Preliminary: please do not cite

DETERMINANTS OF LIFETIME UNEMPLOYMENT - A MICRO DATA ANALYSIS WITH CENSORED QUANTILE REGRESSIONS

ACHIM SCHMILLEN[1,2] AND JOACHIM MÖLLER[3]

Building on a large German administrative micro data set for the time span 1975-2004 we look at lifetime unemployment for selected West German cohorts. Descriptive evidence shows a highly uneven distribution of unemployment in West Germany - more than 60% of the individuals in our sample were not unemployed for a single day over the better part of their professional career while almost half of the total amount of unemployment fell upon 5% of the individuals covered.

We employ censored quantile regressions to explain the total duration of unemployment spells for individuals. Explanatory variables are either characteristics of the individual (like education), of the job (like the wage) or of the employer (like the size of the firm) early in the professional career. A particular emphasis is placed on the importance of the occupation: we find that males working in a disadvantageous occupation at age 25 are ceteris paribus faced with a significantly higher amount of lifetime unemployment. Other factors connected to the amount of men's lifetime unemployment are educational attainment or the wage earned at age 25, amongst others. Some of these variables show very interesting patterns when looking at different quantiles. For women results are in general less clear-cut.

KEYWORDS: Lifetime unemployment, Censored-Quantile Regressions, Occupation-specific human capital
JEL-CLASSIFICATION: J64, J24.

1. INTRODUCTION

Starting with the influential paper of Ljungqvist and Sargent (1998) a growing literature has emphasized the connection between human capital and unemployment. In this literature losing a job is seen as a sudden depreciation of human capital which (possibly together with other factors) might lead to long unemployment spells.

While traditionally the focus was mainly on job- or industry-specific human capital recent studies by Kambourov and Manovskii (2009) and others suggest that it might be more appropriate to consider occupation-specific human capital. Our first contribution is to follow this approach and to evaluate whether the occupation pursued early in the professional career has an effect on the amount of lifetime unemployment for selected West German cohorts. Controlling for a large number of individual or job characteristics we find that men working in a disadvantageous occupation early in their professional career (at age 25) are indeed faced with a significantly higher amount of lifetime unemployment. Using censored-quantile regressions we document that the relationship between the two

[1]Institute for Employment Research (IAB), Osteuropa-Institut Regensburg and University of Regensburg

[2]Correspondence to: Achim Schmillen, Institute for Employment Research (IAB), Regensburger Strasse 104, D-90478 Nürnberg, Germany, e-mail: achim.schmillen@iab.de, phone: +49 (911) 179-5660, fax: +49 (911) 179-3296.

[3]Institute for Employment Research (IAB), IZA and University of Regensburg

variables is especially strong for men with a very high amount of lifetime unemployment. For women our results are less conclusive: over all quantiles examined the advantageousness of the occupation at age 25 has a negative effect on lifetime unemployment. However the coefficient is never statistically significant.

Our second contribution is to show how other attributes affect different quantiles of the distribution function of lifetime unemployment. For men a priori expectations are mostly met: better educated individuals who early in their career hold a well-paying job in a prospering sector are comparatively unlikely to end up with a high amount of lifetime unemployment. For women results are in general less clear-cut. Many variables are either insignificant (like the industrial sector) or do not exhibit expected signs (like some dummy variables covering education).

Our study is related to the literature looking at different factors that might determine the length of unemployment. Examples include Galiani and Hopenhayn (2003) for Argentina, Koenker and Bilias (2001) for the United States, Obben, Englebrecht and Thompson (2002) for New Zealand and Lüdemann, Wilke and Zhang (2006) for Germany. Interestingly all these studies focus on distinct periods of unemployment. In contrast we look at what we call lifetime unemployment: the total length of all unemployment spells over a 25-year period (from age 25 to 50). To the best of our knowledge ours is the first study to look at such a long time span with a rich and reliable administrative micro data set and using multivariate statistics.[1]

The remainder of this paper is structured as follows: section 2 introduces our data set. Section 3 presents descriptive evidence while sections 4 and 5 contain methods and results of our multivariate analyses. Section 6 concludes.

2. DATA

The data set used in our study is the so-called IAB Employment Sample (IABS) of the Institute for Employment Research, Nürnberg (IAB). Its source is the IAB's employment register which covers about 80% of Germany's total workforce. The IABS is based on a 2% random sample of all German employees registered by the social security system and contains detailed longitudinal information exact to the day.

The IABS covers all employment spells associated with the payment of social security contributions. Only employees not covered by social security like civil servants or family workers and self-employed persons are not included in the data. Spells during which workers received unemployment benefits are added to the sample. Because records are used to compute social security contributions and accordingly unemployment benefits data are likely to be highly reliable.

The key variable for our analysis is what we call the individual amount of unemployment or - for the sake of brevity - lifetime unemployment. It is defined

[1] A small number of studies for Germany with similar data exist [e.g. Kurtz and Scherl (2001)]. However these papers are confined to descriptive evidence.

as the total length (in days) of all unemployment spells of an individual from age 25 to age 50. We restrict our sample to this range because of data limitations and because this procedure should limit distorting effects of (un-)employment patterns specific to particularly young or particularly old individuals (e.g. connected to tertiary education or early retirement).

About 90% of those registered as unemployed are eligible for unemployment relief or related benefits. Unfortunately our data do not contain information on unemployed individuals who do not receive any unemployment benefits at all. The same applies to individuals who for some reason have not registered as unemployed but are still willing to take up a job.

Thus we have to restrict our definition of unemployment to spells of unemployment associated with the receipt of benefits. This might somewhat limit the informative value of our analysis. It might especially distort the unemployment pattern of women, a comparatively large number of whom do not qualify for unemployment benefits. This is one reason why we perform our descriptive and multivariate analyses separately for men and women. We are also very careful to compare the respective results.

There is one further drawback of using the receipt of unemployment benefits to define unemployment episodes: regulations concerning unemployment benefits have somewhat varied during the last decades. This makes it difficult to compare the length of unemployment periods from different points in time. Therefore we limit our analysis to a number of selected cohorts. Specifically we focus on those individuals born between 1950 and 1954. Thus our study draws on data from 1975 (when the individuals born 1950 turned 25) to 2004 (when the cohort of 1954 turned 50).

In order to ensure valid and undistorted results and to limit the impact of non-standard employment careers we additionally exclude the following groups from our analyses:

- East Germans because they are only included in our data since the early 1990s.[2]
- Individuals who were employed with coverage by the social security system or recipients of some form of unemployment benefit for the very first time after their 30th birthday.
- Foreigners, i.e. individuals that at the end of their career history did not hold a German passport.

Additionally, it is important to identify a meaningful employment spells. When an individual had multiple jobs at the same time we delete all of these but the one with the highest wage. Also employment spells with the following characteristics are discarded:

- Marginal employed persons who are only covered by our data since 1999.
- Employment spells with a wage below the marginal part-time income thresh-

[2]We label all individuals "East German" whose first employment or unemployment spell registered by the social security system took place in East Germany.

old. We believe that for these employment spells the wage information is corrupt (in fact many of them indicate a daily wage of zero).
- Spells during which the individual was in an apprenticeship because these spells are arguably not comparable to "regular" employment episodes.

