLE II

EARLY-CAREER CHANGES OF EMPLOYER AND PRIME-AGE UNEMPLOYMENT

	Share among 5 %	
Number of changes of employer	direct changes	indirect changes
0	5.05	1.52
1	4.68	2.12
2	4.82	4.6
3	4.84	8.38
4	6.16	12.15
5	6.51	16.03
6	9.11	19.74
7	8.64	23.00
8	11.37	24.18
9	13.89	24.89
10 or more	13.09	29.06

Notes: *Share among 5 %* denotes the share of individuals with a certain number of changes of employer during the early career among the five percent with the highest amount of prime-age unemployment.

The descriptive evidence provided so far suggests that an adjustment process takes place early in the professional career. While this phase tends to be characterized by frequent changes of employer (sometimes accompanied by periods of unemployment), subsequent employment periods appear to be more stable and

TABLE III

Summary statistics on early-career, prime-age and lifetime unemployment

	lifetime unemployment	early-career unemployment	prime-age unemployment
mean	480,2393	195,8603	277,453
s.d.	881,4544	350,5933	663,864
min	0	0	0
max	8,949	3,109	5,842
p30	0	0	0
p35	2	0	0
p40	35	0	0
p45	73	0	0
p50	119	22	0
p55	177	51	0
p60	247	86	0
p65	330	127	55
p70	420	182	122
p75	556	253	223
p80	729	339	357
p85	971	445	541
p90	1,381	622	868
p95	2,230	918	1,574
obs.	312,309	338,264	316,065

mobility in the distribution of annual income almost disappears.

3.4. *Early-career, Prime-age and Lifetime Unemployment*

Table III provides summary statistics on early-career, prime-age and lifetime unemployment. It shows that the average individual in our sample suffers from 196 days of unemployment during the first eight years of the professional career and from 277 days of unemployment over the subsequent 16 years. The mean amount of *lifetime unemployment* — defined as the sum of youth unemployment and prime-age unemployment, cf. Schmillen and Möller (2012) — is 480 days. Its distribution is highly skewed to the right: More than 30 percent of individuals in the sample are never registered as unemployed over the entire observation period. At the same time, 20 percent are registered as unemployed for at least three years and five percent for six years or longer.

The distributions of early-career and prime-age unemployment are even more skewed to the right. The median of the distribution of early-career unemployment is 22 days, its 60$^{\text{th}}$ percentile 86 days and its 95$^{\text{th}}$ percentile 918 days. During prime age, almost two thirds of the individuals in the sample experience no unemployment at all.[6]

[6]The Gini coefficients for lifetime, early-career and prime-age unemployment are 0.74, 0.75 and 0.83, respectively. This confirms Bönke, Corneo and Lüthen's (2011) result that annual

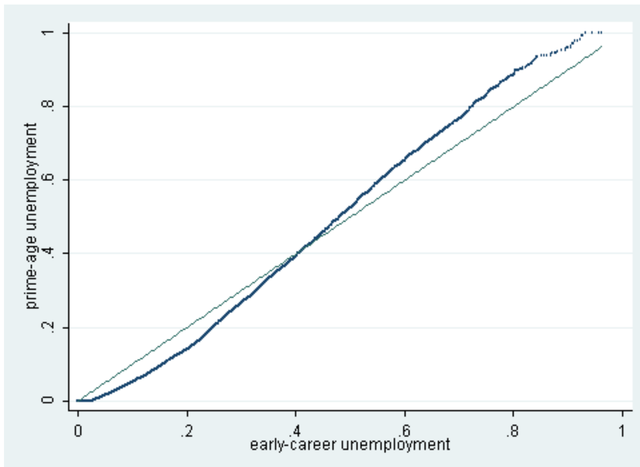


FIGURE 3.— Quantile-quantile plot of early-career vs. prime-age unemployment, measured as proportion of potential time on the labor market.

Figure 3 contains a quantile-quantile plot which helps to compare the probability distributions of early-career and prime-age unemployment by plotting their quantiles against each other. E.g. the smallest amount of early-career unemployment is plotted against the smallest value of prime-age unemployment, the second-smallest against the second-smallest, and so on. Figure 3 shows that comparatively small proportions of unemployment during the early career are plotted against even shorter proportions of prime-age unemployment. At the same time, unemployment proportions higher than 40 percent of the early career — as experienced by less than five percent of the sample — are plotted against even higher proportions of unemployment later in life. Corroborating Table III, the figure shows that the distribution of prime-age unemployment is even more skewed to the right than the distribution of early-career unemployment.

Table IV provides information on the transition probabilities between certain positions in the distributions of early-career and prime-age unemployment. It divides these distributions into cells of equal size (five percentiles) as well as a larger cell that mostly contains individuals with no unemployment in the respective period. If an individual's youth and prime-age unemployment were independent, one would expect roughly five percent of individuals from each early-career unemployment cell to transition into every prime-age unemployment cell (apart from the larger cells containing those with zero unemployment). The table demonstrates this is not what is happening. In contrast, transition probabilities in the table's lower left corner stay below five percent but transition probabilities in

measures of inequality overestimate inequality as compared to measures based on a lifetime perspective.

the table's lower right corner are all much larger. The probability for individuals whose amount of early-career unemployment exceeds the 95th percentile to belong to the ten percent of individuals with the highest amount of prime-age unemployment is almost 50 percent. And almost nobody from this group suffers from no prime-age unemployment at all.

