MAGISTERARBEIT

Titel der Magisterarbeit

„Duration analysis of unemployment spells in the Austrian labor market using linked employer-employee data"

Verfasserin

Sophie Oberhauser, Bakk.rer.soc.oec.

angestrebter akademischer Grad

Magistra der Sozial- und Wirtschaftswissenschaften (Mag.rer.soc.oec.)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 066 913
Studienrichtung lt. Studienblatt:	Magisterstudium Volkswirtschaftslehre
Betreuerin / Betreuer	Dr.rer.pol. Selver Derya Uysal

Contents

List of Figures

List of Tables

Acknowledgements

I thank my supervisor Dr.rer.pol. Selver Derya Uysal for helping me to focus on the big picture, when I started to get lost in details and for all the quick answers, when I was in need of them. I am also thankful to DI Erich Heil for the extraction of the sample and the access to the data. Then, I want to thank Mag. Jörg Dominik Walch, MA for valuable comments. Last but not least I am grateful for the support of my husband and my family.

Chapter 1

Introduction

1.1 Motivation

When reading economic reports by the Federal Ministry of Economy, Family and Youth (2010, 2012) for Austria over the past few years, one can see that unemployment has been affected by the financial crisis of 2008, i.e. the number of unemployed individuals rose sharply in 2009 and declined slowly thereafter. Yet, not only crises can affect unemployment. Recent statistics state that unemployment rates rose due to the long winter in 2012/13 (Federal Ministry of Labor, Social Affairs and Consumer Protection, 2013). The overall unemployment rate for the fourth quarter of 2012 is announced to be 4.4% (4.2% for women and 4.6% for men) and the long-term unemployment rate is given by 1.1% (STATISTICS AUSTRIA, 2013b). Compared to the fourth quarter of 2011 the overall unemployment rate was stable before rising in the fourth quarter of 2012, while gender specific developments were ambiguous (STATISTICS AUSTRIA, 2012).

All statements above are based on a static measure, i.e. the unemployment rate. This measure provides a basic understanding of the Austrian labor market, but this statistic bears some problems. Governatori et al. (2009) give a great example on how static measures may fail to capture the actual labor market situation. In the following I apply their example to the numbers of the Austrian labor market. The unemployment rate of 4.3% in 2012 can be interpreted in two different ways:

- One way, where unemployment is equally divided across the population, i.e. every individual is unemployed for approximately 2.2 weeks (52×0.043) in 2012.
- Second, it could mean that 4.3% of the labor force were unemployed for the whole year (long-term unemployment).

It is obvious that reality lies somewhere between these two extreme cases. Apart from the problem of interpreting the unemployment rate, numbers can also be biased due to dropping out of the register before finding employment or not filing for unemployment at all. An example: The unemployment rate of women in the fourth quarter of 2012 is lower than the unemployment rate of men. This could mean women are less likely to be unemployed or it could be a result of personal choices to stay out of labor force.

Taking the above factors into account the duration of unemployment as additional indicator becomes of interest, but even the interpretation of the duration of a spell is not straight forward. When talking about unemployment, one can either consider the time from one employment to the next or the time spent in the unemployment register, which might lead to different spell lengths. When filling out labor market surveys individuals were found to prefer measuring the duration of unemployment in full months and also to prefer even over odd numbers. The problems related to this type of behavior and the trouble of interpreting incomplete unemployment duration has been discussed by Corak (1996).

The issues related to the interpretation of static measures of unemployment and the availability of data that help circumventing the problem of attrition inspired me to investigate unemployment in the Austrian labor market further. In economic literature several approaches have been used to model unemployment duration. One is *network theory*, a graph theoretic approach, which looks at the structure of an individual's network and the flow of information within it. Examples that underline the importance of network structure and the way individuals are able to pass on information can be found in Calvó-Armengol and Jackson (2004). Network theory can give interesting results and insights, but data for this graph theoretic approach are not available for the Austrian labor market. Another method used in economic literature is *duration analysis*, which focuses on the effects of covariates on the duration of spells, i.e. unemployment or employment.

In the last decades there have been many studies on unemployment using duration analysis. In a study of various countries (Kavkler et al., 2009) the Austrian analysis is based on a small sample from the Microcensus. Other papers, which take the effect of changes in unemployment benefits in Austria into account, are based on individual longitudinal data, i.e. linked employer-employee data (LEED) (Winter-Ebmer, 1998; Lalive and Zweimüller, 2004; Lalive et al., 2006; Lalive, 2008). This paper should contribute in describing the factors influencing unemployment duration in the Austrian labor market with LEED and incorporating measures of industry-specific characteristics into the analysis. The fact that duration analysis as an instrument has gained importance and the availability of linked employer-employee data supported my choice of analysis. The Austrian linked employer-employee database (AMDB) is a dataset that depicts the Austrian population very well.