While section 3 mainly focuses on descriptive evidence on lifetime unemployment and its interpersonal distribution section 5 contains a multivariate analysis of lifetime unemployment. As well as lifetime unemployment as the dependant variable all explanatory variables are constructed with the help of the IABS data set. Firstly they consist of individual characteristics (like education). Secondly we look at the job held by the individual on her 25th birthday or, if the individual was not employed at this date, at the first job taken up after the 25th birthday. We choose the 25th birthday on the one hand because most people aged 25 have finished education and entered the labor force. On the other hand they are still relatively early in their professional career.

A main aim of our study is to assess whether pursuing a "disadvantageous" occupation early in the career affects the amount of lifetime unemployment. For each employment spell our data set contains a 3-digit occupational variable. All in all this allows us to distinguish 341 different occupations, after some data cleaning (discarding occupations that are covered by our data only for certain years etc.) we are left with 327 occupations for which we have consistent data.

In order to decide whether an occupation is regarded as "disadvantageous" we first of all look at the total number of employment relationships on January 1, 1975 and December 31, 2004 and calculate the relative occurrence of the 327 different occupations for these dates. Next we take the difference between the two relative occurrences. This difference serves as a measure for the relative advantageousness of all occupations contained in our data: a positive difference with a large absolute value corresponds to a steep rise in relative employment for a given occupation while a negative difference means relative employment fell from the beginning of 1975 to the end of 2004.[3]

A number of other variables are included in our multivariate analysis in section 5 as controls and also because assessing their effect on the amount of lifetime unemployment might be interesting in itself:

- **Education** level. It is well-known that education is closely related to the occurrence of unemployment. Since education and occupation are strongly connected as well, controlling for education is of outmost importance. We do this by including five dummy variables that measure whether an individual holds a degree from vocational training but no high school diploma, a high school diploma but no degree from vocational training, a high school diploma and a degree from vocational training, a degree from a technical college or a university degree. Our control group consists of those individuals that hold neither a high school diploma nor a degree from vocational

[3]For an overview of the most common occupations on January 1, 1975 and December 31, 2004 as well as on the most advantageous and disadvantageous occupations see appendix A.

training.
We would expect that individuals with more education and especially those with a tertiary degree (from a technical college or a university) are ceteris paribus faced with a lower amount of lifetime unemployment.[4]

- Weekly **wages** earned at the age of the 25. This variable might be interpreted as a proxy for unobserved individual characteristics and we expect that higher wages ceteris paribus lead to a lower amount of lifetime unemployment.[5]
- **Sector** of the firm for which the individual worked on her 25th birthday. Many occupations are for the most part found in a specific sector of the economy (e.g. bricklayers will almost exclusively work in the construction sector). In order to make sure that we do indeed measure the effect of the disadvantageousness of occupations and not that of sectors we have to control for the latter.
  Our data does include sectoral information on a 3-digit-level. However including hundreds of sector dummy variables in our regressions would lead to unfeasibly time-consuming computations. Therefore we only use dummy variables for six aggregated sectors: agriculture, energy and mining, manufacturing, services, construction as well as the public sector and other activities. A priori it is hard to make statements of the different sectors' roles in determining the amount of lifetime unemployment.
- **Region** where the job pursued at age 25 was based. This variable might once again be construed as a proxy for unobserved personal or firm heterogeneity. It is measured by dummy variables for the 10 West German federal states ("*länder*"). A priori we would assume that working in a well-off state (like Bavaria) at age 25 should ceteris paribus be associated with a comparatively small amount of lifetime unemployment.
- The **size of the establishment** for which the individual worked when turning 25. This might indicate whether a company has (otherwise unobserved) positive or negative characteristics. Since generally speaking in Germany the influence of labor unions is strongest in big companies it might also be a signal for whether employees have some bargaining power that might lead to less lay-offs and a lower risk of unemployment. The size of the establishment is measured by simply adding up the number of its employees. We expect that individuals working for a larger firm at the beginning of their professional career ceteris paribus face a rather small amount of lifetime unemployment.

[4] While some information in our data set (for instances on the duration of employment or unemployment periods and wages) is extremely reliable this is not always the case when it comes to education. We use the imputation mechanism suggested by Fitzenberger, Osikominu and Völter (2006) to obtain reliable education information.

[5] A characteristic of the IABS is that wages are censored to the right. Since wages are not at the center of our interest we ignore this censoring.

TABLE I

Summary statistics on the three states for men

	employed	unemployed	out of the labor force	total
average number of days in state	5416.25	377.27	3702.88	9496.40
average number of spells	3.15	1.67	4.02	8.85
average length of spells in days	1718.05	225.29	920.37	1072.99
relative occurrence of state	.570	.040	.390	1
relative occurrence of state (not considering *out of the labor force*)	.935	.065		1

TABLE II

Summary statistics on the three states for women

	employed	unemployed	out of the labor force	total
average number of days in state	4555.54	243.18	4697.27	9496.00
average number of spells	3.25	1.07	3.70	8.03
average length of spells in days	1400.04	226.67	1269.14	1182.88
relative occurrence of state	.480	.026	.495	1
relative occurrence of state (not considering *out of the labor force*)	.949	.051		1

3. DESCRIPTIVE EVIDENCE

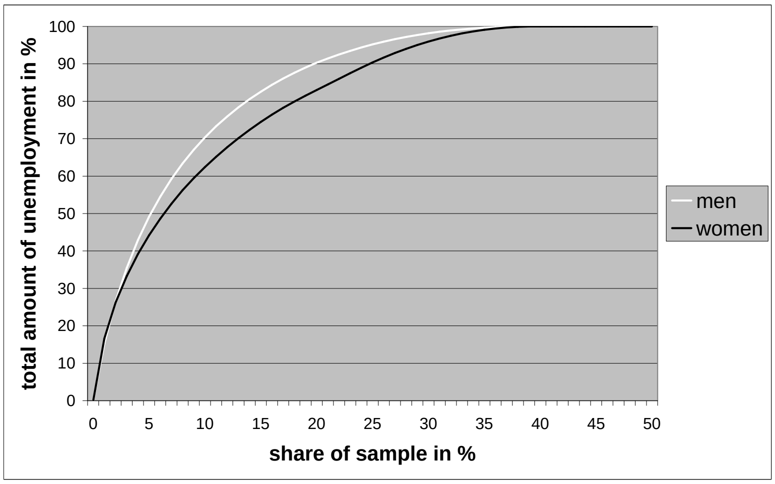
Before turning to our multivariate analyses we now present some descriptive evidence on the amount of personal unemployment and its distribution in West Germany. For this section our samples consist of 33403 men and 28668 women with the characteristics described in section 2. For the regressions in section 5 this numbers reduce to 30334 men and 26036 women for whom we have information on all regressors.