TABLE IV

Transition probabilities between certain positions in the distributions of early-career and prime-age unemployment (in percent).

	early-career unemployment										
prime-age unemployment	*p51 to p55*	*p56 to p60*	*p61 to p65*	*p66 to p70*	*p71 to p75*	*p76 to p80*	*p81 to p85*	*p86 to p90*	*p91 to p95*	*p96 to 1*	*0 to p50*
p61 to p65	5,2	5,6	**6,2**	5,8	**6,2**	6	5,6	5,2	*4,8*	**9**	40,4
p66 to p70	5,4	5,8	6	**6,4**	**6,4**	**6,8**	**6,6**	**6,4**	**6,2**	*4,4*	39,6
p71 to p75	5,2	5,4	**6,2**	**6,4**	**6,6**	**6,6**	**7**	**8,4**	**6,4**	5,6	36,2
p76 to p80	5	5,2	5,8	**6,2**	**6**	**6,6**	**7**	**8,4**	**7**	**6,2**	36,6
p81 to p85	5	5	5,8	**6,4**	**6,4**	**6,8**	**7,6**	**8,2**	**7,6**	**7,6**	33,6
p86 to p90	*4,4*	5	5,6	6	**6,6**	**7,2**	**8,2**	**9,4**	**9,4**	**10,2**	28
p91 to p95	*3,6*	*4,2*	*4,8*	5,4	**6,2**	**7,2**	**8,4**	**10,6**	**11,8**	**15**	22,8
p96 to 1	*2,4*	*2,6*	*3,2*	*4*	*4,8*	5,6	**7,6**	**9,8**	**13,6**	**32,2**	14,2
0 to p60	63,8	61,2	56,4	53,4	50,8	47,2	42	33,6	33,2	9,8	37,5

Notes: Probabilities larger than six percent are marked in bold, those smaller than five percent in italics.

The general picture that emerges from Table IV is that high youth unemployment almost constitutes a necessary condition for having a very elevated amount of prime-age unemployment while those who experience no unemployment during the first years of their professional career often exhibit no prime-age unemployment either. However, there are individuals who manage to transition from a youth characterized by high unemployment to relatively low unemployment levels later in their career or who experience no early-career but a high amount of prime-age unemployment.[7]

To better understand the relationship between early-career and subsequent unemployment, Table V summarizes its temporal dynamics. The table divides our sample into seven groups. The first consists of the 50 percent of individuals with the lowest amount of early-career unemployment. The next four encompass the individuals whose amount of early-career unemployment lies between the 51st and the 60th, the 61st and the 70th, the 71st and the 80th and the 81st and the 90th percentile, respectively. Because of the skewed distribution of early-career unemployment, we divide the individuals between the 91st and the 100th

[7]A rank correlation of 0.38 between early-career and prime-age unemployment confirms a higher long-term mobility in the distributions of unemployment than the short-run mobilities reported above. This supports the hypothesis that early-career unemployment is, to a certain extent, an expression of early job-mobility and does not necessarily have to be damaging in the long-run.

percentile into two groups of equal size.

TABLE V

Relation between early-career unemployment and the occurrence and duration of subsequent unemployment

early-career unemployment	*obs.*	later unemployment	1989 to 1992	1993 to 1996	1997 to 2000	2001 to 2004
0 to p50	169,553	occurrence	0.08	0.10	0.09	0.10
		mean amount	17.86	34.38	33.76	41.61
p51 to p60	33,516	occurrence	0.14	0.16	0.14	0.14
		mean amount	33.41	55.59	53.24	64.09
p61 to p70	33,924	occurrence	0.20	0.19	0.17	0.17
		mean amount	51.29	69.67	71.64	80.90
p71 to p80	33,781	occurrence	0.26	0.24	0.21	0.21
		mean amount	63.06	95.11	97.10	106.45
p81 to p90	33,671	occurrence	0.37	0.31	0.28	0.27
		mean amount	100.88	135.31	136.32	152.24
p91 to p95	16,920	occurrence	0.45	0.36	0.36	0.31
		mean amount	142.00	177.24	183.76	195.18
p96 to 1	16,899	occurrence	0.67	0.51	0.46	0.43
		mean amount	334.87	359.75	360.07	362.28

Notes: *Occurrence* is measured as the proportion of individuals registered as unemployed for at least one day within each time-frame. *Mean amount* denotes the mean total unemployment generated within each time-frame.

Table V lists the proportion of individuals in each of the seven groups that suffer from a positive amount of unemployment in the subsequent 16 years (divided into four periods). Additionally, it displays the mean of total unemployment generated by each group within each time-frame. The table reveals that for each period, the incidence of unemployment strictly increases in early-career unemployment. For those with the lowest amount of youth unemployment, the incidence of unemployment stays almost constant from 1989 to 2004 (at less than ten percent for every four-year period). For all other groups, this incidence falls over time and the decline is strongest for the five percent with the highest amount of early-career unemployment. So at least some members of this group manage to escape the "curse of youth unemployment".[8]

As is evident from Table V, the incidence of unemployment falls over the course of the professional career. At the same time, the mean of total unemployment generated within each time-frame slightly increases with early-career unemployment as well as over time. So with the proportion of people experiencing unemployment declining, a shrinking group of individuals seems to experience longer and longer spells of unemployment. This is in line with true state dependence but is evidence against (time-invariant) heterogeneity as the only link between early and subsequent unemployment. Also, unobserved heterogeneity can prob-

[8]Reassuringly, sample attrition does not seem to depend on early-career unemployment. The share of individuals exiting the sample before the year 2000 is roughly 20 percent for groups one to five and 18 percent for groups six and seven.

ably not explain why the relationship between early-career unemployment and subsequent unemployment weakens over time.

To sum up: The persistent and high inequality in the distribution of annual unemployment in later years documented in this section constitutes a necessary but not sufficient condition for the existence of true state dependence. In fact, evidence against (time-invariant) heterogeneity as the only link between early and subsequent unemployment is provided by the unemployment dynamics summarized in Table V. However, the relatively even distribution of youth unemployment and the frequent changes of employer early in the professional career documented in Section 3.3, strengthen the view that early career unemployment may mostly operate as a "natural" way of building efficient employer-employee matches [cf. Arulampalam, Booth and Taylor (2000) and Footnote 1]. Ultimately, a multivariate analysis that takes account of the potential endogeneity of youth unemployment is needed to decide whether true state dependence exists. This is the goal of the next sections.

4. METHODOLOGY

4.1. *Methodological Challenges*

A regression of prime-age unemployment on early-career unemployment poses at least three methodological challenges. First, as shown in Section 3.4, more than 60 percent of the individuals in our sample are not unemployed for a single day between 1989 and 2004. This is the typical case of a *corner solution outcome* as described by Wooldridge (2002). As a consequence, OLS estimates would be biased and inconsistent because of a correlation between the regressors and the error term.