The thesis aims to describe the duration of *out-of-work spells*[1] (in the following also referred to as unemployment) for the Austrian labor force in the years of 2001-2012. Special focus will be on industry-specific characteristics. The main questions concerning these characteristics are:

- Does previous employment in industries with seasonal peaks increase the chances of exiting unemployment?
- Does coming from industries with high labor turnover lead to longer or shorter unemployment spells?
- Does previous employment in a big industry shorten the unemployment duration?

The structure of the thesis is as follows: First the basic concept of duration analysis will be given and the models will be discussed. Then I will explain the Austrian unemployment system before I will summarize the literature on duration analysis and unemployment. Chapter 1 is concluded with data description and the concept of out-of-work spells. Chapter 2 focuses on the empirical analysis. The first section is about descriptive statistics of the population and of the sample and the restrictions that have been imposed prior to the analysis. After the first look at the overall duration of out-of-work spells, the analysis will be split into two samples of men and women. Reasons for these separate analyses are differences in the availability of a covariate indicating the number of children and the fact that the decision of participating in or remaining out of the labor force can be different for women and men. Following the data description the comparison of results of the nonparametric, semiparametric and parametric analysis is given. Chapter 3 finishes the thesis with concluding remarks and possibilities for further research.

[1]Out-of-work spells are defined as the time period in between two employments. For further explanation see Section 1.5.2.

1.2 Duration Analysis

Duration analysis is known under various names, i.e. survival analysis, event history analysis and reliability analysis. The objective of this analysis is to model time-to-event data, which means the time spent in a state is analyzed until the individual moves to another state. In this thesis the change of state will from now on be referred to as *transition*. The aim of this section is to give an overview over the roots, the basic concepts, distributions and models used in duration analysis.

The method has its roots in bio-statistics, medicine and biology. In this field of research the method was applied to clinical trials to assess whether patients were able to live longer if they get a specific treatment or to analyze the time to death after a certain diagnosis. In the medical context the event of interest would be death, getting a disease or the cure of a disease, which led to the name *survival analysis*. The foundation of the method are *life tables*. These are tables plotting the happening of ordered events over time, giving the possibility to estimate the probability of survival. In engineering –where survival analysis is referred to as *reliability analysis*– the event of interest would be the failure of a machine or a component.

In the economic context the method is called *duration analysis* and its most common applications are in the labor market. In this field the duration of unemployment or employment are the most important. Lancaster (1992) and Kiefer (1988) give a good overview on modeling survival data in the unemployment context. The introduction to duration analysis in this thesis relies heavily on this literature and on Cleves et al. (2010).

The data underlying the theory are called *time-to-event data*. They are individual based data, i.e. the unit of interest is an individual and not the spell itself. Within a predetermined time frame –the *observation period*– spells are observed and can be used for estimation. An individual can have several spells throughout this time period, but the transition from one state to the state of interest can only happen once per spell. Time-to-event data have two important characteristics, which explain the need for duration analysis. First, since time is nonnegative and not necessarily normally distributed standard Ordinary Least Squares regressions should not be applied. Second, in most cases not all individuals experience the event of interest within the observed period of time, i.e. the whole spell duration has not been observed. These observations are called *censored*.