We start by looking at some summary statistics. For this purpose we distinguish three states: *employed*, *unemployed* and *out of the labor force*. The first two states are defined as described in the last section. The remainder of the professional career is labeled *out of the labor force* even though strictly speaking it might encompass episodes of marginal employment, unemployment without receipt of unemployment benefits and similar cases discussed in section 2.[6]

Tables I and II summarize information on the three states. On average employment careers of men encompass 1.67 unemployment spells with an average length of 225.29 days. Women are on average 1.07 times unemployed with average unemployment episodes lasting for 226.67 days. For both genders the state *out of the labor force* plays on average a much greater role than unemployment. Men are on average counted as out of the labor force for more than a third of the

[6]If for an individual information on employment or unemployment is only available some time after her 25th birthday or not right until her 50th birthday these gaps are also included in our notion of *out of the labor force*. Excluding them altogether would not qualitatively alter applicable results.

FIGURE 1.— Inverted Lorenz curves for the interpersonal distributions of the total amount of unemployment for men and women



time span covered (3702.88 of 9496.40 days[7]). The average woman even spends more days out of the labor force than in employment (on average 4697.27 days out of 9496.00 are spent out of the labor force but only 4555.54 in employment)

The large periods reported as out of the labor force can probably be explained not only by actual periods of inactivity but also by relative our restrictive definitions of employment and unemployment. Not counting periods where individuals where out of the labor force we calculate average unemployment rates of 6.5% for men and 5.1% for women. While it is not feasible to compare these figures with unemployment rates defined in a standard way they lie in a plausible range.

We now drop the categories *employed* and *out of the labor force* and focus solely on periods of unemployment. Here, we are most interested to learn more on the long-term interpersonal distribution of unemployment. Our first step is to look at the fraction of the sample that was ever unemployed between age 25 and age 50. We find that "only" about 38% of men and 39% of women were unemployed for at least a single day during this period of life. Conversely more than 60% of the individuals in our sample were not personally affected by unemployment between age 25 and 50 at all. This observation is a first indicator for a very uneven distribution of lifetime unemployment.

The very uneven distribution of lifetime unemployment becomes even more obvious when taking the overall distribution of unemployment into account. This is done by figure 1. The graph draws inverted Lorenz curves for the interpersonal

[7]While for all individuals we look at the time span from their 25th to their 50th birthday leap years have the effect that the total time span differs by up to two days for different cohorts.

distribution of the *total amount of unemployment* separately for men and women. The total amount of employment is defined as the sum of the individual amounts of unemployment for all individuals in our sample. Figure 1 shows two very uneven distributions. This is also the result when looking at Gini-coefficients. Values of 0.842 for men and 0.807 for women signify a very uneven distribution of lifetime unemployment.

For illustrative purpose one can also compare the fact that more than 60% of the individuals in our sample were not affected by unemployment between the age of 25 and 50 with the observation that for men about half of the total amount of unemployment falls upon 5% of the sample. For women 6% of the sample are affected by roughly 50% of the total amount of unemployment.[8]

All in all our data show a very uneven distribution of the total amount of unemployment. This leads to the following question: what determines the individual amount of unemployment? More specifically, it is especially relevant to know which attributes characterize those (say 5%) of individuals who are faced with very high lifetime unemployment. A method particularly suited to address this issue, (censored) quantile regression, is presented in the next section. Subsequently, results of its application to the interpersonal distribution of lifetime unemployment are discussed.

4. METHODS

For a multivariate analysis of the amount of lifetime unemployment it is important to recall that for both men and women more than 60% of individuals in our sample were not unemployed between age 25 and age 50 at all. That means our data are censored and an ordinary least square estimation would lead to biased results. The classical way to deal with censoring would be to use what is called a Tobit estimator [proposed by Tobin (1958)]. Instead we use a more modern alternative with many benefits, the so-called censored quantile regression (CQR) model introduced by Powell (1986).

Compared to a Tobit model CQR offers several advantages: Firstly, as shown by Powell (1986), it does not require homoscedasticity of the error terms. Secondly it is consistent and asymptotically normal whatever the distribution of the error term as long as the conditional quantile of the error term is zero. Thirdly, like the conventional quantile regression model introduced by Koenker and Bassett (1978), it allows marginal effects to differ between lower and higher conditional quantile. This third point is especially relevant in the context of our study since we primarily want to find out whether occupation-specific human capital acquired early in the professional career and other factors are relevant

[8] One might infer from graphs "A" and "B" that the total amount of unemployment is more unevenly distributed for men than for women. However, as was discussed in section 2, such a comparison is highly problematic. The total amount of unemployment for women could in particular be less evenly distributed than shown by the graph if a comparatively large number of women faced with high unemployment are not in fact registered as unemployed.

for individuals with a very high amount of lifetime unemployment.

In general, the CQR estimator for quantile θ assumes the following latent model:

$$y_i^* = x_i'\beta_\theta + \epsilon_{\theta i} \tag{4.1}$$

where x_i is the vector of explanatory variables and $\epsilon_{\theta i}$ denotes the error term with a conditional quantile of zero, $Quant_\theta(\epsilon_{\theta i}|x_i) = 0$. y_i^* is the latent dependent variable.

When estimating the amount of lifetime unemployment we are faced with lower censoring at zero and no upper censoring.[9] In this case the following equation holds between the latent unemployment variable y_i^* and the observable amount of lifetime unemployment y_i:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* \geq 0 \text{ and} \\ 0 & \text{if } y_i^* < 0 \end{cases} \tag{4.2}$$

If lower censoring at zero is present the conditional quantile of y is given by

$$Quant_\theta(y|x) = \max(0, x'\beta_\theta) \tag{4.3}$$

Powell (1986) showed that a consistent estimator for β_θ is obtained as a solution to minimizing

$$\frac{1}{N}\sum_{i=1}^{N}[[\theta - I(y_i < \max(0, x_i'\beta_\theta))][y_i - \max(0, x_i'\beta_\theta)]] \tag{4.4}$$

with respect to β_θ, where I is an indicator function that takes the value of unity when the expression holds and zero otherwise.

In Koenker and Bassett (1978)'s traditional quantile regression models linear programming is used to solve for the regression parameters. Because $\max(0, x_i'\beta_\theta)$ is not linear in β this is not possible for equation (4.4). Fortunately the literature suggests a number of ways to deal with this problem. The most prominent solutions are a an iterative linear programming algorithm proposed by Buchinsky (1994) and a programming algorithm by Fitzenberger (1997). These approaches are, however, not without drawbacks: Fitzenberger (1997) and others point especially to the failure to reach asymptotic efficiency in practice, a high computational burden and a poor performance when a large proportion of the data is censored (as is the case with our application).

Therefore, we make use of improved estimators for censored quantile regressions introduced by Chernozhukov and Hong (2002) that overcomes many of the

[9] Some of the studies on single unemployment episodes mentioned in section 1 face not lower censoring but upper censoring. Koenker and Bilias (2001) and Lüdemann, Wilke and Zhang (2006) use CQR to approach this problem.

shortcomings of the more tested approaches. (Chernozhukov and Hong, 2002, p. 872) report that their "estimators are theoretically attractive (i.e., asymptotically as efficient as the celebrated Powells (...) estimator). At the same time, they are conceptually simple and have trivial computational expenses." In spite of these evident advantages they have not been widely used in the labor literature. Exceptions include Machado and Santos Silva (2008) and Ludsteck and Haupt (2007) who extend the method to censored panel data regressions

The estimating procedure introduced by Chernozhukov and Hong (2002) consists of three steps. Next we briefly describe these steps and how we adjusted it to our specific circumstances.