Second, pure location-shift models based on the mean of the dependent variable's distribution assume marginal effects to be constant over this distribution. In contrast, quantile regression models — proposed by Koenker and Bassett (1978) — not only allow the regressors to alter the location of the dependent variable's distribution but also to impact its shape or scale. This allows an emphasis on the right tail of the (conditional) distribution of prime-age unemployment and a test of whether scarring varies over this distribution. Besides, quantile regression models have further advantages: In particular, they are quite robust with regard to distributional assumptions, outliers or heteroscedasticity.[9]

Third, while we can control for quite a number of socio-demographic or firm-related variables, we cannot rule out the presence of unobserved heterogeneity. Unobserved heterogeneity would be yet another reason for OLS estimates to be

[9] Estimating partial effects of conditional quantiles says nothing about how these effects vary over the distribution of the main explanatory variable, youth unemployment. However, the last section showed that early-career unemployment might (at least to a certain extent) be an expression of early job mobility. In contrast, an individual's position in the distribution of prime-age unemployment is a much better indicator for the overall "success" or "failure" of the professional career.

biased and inconsistent. Relying on fixed-effects techniques would probably not help much because they only control for time-invariant heterogeneity. But the last section made clear that for many individuals in our sample the early years of the professional career provide an opportunity for adjustments and for finding a productive employer-employee match [cf. von Wachter and Bender (2006) for a discussion of this issue]. That is why to us a control function approach in the tradition of Hausman (1978) seems to be a more suitable alternative.

To addresses all three methodological challenges, we use the 4-step censored quantile instrumental variable (CQIV) estimator developed by Chernozhukov, Fernández-Val and Kowalski (2011, p.1) that "combines Powell (1986) censored quantile regression (CQR) to deal semi-parametrically with censoring, with a control variable approach to incorporate endogenous regressors". Now, we will briefly describe the CQIV estimator, followed by a discussion of our identification strategy.

4.2. *Censored Quantile Instrumental Variable Regression*

Assume linearity in parameters and a conditional quantile function of the dependent variable $y^* = Q_{y^*|d,w,v,u}(\tau)$ (prime-age unemployment) at quantile τ that depends on the regressor of interest d (early-career unemployment), a vector of exogenous covariates w (including a constant and possibly the censoring variable), a latent and unobserved variable v which is correlated with y as well as with d and the error term u with a conditional quantile of zero, $\mathrm{Q}_u(\tau|d,w,v) = 0$.[10] Then, with $\tau \in [0,1]$ indexing the quantile and $i = 1,...,N$ indicating the individual, we arrive at the following system of equations:

$$y_i^* = d_i\alpha(\tau) + w_i'\beta(\tau) + v_i\gamma(\tau) + u_i, \tag{4.1}$$

$$d_i = w_i'\dot{\beta} + z_i\pi + v_i, \tag{4.2}$$

where $\alpha(\tau)$, $\beta(\tau)$ and $\gamma(\tau)$ are parameters to be estimated. Further assume conditional independence of u and v, $u \sim U(0,1)|d,w,z,v$ and $v \sim U(0,1)|w,z$. As long as we cannot control for v, estimates of $\alpha(\tau)$ would be biased and inconsistent because v would be absorbed by the new error term "inducing endogeneity or selection bias, so that the conditional quantile of selected [y^*] given the selected [d], is generally not equal to the quantile of potential outcome" [Chernozhukov and Hansen (2006, p.494)].

While we cannot observe v directly, we can estimate it from the residuals of Equation 4.2. To accomplish this, we need to use the "instrumental variable" z that is excluded from Equation 4.1 but influences d through π in Equation 4.2. This instrumental variable enables us to control for any endogenous variation of

[10]Chernozhukov and Hansen (2006) note that neither the hypothetical values of y^* which would evolve under random assignment of treatment nor its corresponding quantiles are actually observable if endogeneity is present. However, CQIV still allows to recover the structural parameters of $Q_{y^*|.}(\tau)$.

d in Equation 4.2 and thus to recover the parameters of interest. This is why v is known as the *control term* and Equation 4.2 as the *control function*.

Here we use labor market conditions at the time of graduation as instruments [cf. the next section]. Therefore, v could be interpreted as the marginal propensity to experience early-career unemployment evaluated at the individual's position in the distribution of prime-age unemployment conditional on the quality of initial matching of apprentices to firms and further exogenous characteristics.

Additionally, we face a corner solution with positive probability mass at zero. That is why we interpret y_i^* as the latent amount of prime-age unemployment as opposed to the actually observed amount of prime-age unemployment, y_i. It holds that

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* \geq 0 \text{ and} \\ 0 & \text{if } y_i^* < 0. \end{cases} \tag{4.3}$$

The conditional quantile function of y is

$$\mathrm{Q}_y(\tau|X) = \max(X'\phi(\tau), 0), \tag{4.4}$$

where $X \equiv [d, w, v]$ and $\phi(\tau) \equiv [\alpha(\tau), \beta(\tau), \gamma(\tau)]$. Equation 4.4 holds because quantiles are equivariant against monotone transformations, such as censoring. In the presence of exogenous regressors, the model presented so far could be consistently estimated with Powell's (1986) estimator. Better applicability is achieved by the semi-parametric estimator developed by Chernozhukov and Hong (2002) which is asymptotically as efficient as Powell's (1986) estimator but far less computationally demanding.