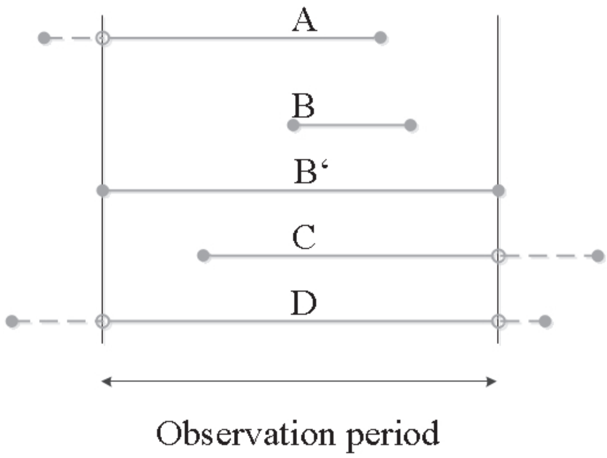


Figure 1.1: Types of censoring

Figure 1.1 depicts the different types of censoring. While spells B and B' represent completed spells (starting and end point are within the period of observation), spell C shows the most common form of censoring in duration analysis: *right censoring*. This means an individual made no transition by the end of the observation period, but could in the future. Other forms of censoring are *left censoring* (Spell A), where within the observation period only the occurrence of the event can be seen (the starting point of the spell is unknown), *censoring at both the beginning and the end of the observation period* (Spell D) and *interval censoring*, where transitions can only be determined to have happened in a certain time interval. As Figure 1.1 shows observed individuals can start their spells at different points in time. In the analysis the starting point $t_0 = 0$ is defined as the time an individual enters the state of interest (Lancaster, 1992).

Both the violation of normally distributed spell lengths and the presence of censoring make it (in most cases) impracticable to use standard Ordinary Least Squares regression for estimation, thus duration analysis is preferred.

The method is based on the distribution of a continuous[2] random variable T, which represents the time until the transition takes place. For this part it is assumed that the population is homogeneous, i.e. every spell length is drawn from the same probability distribution. The realization of the spell length $t > 0$ represents a specific time interval. The probability distribution function of the variable time will be referred to as $F(t)$, where

$$F(t) = P(T < t) \tag{1.1}$$

[2]In economics, time is often treated as continuous, as it is a better fit to the concept, because "*[...] there is typically no natural period in which economic decisions are taken [...].*" (Kiefer, 1988, p.655).

and the corresponding density function will be denoted $f(t)$, where

$$f(t) = \frac{dF(t)}{dt}. \tag{1.2}$$

The two most important measures in duration analysis are the survivor rate function $S(t)$ and the hazard rate function $\lambda(t)$. The former gives the probability that an individual will not experience the event until –survive beyond– time t. The survivor function is therefore the complement of the probability distribution function:

$$S(t) = P(T \geq t) = 1 - F(t). \tag{1.3}$$

The hazard rate function is the instantaneous probability of transition given the individual did not experience the event of interest until t:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{1.4}$$

It can be seen that multiplication with the term Δt leads to the simple conditional probability. The hazard rate function –which can be constant, increasing, decreasing or nonmonotonic over time– is the negative growth rate of the survivor function:

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\frac{dF(t)}{dt}}{S(t)} = -\frac{\frac{dS(t)}{dt}}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d \ln S(t)}{dt}. \tag{1.5}$$

Another function of interest is the cumulative hazard rate function, i.e. the cumulated risk over time. It is given by

$$\Lambda(t) = \int_0^t \lambda(u) du. \tag{1.6}$$

One advantage of duration analysis is that knowing one function allows one to derive every other function through the following relationships: Plugging the last term of Equation (1.5) into the cumulative hazard function in Equation (1.6) we get

$$\Lambda(t) = -\ln S(t) \Leftrightarrow S(t) = e^{-\Lambda(t)} \tag{1.7}$$

and therefore

$$F(t) = 1 - e^{-\Lambda(t)}, \tag{1.8}$$

$$f(t) = \lambda(t) e^{-\Lambda(t)}, \tag{1.9}$$

$$\lambda(t) = \frac{d\Lambda(t)}{dt}. \tag{1.10}$$

The hazard rate function has another important feature for estimation. Through the shape of the hazard function one can determine *duration dependence*. There can be either positive or negative duration dependence or the risk can be constant or vary over time. Positive duration dependence corresponds to an increasing risk over time and is analytically defined by $\frac{d\lambda(t)}{dt} > 0 \; \forall t \in \mathcal{T}$. Negative duration dependence ($\frac{d\lambda(t)}{dt} < 0 \; \forall t \in \mathcal{T}$) is equal to a declining risk of transition over time. If there is no duration dependence (a constant hazard rate function) then $\frac{d\lambda(t)}{dt} = 0 \; \forall t \in \mathcal{T}$ (Kiefer, 1988).