Step 1. Our first goal is to to chose a subset of observations where the quantile line $x_i'\beta_\theta$ is above the censoring point. We start with a logit estimation of the model

$$\delta_i = \dot{x}_i'\gamma + \epsilon_{\gamma i} \tag{4.5}$$

where δ_i is an indicator of not-censoring and $\dot{x}_i$ is a transform of x_i. It is crucial that censoring is predicted as good as possible. Therefore we include a large number of explanatory variables in $\dot{x}_i$: a cubic polynomial in wage and establishment size, the advantageousness of the occupation, education and professional status dummies, three-digit occupation dummies as well as dummies for 326 West German administrative districts ("Kreise").

Next we select the sample

$$J_0(c) = \{i : \dot{x}_i'\hat{\gamma} > 1 - \theta + c\} \tag{4.6}$$

with c strictly between 0 and θ. We choose c such that $\#J_0(c)/\#J_0(0) = 0.9$. According to Chernozhukov and Hong (2002) this somewhat ad-hoc rule works well in simulations.

Step 2. Now we obtain an initial estimator $\hat{\beta}_\theta^0$ by running an ordinary quantile regression

$$y_i = x_i'\beta_\theta^0 + \epsilon_{\theta i}^0 \tag{4.7}$$

on the sample J_0. Chernozhukov and Hong (2002) show that the resulting estimator is consistent and useful for building up the efficiency of the last step. For step 3 we calculate a sample with the properties

$$J_1(k) = \{i : x_i'\hat{\beta}_\theta^0 > 0 + k\} \tag{4.8}$$

where k plays a similar role as c did in step 2. Much of the literature sets $k = 0$. We follow this approach but also make sure (as suggested by Gustavsen, Jolliffe and Rickertsen (2008)) that $\#J_1/\#J_0 > 0.66$ and $\#\{J_0 \not\subset J_1\}/\#J_1 < 0.1$ in order to avoid using too small a sample and ensure robustness.

Step 3. Finally we run another ordinary quantile regression

$$y_i = x_i'\beta_\theta^1 + \epsilon_{\theta i}^1 \tag{4.9}$$

using observations from J_1 this time. Chernozhukov and Hong (2002) show that the resulting estimator $\hat{\beta}^1_\theta$ not only and works well in their simulations but is also is consistent and asymptotically efficient.

The next section summarizes our benchmark regressions. Quantile regressions where calculated with Stata and its *qreg*/*sqreg* commands. Because *qreg*'s analytical standard errors have frequently been criticized (e.g. by Koenker and Hallock (2001)) for the quantile regressions in step 3 we relied on bootstrap standard errors with 200 replications obtained with the command *sqreg*.

5. RESULTS

Results of our benchmark regressions for men and women are summarized in tables III and IV, respectively. Additionally, results are visualized in figures 2 and 3. We focus on higher ends of the distribution of the amount of personal unemployment because we are most interested in finding the factors that are associated with a very high amount of lifetime unemployment. Specifically we estimate CQR models for the 75th, 80th, 85th, 90th and 95th percentile.

The dependent variable of all our regressions is the amount of lifetime unemployment (measured in days). That means a negative sign of an explanatory variable's coefficient implies this variable is ceteris paribus associated with a smaller amount of lifetime unemployment.

Since one focus of our analysis is to assess whether pursuing an advantageous occupation early in the professional career does lead to a significantly lower amount of lifetime unemployment we discuss results concerning the advantageousness of the occupation in detail. Results on other variables are presented somewhat briefer. In the next subsection we look at the CQRs for men and then turn to the results for women.

5.1. *Men*

Table III and figure 2 show that for men the **advantageousness of the occupation** held early in the professional career is clearly related to the amount of lifetime unemployment. The more advantageous the occupation held on the 25th birthday the smaller the expected amount of lifetime unemployment. This relationship is always statistically significant at the 5% level and especially pronounced for higher quantiles of the distribution function of the amount of lifetime unemployment.

This result lends support to the hypothesis that occupation-specific human capital plays a role in determining success or failure of a professional career. Our regressions control for a number of other factors (region, education, ...). Therefore we can be pretty confident that there is indeed a positive connection between the advantageousness of the occupation and the amount of lifetime unemployment and that the significant coefficient of the advantageousness of the occupation held at the age 25 is not caused by a hidden variable.