Chernozhukov, Fernández-Val and Kowalski (2011) combine Chernozhukov and Hong's (2002) estimator with a control function approach. The authors show that under mild regularity assumptions, $\sqrt{n}$-consistent and asymptotically normal estimates for $\phi(\tau)$ at every quantile τ can be obtained by

$$\hat{\phi}(\tau) = arg \min_{\phi \in \mathbb{R}^{dim(X)}} \frac{1}{N} \sum_{i=1}^{N} I(\hat{S}_i'\hat{\delta} > k)\rho_\tau(y_i - \hat{X}_i'\phi). \tag{4.5}$$

Here $I(.)$ is an indicator function taking on unity when the expression holds and zero otherwise, $\rho_\tau(u)$ is Koenker and Bassett's (1978) absolute asymmetric loss function, $\hat{X}_i = x(d_i, w_i, \hat{v}_i)$, $\hat{S}_i = s(\hat{X}_i, 0)$ and $x(.)$ as well as $s(.)$ are vectors of transformations of (d, w, v) or $(X, 0)$, respectively. $I(\hat{S}_i'\hat{\delta} > k)$ is called "selector" by Chernozhukov, Fernández-Val and Kowalski (2011) because it selects the subset of observations for which a linear form of the conditional quantile function can be assumed. Unfortunately, linear programming cannot be used to solve Equation 4.5. Instead, one may rely on an algorithm proposed by Chernozhukov, Fernández-Val and Kowalski (2011) which augments the 3-step procedure of Chernozhukov and Hong (2002) by an additional step. The resulting four steps are as follows:

Step 1. Run an OLS regression of d on the instrument z and exogenous regressors w and obtain a prediction for the control term $\hat{v} = \hat{F}_d(d|w, z)$ from the residuals. This allows the construction of $\hat{X}_i = x(d_i, w_i, \hat{v}_i)$.

Step 2. Identify the linear part of the conditional quantile function $X_i'\phi_0(\tau)$. To do so, choose a subset of observations for which the conditional quantile line is "sufficiently" above zero, $\{i : X_i'\phi_0(\tau) > 0\}$. Estimating a logit model for the conditional probability of non-censoring $P(y = 1|S)$,

$$P(y_i = 1|\hat{S}_i) = \Lambda(\hat{S}_i'\delta_0), \tag{4.6}$$

allows to choose a sample $J_0(c)$ that contains those observations which satisfy

$$J_0(c) = \{i : \Lambda(\hat{S}_i'\hat{\delta}_0) > 1 - \tau + c\}, \tag{4.7}$$

with $0 < c < \tau$. Chernozhukov and Hong (2002) suggest to choose c such that $\#J_0(c)/\#J_0(0) = 0.9$.

Step 3. Run an ordinary quantile regression on subsample $J_0(c)$. This gives

$$\hat{\phi}_0(\tau) = arg \min_{\phi \in \mathbb{R}^{dim(X)}} \sum_{i \in J_0(c)} \rho_\tau(y_i - \hat{X}_i'\phi), \tag{4.8}$$

a consistent but inefficient estimate. To gain efficiency, the subset of observations used in Step 2 is updated by choosing $J_1(k)$ according to:

$$J_1(k) = \{i : \hat{X}_i'\hat{\phi}_0(\tau) > k\}, \tag{4.9}$$

where the fitted values from Equation 4.8 are used and the cut-off k plays a similar role as c did in Step 2.

Step 4. Finally, repeat Step 3 but this time on subsample $J_1(k)$.[11]

4.3. *Identification Strategy*

Identifying true state dependence with Equation 4.1 is highly challenging because inter-temporal correlation in unemployment may arise because of — possibly unobserved — factors which vary across individuals, that are not influenced by the experience of unemployment themselves but have an impact on a person's propensity to become and/or stay unemployed [cf. Section 1 and Heckman and Borjas (1980)]. If one does not control for these factors, estimates of scarring effects will be biased and inconsistent.

Important factors correlated with both early-career and prime-age unemployment that are hard to observe may be an individual's productivity level, his or her preferences/norms or the quality of the employer-employee match [cf.

[11] Quantile regressions were calculated using Stata's *qreg* command. For inference, 95% confidence intervals on the coefficients were obtained via bootstrapping the whole 4-step procedure with 100 replications.

Kambourov and Manovskii (2009)]. The IEB allows us to control for socio-demographic and employment-related variables at the time of graduation. However, like probably all other data sets, it does not include perfect measures of individual motivation or ability. This might result in scarring effects estimated solely by Equation 4.1 to be upward-biased. What is more, mobility patterns and matching processes early in the professional career may — if not controlled for appropriately — impose a negative correlation between early-career and prime-age unemployment if early job-mobility prevents subsequent unemployment by establishing more stable employer-employee matches. This in turn might lead to downward-biased estimates of scarring effects. Lastly, differences in preferences or norms might also affect both early-career and prime-age unemployment and might bias estimates of scarring effects in an unknown direction.

According to Heckman and Borjas (1980), what we need for identifying the causal impact of early-career on prime-age unemployment is at least one exogenous determinant of early-career unemployment not correlated with prime-age unemployment. Conditioning on this *instrument* in the control function (Equation 4.2) allows to capture the portion of the variation in early-career unemployment that is independent of relevant but unobservable characteristics.

Loosely following Gregg (2001), our identification strategy relies on instrumenting early-career unemployment with local labor market conditions prevailing at the training firm's location right before graduation.[12] More specifically, the local unemployment rate is used as an instrument, where locations are defined by the administrative districts of Germany's Federal Employment Agency. For 1980, we can distinguish 141 such districts with unemployment rates varying considerably from 1.2 (Nagold) to 7.1 percent (Saarbrücken). We argue that the conditions that prevail just before labor market entry influence the quality of initial matching of apprentices to firms and are thus relevant for the early-career adjustment process described in Section 3.

At the beginning of training, individuals are on average 17 years old. Following the reasoning in Gregg (2001), choice of location can assumed to be exogenous at that age because most individuals still reside with their parents and do not have the means to move to another region. What is more, 97.5 percent of individuals in our sample do not change districts during their apprenticeship. Therefore, we consider their location at graduation to be exogenous.[13]

High job-mobility, on-the-job training (accumulation of specific human capital) and time-varying patterns of economic conditions suggest that the instrument has no direct impact on the duration of prime-age unemployment. But, to avoid

---

[12] Local labor market conditions are captured on June 30th 1980 if the apprenticeship is completed on or after that day and on June 30th 1979 if the graduation happens earlier.