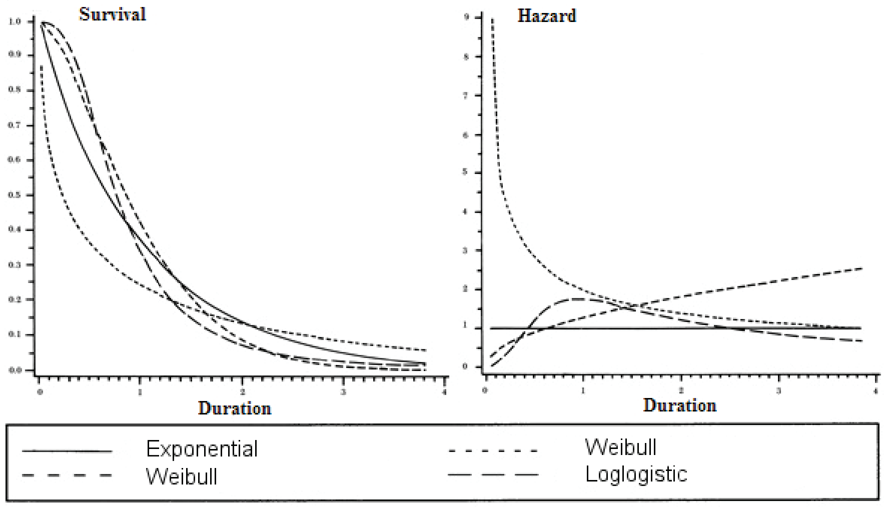
For parametric estimation it is necessary to make assumptions about the distribution of risk over time. Following the paper of Kiefer (1988) three distributions shall be briefly discussed. The Weibull distribution allows for a monotonic increasing (positive duration dependence) or decreasing hazard rate function (negative duration dependence). A special case of this distribution –with no duration dependence, i.e. a constant risk of transition– is the Exponential distribution. With the Loglogistic distribution the hazard rate function can be nonmonotonic.

Table 1.1: Cumulative distribution, density, survivor, hazard and cumulative hazard function in the Exponential, Weibull and Loglogistic model

Exponential	**Weibull**	**Loglogistic**
$F(t) = 1 - e^{-\phi t}$	$F(t) = 1 - e^{-\phi t^p}$	$F(t) = 1 - [1/(1 + \phi t^{\frac{1}{\gamma}})]$
$f(t) = \phi e^{-\phi t}$	$f(t) = \phi p t^{p-1} e^{-\phi t^p}$	$f(t) = \phi \frac{1}{\gamma} t^{\frac{1-\gamma}{\gamma}} / (1 + \phi t^{\frac{1}{\gamma}})^2$
$S(t) = e^{-\phi t}$	$S(t) = e^{-\phi t^p}$	$S(t) = 1/(1 + \phi t^{\frac{1}{\gamma}})$
$\lambda(t) = \phi$	$\lambda(t) = \phi p t^{p-1}$	$\lambda(t) = \phi \frac{1}{\gamma} t^{\frac{1-\gamma}{\gamma}} / (1 + \phi t^{\frac{1}{\gamma}})$
$\Lambda(t) = \phi t$	$\Lambda(t) = \phi t^p$	$\Lambda(t) = \ln(1 + \phi t^{\frac{1}{\gamma}})$

Source: Kiefer (1988), own representation

Table 1.1 gives the relationships of functions for these distributions in the notations established in Equations (1.7) to (1.10). In the Weibull model the shape parameter is referred to as p, in the Loglogistic model the shape parameter is called γ. The function ϕ incorporates heterogeneity. It consists of the baseline hazard –the underlying risk, which is the same for every individual– and the risk corresponding to certain characteristics. The Exponential distribution is a special case of the Weibull distribution (with $p = 1$) with no duration dependence, since the derivative of a constant is equal to zero. The assumption that duration has no effect on the length of an unemployment spell is in principle not realistic. As already explained the Weibull distribution corresponds to a monotonic hazard rate function, where both positive ($p > 1$) and negative duration dependence ($p < 1$) is possible. The hazard rate function of the Loglogistic model does not have to be monotonic and therefore allows modeling in such a way that hazard rates can change in both directions over time. Depending on the parameter γ, the risk of transition can be monotonic decreasing from ∞ if $\gamma > 1$ or from one if $\gamma = 1$. The hazard is nonmonotonic (increasing from 0 and then decreasing) if $\gamma < 1$.



Source: Kiefer (1988) p.653-654

Figure 1.2: Survivor functions (left) and hazard rate functions (right) to the corresponding distribution functions

Figure 1.2 gives the survivor and hazard rate functions for the three distributions. The shapes of the hazard functions following various distributions can be found on the right hand side, the corresponding survivor rate functions are depicted on the left hand side. The solid horizontal line gives no duration dependence, i.e. a constant hazard over time (Exponential distribution). The short-dashed line shows the course of the hazard rate function when negative duration dependence is present in the Weibull distribution, whereas the medium-dashed line gives the hazard with positive duration dependence. The long-dashed line shows an increasing and decreasing risk over time under the Loglogistic distribution.