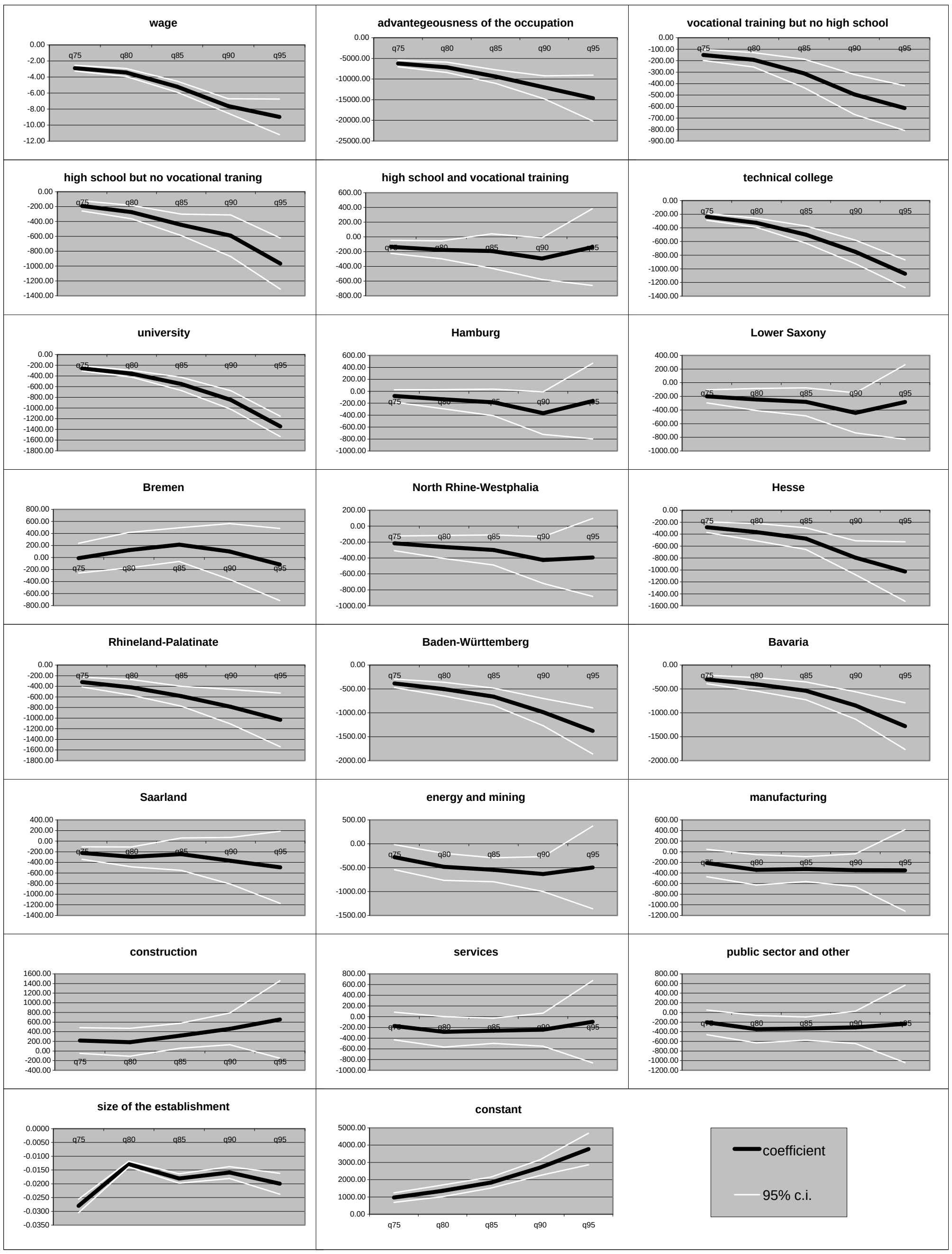
TABLE III

Censored quantile regressions for men, dependant variable: lifetime unemployment in days

variable	75th percentile	80th percentile	85th percentile	90th percentile	95th percentile
advantageousness of the occupation	-6244.95*** (408.46)	-7151.12*** (640.81)	-9398.77*** (841.30)	-12026.18*** (1425.93)	-14633.01*** (2835.41)
wage	-2.89*** (0.18)	-3.44*** (0.25)	-5.20*** (0.35)	-7.62*** (0.46)	-8.97*** (1.13)
vocational training but no high school	-150.34*** (24.47)	-192.35*** (32.06)	-310.21*** (62.95)	-494.32*** (89.01)	-613.00*** (99.43)
high school but no vocational training	-189.96*** (34.76)	-274.74*** (45.81)	-445.24*** (73.44)	-591.86*** (143.15)	-966.73*** (175.42)
high school and vocational training	-135.06*** (45.28)	-175.13*** (62.35)	-192.51 (119.42)	-295.00** (143.69)	-138.76 (265.54)
technical college	-238.00*** (25.81)	-324.52*** (33.83)	-499.00*** (64.40)	-752.27*** (88.15)	-1072.42*** (103.83)
university	-258.90 (24.57)	-357.74*** (32.40)	-548.78*** (63.52)	-848.39*** (87.95)	-1342.72 (97.17)
Hamburg	-80.05 (53.04)	-134.60* (81.77)	-187.30* (113.65)	-366.55** (181.52)	-164.67 (323.48)
Lower Saxony	-199.01*** (48.80)	-245.65*** (82.56)	-280.54** (104.51)	-442.97*** (149.91)	282.13 (278.26)
Bremen	-11.76 (126.14)	122.44 (150.77)	213.35 (143.32)	99.45 (236.39)	-115.92 (305.85)
North Rhine-Westphalia	-214.94*** (46.79)	-261.49*** (73.45)	-299.77*** (96.46)	-425.42*** (149.90)	393.07 (249.89)
Hesse	-284.51*** (45.29)	-367.00*** (74.37)	-475.27*** (92.61)	-795.62*** (145.22)	-1026.79*** (253.48)
Rhineland-Palatinate	-321.75*** (46.86)	-421.56*** (74.94)	-585.93*** (95.49)	-785.96*** (166.36)	-1032.87*** (257.81)
Baden-Württemberg	-383.10*** (45.83)	-501.83*** (73.15)	-661.89*** (92.01)	-985.34 (145.72)	-1377.88*** (246.85)
Bavaria	-301.64*** (45.11)	-403.13*** (72.36)	-538.84*** (95.34)	-846.94*** (145.84)	-1278.73 (249.28)
Saarland	-224.29*** (60.27)	-296.90*** (95.33)	-244.54 (155.32)	-370.19* (224.74)	-492.30 (346.72)
energy and mining	-280.12** (133.41)	-483.49*** (144.22)	-543.98*** (126.83)	-634.27*** (186.82)	-494.77 (440.83)
manufacturing	-212.03 (131.83)	-340.35** (145.62)	-326.75*** (118.91)	-348.71 (158.64)	-350.23 (391.00)
construction	216.88 (135.90)	181.39 (145.62)	316.63** (118.91)	459.47*** (158.64)	653.68 (391.00)
services	-172.71 (131.62)	-282.54* (144.63)	-261.42** (120.52)	-242.52 (157.95)	-94.53 (392.16)
public sector and other	-206.06 (130.97)	-347.52** (145.67)	-335.07*** (123.27)	-309.35* (171.70)	-237.44 (408.86)
size of the establishment	-0.0280*** (0.0013)	-0.0128*** (0.0005)	-0.0181*** (0.0009)	-0.0159*** (0.0011)	-0.0200*** (0.0019)
constant	964.18*** (141.52)	1354.55*** (174.10)	1836.74*** (162.10)	2697.52*** (230.33)	3773.81 (467.01)

Notes: Standard errors in parentheses. * , (**), (***) indicates significance at the 10, (5), (1) per cent level. For a detailed description of variables used see section 2.

FIGURE 2.— Censored quantile regressions for men, dependant variable: lifetime unemployment in days, confidence intervals



Nevertheless, a word of caution might be appropriate: we use relatively crude measures to control for individual or job characteristics. Furthermore our regressions cannot guarantee that occupation-specific human capital gained early in the professional career causally affects the amount of lifetime unemployment. As always, we cannot completely rule out that our results are caused by unobserved personal heterogeneity: It could be the case that individuals with unobservable characteristics rewarded by the labor market but not captured by control variables like the wage earned when turning 25 are more likely to get into advantageous occupations in the first place.

We now turn to the coefficients of our various control variables. When estimating the amount of lifetime unemployment for men we find that most coefficients are statistically significant and have the expected sign.

This is by all means the case for the **wage** earned when turning 25. This wage does have a statistically significant effect on the amount of lifetime unemployment and the coefficient does have the expected sign: Higher wages (which might be interpreted as a proxy for otherwise unobserved favorable personal characteristics) are associated with fewer days of unemployment over the professional career. This relationship is especially pronounced for higher quantiles of the amount of lifetime unemployment's distribution function.

The level of **education** is strongly related to the amount of lifetime unemployment. Our control group consists of individuals with neither high school diploma nor vocational training. Its members have by far the highest expected amount of lifetime unemployment. Individuals with vocational training and particularly those with a tertiary degree do on average much better. The only parameter estimates not statistically significant on the 1% level are related to individuals who hold a high-school diploma and have completed vocational training. Since this group has historically been relatively small the large confidence intervals found for its member are perhaps not surprising.