[13] Of the 40 percent of individuals who do not stay at their training firm after graduation [cf. Section 3], 44 percent change districts between graduation and their first job subject to social security contributions. Therefore, the location of the first employment or unemployment spell has to be considered as endogenous. This is why our identification strategy relies on the local labor market conditions right before graduation.

possible correlations between the control term and the error, we follow Gregg (2001) and control for local labor market conditions in 1988. We argue that identification is achieved because, conditional on the local labor market conditions at the transition to prime-age and our other control variables, the instrument can influence prime-age unemployment only indirectly through scarring effects caused by early-career unemployment.

5. RESULTS

5.1. *CQR Results*

As a starting point, Table VI summarizes the outputs of ten censored quantile regressions of prime-age unemployment on early-career unemployment and the control variables introduced in Section 2 (to save space neither the regional dummies nor the constant are included in the table). A negative sign of an explanatory variable's coefficient implies that a larger value of this explanatory variable corresponds to a smaller amount of prime-age unemployment *ceteris paribus*.

In this and the following tables, results are presented for selected quantiles of the conditional distribution of prime-age unemployment. Because a large proportion of sampled individuals exhibit no or little prime-age unemployment and because we are most interested in those that suffer from a very elevated amount of unemployment, our regressions start at the median and proceed in steps of five percentiles all the way to the 95^{th} percentile.

Table VI shows that even if all our control variables are taken into account, a significant and positive relationship between early-career unemployment and prime-age unemployment exists. This relationship is especially pronounced in the right tail of the prime-age unemployment distribution: for individuals at the 95^{th} percentile an additional day of early-career unemployment goes hand in hand with an increase of prime-age unemployment by 1.79 days. The next section will discuss whether this positive relationship between early-career and prime-age unemployment can be interpreted as causal, but for now we turn to a brief discussion of the control variables.

Confirming strong relationships between early-career conditions and later labor market outcomes, almost all control variables exhibit statistically and economically significant coefficients. This is certainly the case for the variables measuring education and age. Interestingly, only for the highest quantiles studied do we find that holding a high school diploma is associated with less and having a higher graduation age with more prime-age unemployment. For other quantiles, coefficients are either insignificant or even exhibit the opposite signs.

Concerning the wage earned at graduation, all coefficients are negative and statistically significant. This result confirms findings by empirical studies focusing on single unemployment spells [like Lüdemann, Wilke and Zhang (2006)].