Up to this point, the incorporation of heterogeneity has not been discussed. Similar to standard OLS regression, this can be done by including covariates. To do so, one defines that in Table 1.1 $\phi = \lambda_0(t)e^{X\beta}$, where $\lambda_0(t)$ represents the baseline hazard –the hazard that is the same for all individuals– and $e^{X\beta}$ is the risk associated with the characteristics. $X\beta$ is defined as

$$X\beta = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{N1} \end{pmatrix} \beta_1 + \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{N2} \end{pmatrix} \beta_2 + \cdots + \begin{pmatrix} x_{1K} \\ x_{2K} \\ \vdots \\ x_{NK} \end{pmatrix} \beta_K, \tag{1.11}$$

where X is a $\big(N \times (K+1)\big)$ matrix, where each row vector x_n represents a spell $n = 1, 2, 3, \ldots, N$ with a set of K characteristics and $\beta = (\beta_0, \beta_1, \ldots, \beta_K)$ is a $\big((K+1) \times 1\big)$ vector of coefficients.

Making assumptions about the distribution of risk is not always of interest in modeling survival and can lead to wrong results, if the assumption is wrong. Therefore there have been approaches to estimate the effects of covariates without specifying the distribution of the baseline hazard. These models are called *semiparametric*. A popular semiparametric approach is the *Cox Proporional Hazards (PH) model* (Cox, 1972). The basic concept behind the model is the assumption that characteristics have a multiplicative effect, i.e. the baseline hazard is the same for everyone and is just shifted. This assumption is called the *proportionality assumption*. The Cox PH model focuses on the effect of predefined variables on the hazard rate.

Some parametric models are also based on the proportionality assumption. These models are in the Proportional Hazards metric. If the assumption does not hold, another approach, which observes the impact of covariates on the time of survival rather than on the risk of transition, is used. These models are called *Accelerated Failure Time (AFT) models*. When only a comparative analysis of the survival rate is of interest nonparametric analysis can be applied.

1.2.1 Nonparametric models

This subsection focuses on nonparametric estimators, i.e. estimators that compare categorical covariates in the sample. At first the Kaplan-Meier estimator of survivor functions will be presented. In the second part the Nelson-Aalen estimator for the hazard functions will be shortly explained.

Kaplan-Meier estimator

The *Kaplan-Meier estimator* (Kaplan and Meier, 1958), also called the *product limit estimator*, is a function for estimating the survivor function based on *life tables*. These tables describe the number of spells at risk, transitions and censored spells for all time intervals. In a first step the individual spells are ordered according to their duration and then assigned a consecutive number. For each ordering number the corresponding duration, number of individuals who failed and the number of individuals at risk (individuals who survived until the end of the interval) are plotted. Using the information from life tables, the survivor function and the hazard rate function can be estimated. Nonparametric analysis does therefore give comparative results for the values of the categorical covariates (Cleves et al., 2010).

The estimated hazard rate function is the probability of instantaneous failure, given that an individual's survival time equals the interval length t_j, where $j = 1, 2, \ldots, J$ and $t_{J+1} = \infty$. Every recorded censoring or transition τ_i is counted within the interval $t_j \leq \tau < t_{j+1}$. Embedding this into life tables, the estimate of the hazard rate at the end of interval t_j is given by the number of transitions divided by the number of individuals at risk. Analytically this is equal to

$$\hat{\lambda}(t_j) = \frac{d_j}{n_j}, \tag{1.12}$$

where d_j is the number of transitions at t_j and $n_j \in R(t_j)$ denotes the number of individuals in the risk set, i.e. individuals at risk of transition at time t_j. Thus the estimate of the survivor function is then given by

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j} = \prod_{j|t_j \leq t} 1 - \frac{d_j}{n_j} = \prod_{j|t_j \leq t} 1 - \hat{\lambda}(t_j). \tag{1.13}$$

Plotting the survivor function in a graph leads to a monotonic nonincreasing step function, where the step height varies with the number of failures at the end of t_j. In this model censored spells are assumed to happen *after* transitions, if they coincide with them. When censoring occurs, the number of individuals at risk is reduced by the number of censored spells.