When it comes to the **region** the control group is the North German *land* of Schleswig-Holstein. As could be expected individuals who early in their professional career work in Schleswig-Holstein or other *länder* known for a poor economic performance tend to be faced with a comparatively high amount of lifetime unemployment. By contrast those who work in Bavaria or Baden-Württemberg, Germany's two most prosperous *länder* on their 25th birthday ceteris paribus accumulate a significantly smaller number of unemployment days during their professional career. Once again this effect is strongest for higher quantiles of the distribution function of the amount of lifetime unemployment. The exception is the 95th percentile where a number of coefficients are statistically insignificant and confidence intervals in general relatively wide. This might be explained by the modest number of individuals with lifetime unemployment higher than the 95th percentile of its distribution function.

The same phenomenon might be responsible for the observation that result for the **sector** variable are not altogether clear-cut. The reference category is the agricultural sector and now none of the coefficients for the 95th percentile differs statistically significantly from zero. Besides, for only two of the five other sectors do we find coefficients that are more or less statistically different from zero over all other estimated quantiles: individuals engaged in the energy/mining sector at age 25 as well as those working for the "public sector and other" category can expect a comparatively small amount of lifetime unemployment. For other sectors many coefficients are insignificant at the 5% level. A priori we would have assumed a bigger role for our sectoral variable. A reason for their apparent unimportance might be the crudeness of our definition of the different sectors.

While the effects of a number of explanatory variables are especially pronounced for higher quantiles of the the amount of lifetime unemployment's distribution function, this is not the case for the variable representing **size of the establishment**. As detailed above this variable counts the number of employees of the firm for which the individual worked at age 25 and is amongst other things meant to catch otherwise unobserved heterogeneity. As could be expected working in a large firm early in the professional career is ceteris paribus associated with a smaller amount of lifetime unemployment.

5.2. *Women*

While the coefficients estimated for men usually hold the expected sign and are statistically significant table IV and figure 3 show that this is not always the case for the CQRs for women. Here, our results do not imply a statistically significant relationship between lifetime unemployment and a relatively large number of explanatory variables.

Regarding the **advantageousness of the occupation** held early in the professional career results from the estimations for men are to some extent confirmed when looking at the data for women. That is, we again find consistently negative coefficients associated with our measure of the advantageousness of the occupation held when turning 25 (with the exception of the 90th percentile). However, none of the five estimated coefficients is statistically significant at the 10% level.

As has already been mentioned we obtain relatively few significant and some outright counterintuitive results for some of the other explanatory variables when estimating the amount of lifetime unemployment for women. This is for instance the case for **wages**: for women wages seem not to play a statistically significant role in determining the amount of lifetime unemployment. Coefficients are all very small and confidence intervals pretty large.

At first glance, some of the results for **education** seem puzzling: while for all quantiles but the 95th percentile university graduates face a smaller probability of being affected by high unemployment than individuals with neither high school diploma nor professional education this is not true for graduates from technical colleges. Even though these individuals also hold a tertiary degree our CQRs consistently say that they are more likely to be faced with a high amount of lifetime unemployment than members of the control group (though this difference is only significant at the 5% level for the 75th, 80th and 85th percentiles). Similarly

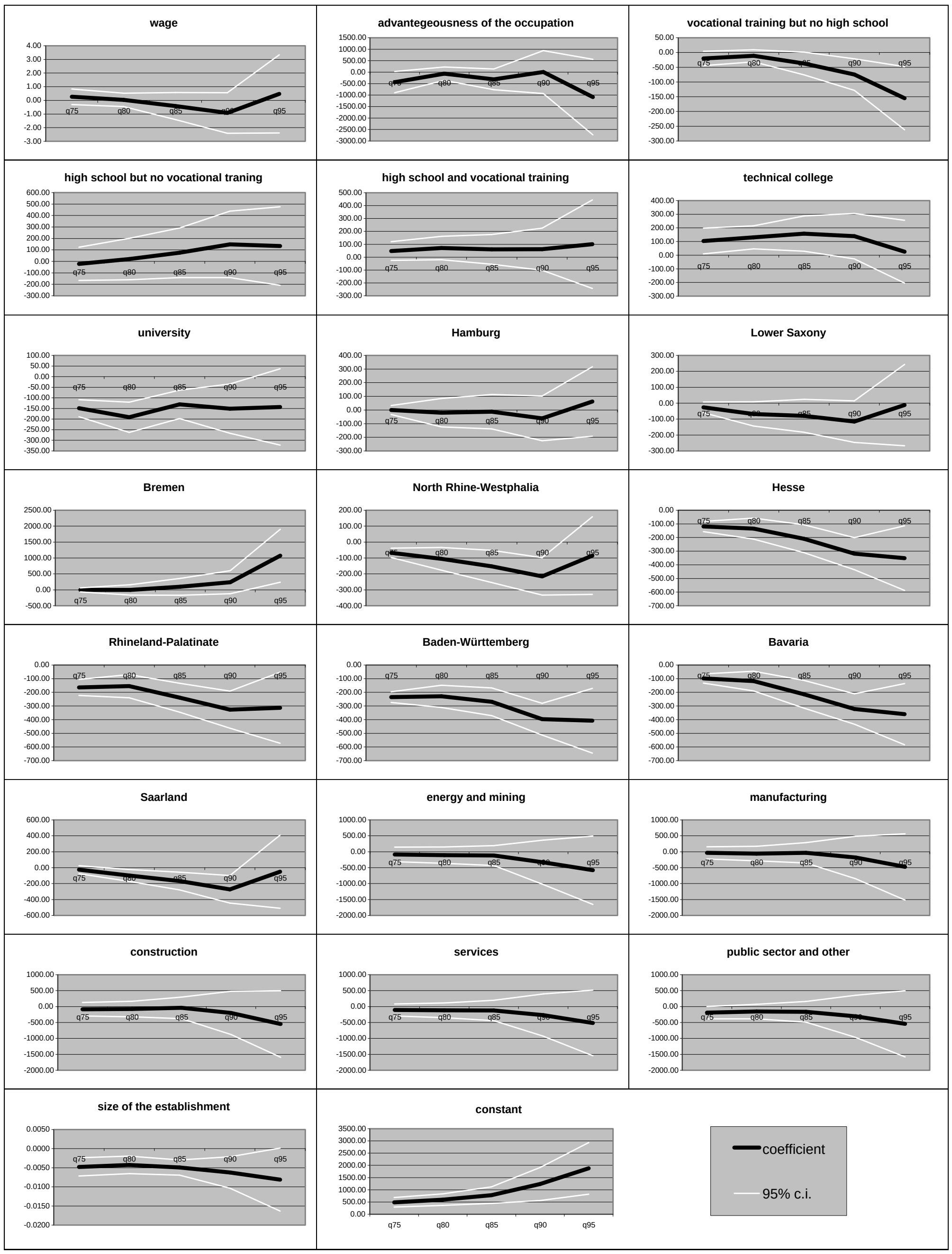
TABLE IV

Censored quantile regressions for women, dependant variable: lifetime unemployment in days