Also in line with existing research — e.g. by Schmillen and Möller (2012) —

TABLE VI
THIRD STAGE CQR RESULTS

	prime-age unemployment									
	p50	p55	p60	p65	p70	p75	p80	p85	p90	p95
early-career unemployment	0.57***	0.61***	0.67***	0.72***	0.82***	0.96***	1.16***	1.36***	1.59***	1.79***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
age	-13.6***	-12.1***	-9.6***	-7.2***	-3.8***	-1.5***	0.2	2.1***	4.3***	10.4***
	(0.9)	(1.0)	(0.7)	(0.5)	(0.5)	(0.2)	(0.4)	(0.8)	(1.5)	(2.5)
German nationality	-151.7***	-165.6***	-199.6***	-204.2***	-203.2***	-234.5***	-280.9***	-320.5***	-369.6***	-406.6***
	(8.3)	(8.8)	(6.6)	(5.1)	(8.3)	(22.4)	(26.4)	(36.2)	(37.6)	(50.3)
high school diploma	39.8***	26.3***	21.3***	15.0***	15.5***	6.4**	1.5	4.0	-9.6	-47.3***
	(9.6)	(7.8)	(5.4)	(5.4)	(2.8)	(2.6)	(3.4)	(3.7)	(6.4)	(16.6)
wage	-1.01***	-1.03***	-1.66***	-1.25***	-1.21***	-0.87***	-1.42***	-1.7***	-3.1***	-5.3***
	(0.22)	(0.23)	(0.17)	(0.13)	(0.12)	(0.08)	(0.10)	(0.16)	(0.37)	(0.69)
mining/energy	233.8***	335.6***	330.8***	309.2***	247.1***	196.0***	161.2***	226.9***	382.4***	372.9***
	(20.4)	(15.7)	(10.8)	(11.4)	(11.9)	(11.8)	(17.4)	(33.6)	(50.6)	(80.1)
manufacturing	56.2***	88.0***	91.4***	96.0***	71.5***	82.3***	65.7***	110.3***	174.9***	166.7***
	(17.6)	(14.4)	(10.6)	(9.7)	(10.4)	(10.6)	(14.0)	(18.7)	(38.6)	(72.5)
construction	73.9***	105.6***	114.0***	129.6***	111.5***	107.9***	111.6***	142.7***	212.6***	196.3***
	(16.9)	(14.5)	(10.6)	(9.8)	(10.3)	(10.7)	(14.0)	(16.6)	(39.9)	(76.5)
trade	56.4***	87.6***	83.6***	93.7***	74.01***	84.4***	76.7***	125.6***	199.2***	187.0***
	(18.2)	(13.0)	(11.2)	(10.4)	(10.2)	(10.7)	(14.3)	(16.6)	(41.2)	(76.2)
transport/ communication	80.8***	95.5***	80.8***	80.3***	42.9***	34.6***	19.7	53.1***	126.7***	108.8
	(19.1)	(14.1)	(13.6)	(11.5)	(15.0)	(11.1)	(14.7)	(17.8)	(44.4)	(82.2)
financial intermediation	192.7***	160.9***	126.6***	92.5***	79.5***	50.3***	17.5	33.2*	102.7***	-34.1
	(20.1)	(19.3)	(11.6)	(11.5)	(11.4)	(11.1)	(14.2)	(17.4)	(39.6)	(81.5)
other services	122.7***	149.2***	150.8***	152.4***	119.9***	113.2***	111.5***	148.6***	222.1***	204.3***
	(17.3)	(13.9)	(10.2)	(9.7)	(10.1)	(10.3)	(14.1)	(17.9)	(40.9)	(73.3)
non-profit/ households	154.0***	170.2***	148.8***	171.0***	141.1***	153.5***	150.7***	226.0***	333.0***	434.6***
	(23.7)	(13.9)	(17.3)	(16.1)	(16.1)	(16.4)	(23.1)	(38.6)	(69.2)	(113.5)
public administration	11.8	101.8***	101.8***	110.7***	115.4***	59.7***	33.3***	32.8	90.6**	19.3***
	(35.4)	(30.1)	(26.4)	(10.8)	(9.4)	(12.0)	(14.2)	(20.0)	(39.6)	(68.0)
unskilled manual occ.	58.5***	37.7***	29.2**	11.0	10.8	23.0*	63.6***	92.1***	94.2**	196.3***
	(16.8)	(13.2)	(11.4)	(9.0)	(10.1)	(12.4)	(16.6)	(25.8)	(40.1)	(67.1)
skilled manual occ.	-67.7***	-93.4***	-102.1***	-120.2***	-101.3***	-94.0***	-97.2***	-105.3***	-164.3***	-189.6***
	(16.4)	(13.0)	(10.9)	(9.0)	(9.7)	(11.0)	(15.0)	(19.7)	(36.3)	(60.5)
technicians	-155.1***	-175.6***	-161.0***	-157.8***	-141.7***	-118.7***	-127.9***	-153.9***	-220.9***	-298.5***
	(17.7)	(15.8)	(12.4)	(9.7)	(11.0)	(10.3)	(14.6)	(19.9)	(40.0)	(63.6)
engineers	-56.7**	-154.0***	-191.3***	-143.0***	-99.3***	-101.2***	-83.7***	-115.1***	-152.7***	-144.1**
	(25.7)	(26.7)	(19.8)	(17.6)	(17.8)	(11.7)	(20.9)	(24.8)	(48.8)	(66.9)
unskilled services	60.0***	30.9**	17.2	3.7	11.6	25.1	43.8**	70.6***	112.1**	192.3**
	(21.3)	(12.9)	(15.7)	(9.9)	(10.2)	(17.8)	(22.2)	(27.4)	(53.3)	(85.6)
skilled services	-12.4	-51.8***	-79.0***	-90.7***	-74.3***	-55.5***	-64.5***	-58.8***	-51.7	-2.0
	(21.7)	(13.7)	(15.6)	(10.9)	(12.0)	(11.7)	(16.3)	(21.3)	(51.8)	(88.4)
semiprofessions	-80.8***	-106.5***	-97.1***	-119.9***	-98.9***	-101.8***	-105.0***	-144.8***	-222.9***	-273.5***
	(24.9)	(15.7)	(16.3)	(9.8)	(11.1)	(9.9)	(15.3)	(18.1)	(39.1)	(57.8)
professions	-474.1	-425.7**	-524.3***	-528.7***	-256.9*	-213.9**	-214.5***	-292.6***	-384.8***	-425.6***
	(585.9)	(244.7)	(158.6)	(200.8)	(150.8)	(96.0)	(23.1)	(32.3)	(55.4)	(65.2)
unskilled commercial occ.	66.7***	43.8***	35.9***	20.3**	7.5	-2.6	18.9	26.1	18.5	110.9
	(21.9)	(15.4)	(12.8)	(10.2)	(9.6)	(11.5)	(20.1)	(29.5)	(55.6)	(76.3)
skilled commercial occ.	-75.4***	-117.3***	-105.6***	-129.9***	-121.8***	-118.1***	-121.8***	-138.7***	-204.4***	-252.8***
	(18.6)	(13.5)	(10.4)	(10.1)	(10.5)	(11.3)	(14.7)	(19.2)	(37.9)	(63.7)
managers	-111.5**	-129.7**	-124.2**	-129.3***	-132.3***	-124.2***	-130.8***	-173.1***	-325.3***	-475.6***
	(60.6)	(55.3)	(50.9)	(31.8)	(16.6)	(12.8)	(15.9)	(21.4)	(37.1)	(76.4)
direct changes of employer	7.1***	8.9***	11.1***	13.1***	15.2***	12.1***	15.6***	9.4***	3.9**	-9.0**
	(0.6)	(0.6)	(0.5)	(0.5)	(0.5)	(0.6)	(0.6)	(11.1)	(1.7)	(4.1)
indirect changes of employer	88.4***	92.4***	95.2***	94.9***	91.4***	91.9***	98.0***	114.2***	137.0***	179.9***
	(1.4)	(1.2)	(1.0)	(1.4)	(1.5)	(1.6)	(1.6)	(75.7)	(2.6)	(4.3)
unemployment rate (1988)	-0.18	-0.85*	-1.82***	-1.86***	-2.28***	-1.91***	-1.94***	0.12	3.3***	11.5***
	(0.54)	(0.49)	(0.48)	(0.31)	(0.26)	(0.19)	(0.32)	(0.72)	(1.20)	(2.4)

Notes: Bootstrap standard errors in parentheses. *, (**), (***) indicates significance at the 10, (5), (1) per cent level.

we find significant long-term unemployment effects of the occupation pursued early in the professional career. In particular, individuals who are apprentices in a managerial, professional or technical occupation are likely to suffer from comparatively little prime-age unemployment. In contrast, apprentices in an unskilled manual or services occupation tend to fare much worse.

Statistically and economically significant results are found for the sector of the employer, too, and in general these are strongest in the right tail. Starting the professional career in the agricultural sector (the reference category) tends to be associated with the highest amount of prime-age unemployment throughout all quantiles studied while apprentices in the energy and mining sector do best.

As conjectured, individuals with many indirect changes of employer are generally faced with a comparatively high amount of prime-age unemployment. Every indirect change is associated with 88 to 180 additional days of unemployment, depending on the quantile studied. At the same time, the picture for the number of direct changes of employer is more complex: These are associated with less prime-age unemployment for the 95$^{\text{th}}$ percentile of the unemployment distribution. But their coefficients are insignificant or positive for all other quantiles. Also, in absolute values coefficients are far smaller than those for indirect changes of employer.

Lastly, the signs of the coefficients associated with local unemployment rates at the transition from youth to prime age also differ by quantile. If local unemployment rates captured local labor demand one would probably expect positive signs, but such positive signs are only recorded for those individuals with a very elevated amount of prime-age unemployment.