variable	75th percentile	80th percentile	85th percentile	90th percentile	95th percentile
advantageousness	-435.65*	-65.23	-314.14	4.43	-1086.91
of the occupation	(235.29)	(147.40)	(231.24)	(475.95)	(840.23)
wage	0.026	0.03	-0.42	-0.93	0.48
	(0.29)	(0.25)	(0.51)	(0.76)	(1.46)
vocational training	-20.12	-11.34	-37.08*	-74.77***	-155.21***
but no high school	(12.48)	(10.35)	(19.62)	(27.26)	(54.26)
high school but no	-21.16	20.36	75.02	148.66	133.33
vocational training	(74.18)	(91.80)	(110.00)	(147.85)	(174.63)
high school and	47.51	71.27	60.42	61.72	101.04
vocational training	(36.85)	(46.35)	(58.86)	(83.54)	(175.30)
technical college	103.50**	130.85***	158.05**	138.99	25.46
	(48.00)	(43.08)	(65.86)	(84.99)	(117.49)
university	-148.60***	-191.23***	-130.44***	-150.65**	-142.87
	(20.49)	(36.38)	(34.15)	(59.22)	(91.67)
Hamburg	-0.31	-20.13	-12.74	-61.77	62.03
	(17.09)	(52.99)	(64.43)	(83.57)	(129.77)
Lower Saxony	-26.61	-68.12*	-79.20	-115.83*	-11.78
	(17.07)	(38.79)	(52.81)	(66.85)	(130.29)
Bremen	-2.98	-3.95	96.30	239.75	1073.35**
	(32.29)	(82.94)	(135.32)	(184.40)	(424.14)
North	-67.67***	-105.38***	-152.76***	-215.24***	-84.92
Rhine-Westphalia	(14.93)	(36.20)	(51.71)	(59.84)	(124.32)
Hesse	-119.14***	-134.13***	-208.96***	-318.37***	-350.77***
	(19.12)	(39.30)	(52.24)	(59.84)	(124.32)
Rhineland-	-163.81***	-154.70***	-238.82***	-326.35***	-313.00**
Palatinate	(30.12)	(43.30)	(54.08)	(69.37)	(132.74)
Baden-	-235.17***	-229.09***	-269.10***	-396.94***	-408.06***
Württemberg	(19.87)	(41.57)	(51.81)	(59.25)	(121.05)
Bavaria	-98.50***	-117.52***	-213.75***	-321.93***	-359.71***
	(16.93)	(36.71)	(51.88)	(56.58)	(114.37)
Saarland	-25.12	-97.84**	-167.58***	-272.53***	-50.21
	(25.50)	(38.38)	(56.40)	(88.22)	(234.41)
energy and mining	-83.23	-107.24	-119.61	-330.14	-579.83
	(115.93)	(130.53)	(159.38)	(355.77)	(547.47)
manufacturing	-35.31	-57.09	-33.40	-180.95	-475.67
	(98.55)	(116.65)	(162.87)	(339.33)	(530.70)
construction	-84.26	-77.67	-40.61	-199.24	-547.18
	(108.52)	(123.69)	(173.15)	(346.83)	(532.21)
services	-107.37	-115.87	-122.32	-267.44	-513.69
	(96.50)	(117.05)	(162.84)	(338.16)	(525.80)
public sector and	-190.71*	-156.25	-126.98	-306.35	-541.81
other	(99.52)	(117.22)	(166.10)	(337.56)	(530.45)
size of the	-0.0048***	-0.0043***	-0.0049***	-0.0062***	-0.0081*
establishment	(0.0012)	(0.0012)	(0.0010)	(0.0021)	(0.0042)
constant	487.92***	596.60***	781.22***	1238.00***	1876.51***
	(99.04)	(129.95)	(172.76)	(344.53)	(537.16)

Notes: Standard errors in parentheses. * , (**), (***) indicates significance at the 10, (5), (1) per cent level. For a detailed description of variables used see section 2.

FIGURE 3.— Censored quantile regressions for women, dependant variable: lifetime unemployment in days, confidence intervals



surprising, a high school diploma does not bring a statistically significant decline in the amount of lifetime unemployment. But even though individuals with a high school diploma and vocational training do not face significantly lower lifetime unemployment than the control group, women without high school diploma but with professional training do (at least when looking at the 85th to 95th percentile). A likely explanation for the puzzling results for women with high school diploma but no vocational training, high school diploma and vocational training and a degree from a technical college is that all these groups are rather small in our sample. Altogether only 1179 women fall in one of these three categories, that is less than 5% of our sample.

The dummies indicating the **region** where the individual works on her 25th birthday is associated with more intuitive results than the education dummies. Women who early in their professional career work in a *land* that is economically well-off like Bavaria and Badem-Württemberg tend to face a comparatively small amount of lifetime unemployment. In contrast those employed in Schleswig-Holstein, Bremen or other *länder* with a more difficult economic environment accumulate a significantly larger number of unemployment days during their professional career. All in all regional effects are similar for men and women.

In a way this can also be said about our **sector** variables. For men only two sectors exhibit coefficients that differ significantly from those of our reference category (agriculture) over all estimated quantiles. For women results are even more sobering: for all 5 sector variables not a single coefficient differs from zero at the 10% level. Apparently our rather crude definition of the different sectors (already mentioned above) combined with the more general problems of our regressions for women (also discussed above) leads to a very low explanatory power of the sector variables in our estimates of women's amount of lifetime unemployment.

Apart from the constant, the last variable in our regression is again the **size of the establishment**. Here, our results from the regressions for women are as clear as those for men. For women working at a large firm at the age of 25 is ceteris paribus associated with a statistically significantly reduction in the amount of lifetime unemployment. This is the case for all quantiles of the distribution function of the amount of lifetime unemployment used for our estimations and especially pronounced for the 90th and 95th percentile.

6. CONCLUSION

In this paper we first of all showed that lifetime unemployment is very unevenly distributed in West Germany. Looking at selected cohorts we found that more than 60% of the individuals in our sample were not affected by unemployment between the age of 25 and 50. On the other hand half of the total amount of unemployment falls upon 5% of the men and 6% of the women in our sample. This result makes it politically highly relevant to find out more about the factors that are associated with very high lifetime unemployment.

Using censored quantile regressions we found that education and several characteristics of the job held when turning 25 have a statistically significant effect on the amount of lifetime unemployment. We also documented that the advantageousness of the occupation held early in the professional career is negatively associated with the amount of lifetime unemployment and that this connection is especially strong for higher quantiles of the distribution of the amount of lifetime unemployment.

If this finding is really caused by interpersonal differences in occupation-specific human capital it has two important implications: It firstly lends support to theories by Ljungqvist and Sargent (1998) and others that stress the connection between human capital and unemployment. Secondly it has direct policy implications. If some individuals are faced with high unemployment because they invested in disadvantageous occupation-specific human capital in the past it might be advisable to publicly fund re-training programs that provide these individuals with more advantageous occupation-specific human capital.