5.2. *CQIV Results*

Outputs from CQIV regressions that address the issue of unobserved heterogeneity are presented in Tables VII and VIII as well as in Figure 4. In the CQIV regression's first stage, individual early-career unemployment is regressed on the county-specific unemployment rate in 1980 (the instrument) and the control variables already used in the CQR estimations summarized in the last section. Later on, prime-age unemployment is the dependent variable while early-career unemployment, a control term generated from the first stage and the same covariates as before serve as regressors (cf. Section 4).

Concerning the first stage, Table VII saves space by not displaying constant and control variables but instead focuses on the coefficient associated with the county-specific unemployment rates in 1980. It shows that this instrument is positively and statistically significantly associated with early-career unemployment.

Table VIII jumps to our CQIV regressions' fourth stage results. These results are also visualized in Figure 4. Once again, constant and control variables are not displayed and once again the focus is on the upper tail of the distribution of prime-age unemployment.

TABLE VII

First stage CQIV regression results

	early-career unemployment
unemployment rate (1980)	18.54***
	(0.79)

Notes: Standard errors in parentheses. *** indicates significance at the 1 per cent level. Constant and control variables not displayed. For a detailed description of variables used, see Section 2.

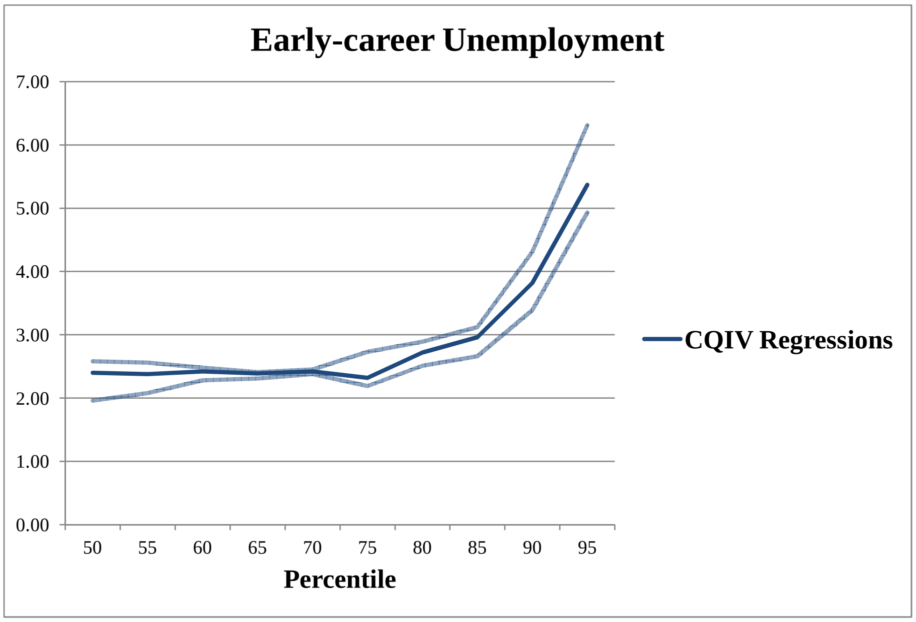
TABLE VIII

Fourth stage CQIV regression results

	prime-age unemployment									
	p50	p55	p60	p65	p70	p75	p80	p85	p90	p95
early-career unemployment	2.40**	2.38**	2.42**	2.39**	2.42**	2.32**	2.72**	2.96**	3.82**	5.37**
	(1.96)	(2.08)	(2.28)	(2.31)	(2.38)	(2.19)	(2.51)	(2.66)	(3.39)	(4.73)
	[2.58]	[2.56]	[2.48]	[2.41]	[2.45]	[2.73]	[2.89]	[3.12]	[4.31]	[6.31]
control term	-1.82**	-1.76**	-1.75**	-1.67**	-1.62**	-1.36**	-1.57**	-1.61**	-2.23**	-3.59**
	(-2.05)	(-2.23)	(-1.84)	(-1.75)	(-1.66)	(-1.77)	(-1.75)	(-1.80)	(-2.71)	(-4.54)
	[-1.36]	[-1.55]	[-1.60]	[-1.56]	[-1.54]	[-1.23]	[-1.37]	[-1.33]	[-1.82]	[-2.95]

Notes: Lower bounds of bias-corrected 95% confidence intervals from 100 bootstrap replications in parentheses, upper bounds in brackets. ** indicates the 95% confidence interval does not include zero. Constant and control variables not displayed. For a detailed description of variables used, see Section 2.

FIGURE 4.— Fourth stage CQIV regression results



Notes: Coefficients from CQIV and bias-corrected 95% confidence intervals from 100 bootstrap replications. For a detailed description of variables used, see Section 2.

Qualitatively, the table confirms the CQR estimations' main result, the existence of a significant and positive relationship between early-career unemployment and prime-age unemployment. Moreover, because of the control variable approach we can now interpret this relationship as causal: more youth unemployment causally leads to a higher amount of prime-age unemployment and thus unemployment early in the professional career has a long-term scarring effect. This scarring effect is present not only at the median but at all of the estimated quantiles. And for all these quantiles it is statistically significant.

What is more, the scarring effect of early-career unemployment varies considerably across the quantiles studied here, in particular across the highest quantiles of the (conditional) prime-age unemployment distribution. In fact, scarring is strongest in the right tail of the distribution. While at the median an additional day of youth unemployment leads to an increase of prime-age unemployment by 2.40 days, for individuals at the 95$^{\text{th}}$ percentile another day of early-career unemployment induces 5.37 days of prime-age unemployment. These pretty high numbers imply that the long-term scarring effect of youth unemployment is not only statistically significant but also economically important.

Besides, for all quantiles studied, coefficients are larger than those found with the help of censored quantile regressions. This means our CQR estimates were downward-biased, a conclusion that is mirrored by the consistently negative control terms in the CQIV regressions' fourth steps. A closer look at these different control terms reveals that the downward bias is most pronounced around the 60^{th}, 90^{th} and 95^{th} percentile of the distribution of prime-age unemployment and smallest for the 75^{th} to the 85^{th} percentile.

6. CONCLUSIONS

This study formally tested whether early-career unemployment has a causal impact on long-term labor market outcomes with German administrative matched employer-employee data that allowed us to follow 300,000 individuals over 25 years. Using an innovative censored quantile instrumental variable estimator introduced by Chernozhukov, Fernández-Val and Kowalski (2011) and instrumenting early-career unemployment with local labor market conditions at labor market entry, it showed that youth unemployment has significant and persistent scarring effects. These effects are especially pronounced in the right tail of the (conditional) distribution of prime-age unemployment. While at the median an additional day of youth unemployment leads to an increase of prime-age unemployment by 2.40 days, for individuals at the 95^{th} percentile another day of early-career unemployment induces 5.37 days of prime-age unemployment.

ACKNOWLEDGEMENTS

We thank Joachim Möller, Stefan Bender and Philipp vom Berge for helpful comments and suggestions. The usual disclaimer applies.

REFERENCES

Arulampalam, Wiji; Booth, Alison and Taylor, Mark (2000). Unemployment Persistence. *Oxford Economic Papers,* **52** 24–50.

Blossfeld, Hans-Peter (1987). Labor-Market Entry and the Sexual Segregation of Careers in the Federal Republic of Germany. *American Journal of Sociology* **93** 89–118.

Bönke, Timm; Corneo, Giacomo and Lüthen, Holger (2011). Lifetime Earnings Inequality in Germany. *IZA Discussion Paper,* **6020**.

Burgess, Simon; Propper, Carol; Rees, Hedley and Shearer, Arran (2003). The Class of 1981: The Effects of Early Career Unemployment on Subsequent Unemployment Experiences. *Labour Economics,* **10** 291–309.

Chernozhukov, Victor and Hansen, Christian (2006). Instrumental Quantile Regression Inference for Structural and Treatment Effect Models. *Journal of Econometrics* **132** 491–525.

Chernozhukov, Victor and Hong, Han (2002). Three-Step Censored Quantile Regression and Extramarital Affairs. *Journal of the American Statistical Association* **97** 872–882.

Chernozhukov, Victor, Fernández-Val, Iván and Kowalski, Amanda (2011). Quantile Regression with Censoring and Endogeneity. *NBER Working Paper 16997.*

The Economist (2011) Left Behind, September 10th, 60–62.

Gregg, Paul (2001). The Impact of Youth Unemployment on Adult Unemployment in the NCDS. *Economic Journal* **111** F626–F653.

Hausman, Jerry (1978). Specification Tests in Econometrics. *Econometrica* **46** 1251–1271.

Heckman, James and Borjas, George (1980). Does Unempoyment Cause Future Unemployment? Definitions, Questions and Answers from a Continous Time Model of Heterogeneity and State Dependence. *Economica,* **47** 247–283.

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

Koenker, Roger and Bassett, Gilbert (1978). Regression Quantiles. *Econometrica* **46** 33–50.

Lockwood, Ben (1991). Information Externalities in the Labour Market and the Duration of Unemployment. *Review of Economic Studies* **58** 733–753.

Lüdemann, Elke; Wilke, Ralf and Zhang, Xuan (2006). Censored Quantile Regressions and the Length of Unemployment Periods in West Germany. *Empirical Economics* **31** 1003–1024.

Mortensen, Dale (1986). Job Search and Labor Market Analysis. In Ashenfelter, O. and Layard, R. (Eds.). *Handbook of Labor Economics, Vol. 2*, Elsevier, Amsterdam.

Mroz, Thomas and Savage, Timothy (2006). The Long-Term Effects of Youth Unemployment. *Journal of Human Resources.* **41** 259–293.

Mühleisen, Martin and Zimmermann, Joachim (1994). New Patterns of Labour Mobility: A Panel Analysis of job Changes and Unemployment. *European Economic Review.* **38** 793–801.

Niedergesss, Markus (2011). Duration Dependence, Lagged Duration Dependence, and Occurrence Dependence in Individual Employment Histories. *Mimeo, University of Tübingen.*

Nilsen, Øivind and Reiso, Katrine(2011). Scarring Effects of Unemployment. *CESifo Working Paper 3675*

Nordström Skans, Oskar (2011). Scarring Effects of the First Labor Market Experience. *IZA Discussion Paper 5565*

Oberschachtsiek, Dirk; Scioch, Patrycja; Seysen, Christian and Heining, Jörg (2009). Integrated Employment Biographies Sample IEBS: Handbook for the IEBS in the 2008 Version. *FDZ Datenreport 03/2009.*

OECD (2012). *Labor Force Statistics by Sex and Age*, retrieved in January 2012, OECD, Paris.

Oreopoulos, Philip; von Wachter, Till and Heisz, Andrew (2008). The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates. *IZA Discussion Paper 3578.*

Pissarides, Christopher (1992). Loss of Skill during Unemployment and the Persistence of Employment Shocks. *Quarterly Journal of Economics* **107** 1371–1391.

Powell, James (1986). Censored Regression Quantiles. *Journal of Econometrics* **32** 143–155.

Raaum, Oddbjørn and Røed, Knut (2006). Do Business Cycle Conditions at the Time of Labor Market Entry Affect Future Employment Prospects? *The Review of Economics and Statistics* **88** 193-210.

Schmelzer, Paul (2010). The Consequences of Job Mobility for Future Earnings in Early Working Life in Germany: Placing Indirect and Direct Job Mobility into Institutional Context. *European Sociological Review* **28** 85–95.

Schmillen, Achim (2012). Long-Term Effects of Occupational Choice on Unemployment: First Evidence from German Registry Data. *Mimeo, Institute for Employment Research.*

Schmillen, Achim and Möller, Joachim (2012). Distribution and Determinants of Lifetime Unemployment. *Labour Economics.* **19** 33–47.

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

Vishwanath, Tara (1989). Job Search, Stigma Effect, and Escape Rate from Unemployment. *Journal of Labor Economics* **7** 487–502.

Wachter, Thill von and Bender, Stefan (2006). In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers' Careers. *The American Economic Review* **96** 1679-1705.

Wooldridge, Jeffrey (2002). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.