ACKNOWLEDGEMENTS

We thank Martin Dietz, Helmut Rudolph, Philipp vom Berge and Ulrich Wenzel for helpful comments on an earlier draft of this paper.

REFERENCES

Buchinsky, Moshe (1994). Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica* **62** 405–458.

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

Fitzenberger, Bernd (1997). A Guide to Censored Quantile Regression. In Madalla, G.S. and Rao, C.R. (Eds.). *Handbook of Statistics*, Elsevier, Amsterdam.

Fitzenberger, Bernd; Osikominu, Aderonke and Völter, Robert (2006). Imputation Rules to Improve the Education Variable in the IAB Employment Subsample. *Schmollers Jahrbuch: Journal of Applied Social Science Studies* **126** 405–436.

Galiani, Sebastian and Hopenhayn, Hugo (2003). Duration and risk of unemployment in Argentina. *Journal of Development Economics* **71** 199–212.

Gustavsen, Geir Waehler; Jolliffe, Dean and Rickertsen, Kyrre (2008). Censored Quantile Regression and Purchases of Ice Cream. *presented at the American Agricultural Economics Association Annual Meeting*

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.

Koenker, Roger and Bilias, Yannis (2001). Quantile regression for duration data: A reappraisal of the Pennsylvania Reemployment Bonus Experiments. *Empirical Economics* **26** 199–220.

Koenker, Roger and Hallock, Kevin (2001). Quantile Regression. *Journal of Economic Perspectives,* **15** 143-156.

Kurtz, Beate and Scherl, Hermann (2001). Zur interpersonalen Verteilung von Arbeitslosigkeit in kohortenbezogener langfristiger Betrachtung: Untersuchung am Beispiel männlicher Arbeitnehmer der Jahrgänge 1925 bis 1930. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* **34** 127–136.

Ljungqvist, Lars and Sargent, Thomas (1998). The European Unemployment Dilemma. *Journal of Political Economy* **106** 514–550.

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.

Ludsteck, Johannes and Haupt, Harry (2007). An Empirical Test of the Reder Hypothesis and Specific Human Capital Against Standard Wage Competition. *University of Munich, Dicussion Papers in Economics 1977* **34** 127–136.

Machado, Jose and Santos Silva, Joao (2008). Quantiles for Fractions and Other Mixed Data. *University of Essex, Economics Discussion Paper 656*

Obben, James; Engelbrecht, Hans-Jürgen and Thompson, Wesley (2002). A Logit Model of the Incidence of Long-Term Unemployment. *Applied Economics Letters* **9** 43–46.

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

Tobin, James (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica* **26** 68–78.

APPENDIX A: OCCUPATIONS

Tables V and VI detail the most common occupations on January 1, 1975 and December 31, 2004 for men and women, respectively. Table VII and table VIII list the ten most advantageous and the ten most disadvantageous occupations.

As could be expected many of the most disadvantageous occupations for men are associated with manual tasks while advantageous occupations often involve knowledge of IT or other new technologies. For women many of the most advantageous occupations are associated with health care. Besides occupations that necessitate a relatively low qualification are often replaced by related occupations that in general require more skills (for instances *data typists* are among the ten most disadvantageous occupations while *data processing specialists* make the top ten of the most advantageous occupations).

TABLE V

10 most common occupations for men

	occupation	relative frequency[10]
January 1, 1975		
1	Office specialists	6.87
2	Motor vehicle drivers	4.76
3	Electrical fitters, mechanics	3.23
4	Bricklayers	2.64
5	Entrepreneurs, managing directors, divisional managers	2.54
6	Warehouse managers, warehousemen	2.31
7	Engine fitters	2.30
8	Motor vehicle repairers	2.23
9	Stores, transport workers	1.82
10	Other technicians	1.76
December 31, 2004		
1	Office specialists	8.15
2	Motor vehicle drivers	4.61
3	Data processing specialists	3.07
4	Electrical fitters, mechanics	3.05
5	Other technicians	2.30
6	Stores, transport workers	2.23
7	Motor vehicle repairers	2.20
8	Bank specialists	2.17
9	Entrepreneurs, managing directors, divisional managers	2.14
10	Salespersons	2.09

TABLE VI

10 most common occupations for women

	occupation	relative frequency[11]
January 1, 1975		
1	Office specialists	19.44
2	Salespersons	11.01
3	Household cleaners	6.68
4	Stenographers, shorthand-typists, typists	6.3
5	Nurses, midwives	2.67
6	Bank specialists	2.61
7	Medical receptionists	2.45
8	Accountants	2.41
9	Clothing sewers	2.09
10	Other housekeeping attendants	1.92
December 31, 2004		
1	Office specialists	23.14
2	Salespersons	8.29
3	Nurses, midwives	5.02
4	Medical receptionists	4.67
5	Bank specialists	3.31
6	Nursery teachers, child nurses	3.27
7	Household cleaners	3.18
8	Stenographers, shorthand-typists, typists	2.79
9	Social workers, care workers	2.72
10	Wholesale and retail trade buyers, buyers	2.03

TABLE VII

10 MOST DISADVANTAGEOUS AND ADVANTAGEOUS OCCUPATIONS FOR MEN

occupation		relative employment[12]
most disadvantageous occupations		
1	Bricklayers	-1.65228
2	Warehouse managers, warehousemen	-.71962
3	Plant fitters, maintenance fitters	-.66777
4	Building laborer, general	-.60741
5	Engine fitters	-.5484
6	Miners	-.53433
7	Foremen, master mechanics	-.5128
8	Building fitters	-.46225
9	Carpenters	-.43864
10	Concrete workers	-.41174
most advantageous occupations		
10	Other technicians	.53351
9	Locksmiths, not specified	.53766
8	Cooks	.54361
7	Bank specialists	.60747
6	Electrical engineers	.72768
5	Other engineers	.76657
4	Workforce not further specified	.89973
3	Assistants (no further specification)	1.01128
2	Office specialists	1.2758
1	Data processing specialists	2.39356

TABLE VIII

10 MOST DISADVANTAGEOUS AND ADVANTAGEOUS OCCUPATIONS FOR WOMEN

occupation		relative employment[13]
most disadvantageous occupations		
1	Stenographers, shorthand-typists, typists	-3.50873
2	Household cleaners	-3.49541
3	Salespersons	-2.7173
4	Clothing sewers	-1.97106
5	Accountants	-1.32408
6	Packagers, goods receivers, despatchers	-.98757
7	Other housekeeping attendants	-.95379
8	Electrical appliance, electrical parts assemblers	-.8596
9	Goods examiners, sorters, n.e.c.	-.56314
10	Data typists	-.52097
most advantageous occupations		
10	Data processing specialists	.64897
9	Bank specialists	.70065
8	Chartered accountants, tax advisers	.70782
7	Workforce not further specified	.78888
6	Home wardens, social work teachers	1.34032
5	Nursery teachers, child nurses	1.69596
4	Medical receptionists	2.22013
3	Social workers, care workers	2.28646
2	Nurses, midwives	2.34674
1	Office specialists	3.696