In my thesis the Kaplan-Meier estimator will be used to compare the survival functions of the categorical variables.

Nelson-Aalen estimator

When the hazard rate is of interest, the Nelson-Aalen estimator (Nelson, 1972; Aalen, 1978) is applied, for it has better small-sample properties than the Kaplan-Meier estimator given by Equation (1.12) (Cleves et al., 2010). Following the previous notation, the estimate for the cumulative hazard is given by

$$\hat{\Lambda}(t) = \sum_{j|t_j \leq t} \frac{d_j}{n_j}, \tag{1.14}$$

where d_j is the number of failures and $n_j \in R(t_j)$ the number of individuals at risk at the end of tj. Since the estimated cumulative hazard is not a continuous function, and taking a derivative of Equation (1.14) is not possible, the height of the steps is used to calculate the hazard function:

$$\hat{\lambda}(t) = b^{-1} \sum_{j=1}^{D} K_t\Big(\frac{t - t_j}{b}\Big) \Delta\hat{\Lambda}(t_j), \tag{1.15}$$

where $\Delta\hat{\Lambda}(t_j)$ is equal to the step height, b is a bandwidth, $K_t(.)$ is a kernel function (for smoothing the hazard function estimate) and D is the number of times at which transitions occur. In this thesis the Nelson-Aalen smoothed hazard estimate is of minor importance, as it is only applied to gain further insights on the influence of benefit payments on unemployment duration.

1.2.2 Semiparametric model

When the effect of covariates on the risk of transition or on the survival is of interest, the focus has to be on semiparametric and parametric models. In this subsection focus lies on models for which no assumptions about the distribution of the baseline hazard have to be made, i.e. semiparametric models.

Cox Proportional Hazards model

The duration model which experienced the most popularity in economics is the Cox Proportional Hazards model. It was developed by David Cox (1972). The partial likelihood (Cox, 1975) has been applied to estimate the model. Similar to linear regression models, the interest lies in estimating coefficients of covariates which might influence the explained variable, i.e. duration of a spell. The hazard function is given by

$$\lambda(t, X, \beta, \lambda_0) = \lambda_0(t) e^{X\beta}, \tag{1.16}$$

where $\lambda_0(t)$ represents the baseline hazard –the probability of transition at time t, when all covariates are equal to zero and $e^{X\beta}$ is the risk associated with the covariates. In the Cox Proportional Hazards model there is no intercept included, or to be more precise, the intercept would be absorbed by the baseline hazard, which is not estimated in the model (Cleves et al., 2010). Using Equation (1.7) and the relationship from Equation (1.16) we get the corresponding survivor function

$$S(t) = e^{-\int_0^t \lambda_0(t)e^{X\beta}} = e^{-e^{X\beta}\int_0^t \lambda_0(t)} = e^{-e^{X\beta}\Lambda_0(t)}. \tag{1.17}$$

It is assumed that coefficients have a multiplicative effect on the baseline hazard. Thus a partial derivative of Equation (1.16) by a characteristic x_k

$$\frac{\partial \ln \lambda(t, X, \beta, \lambda_0)}{\partial x_k} = \beta_k \tag{1.18}$$

Equations (1.16) and (1.17) consist of unknown parameters λ_0 and β. The aim of the semiparametric model is to estimate the coefficients of the covariates without making assumptions about the baseline hazard. The hazard rate function for an individual m to end the spell at time t_j is given by

$$\frac{\lambda(t_j, x_m, \beta)}{\sum_{n\in R(t_j)} \lambda(t_j, x_n, \beta)}, \tag{1.19}$$

where x_m represents the m^{th} row vector of the matrix X with $n = 1, 2, \ldots, m, \ldots, N$. Recall that the proportional hazard function states that the hazard rate consists of the product of a baseline hazard and a function $e^{X\beta}$ that depends on the covariates (Equation (1.16)). Inserting this property into the estimate, we get

$$\frac{\lambda_0(t_j)e^{x_m\beta}}{\sum_{n\in R(t_j)} \lambda_0(t_j)e^{x_n\beta}} = \frac{e^{x_m\beta}}{\sum_{n\in R(t_j)} e^{x_n\beta}}, \tag{1.20}$$

where $R(t_j)$ gives the number of individuals at risk at time t_j. The partial likelihood function is given by

$$\prod_{m=1}^{N} \frac{e^{x_m\beta}}{\sum_{n\in R(t_j)}^{N} e^{x_n\beta}}. \tag{1.21}$$

Censored spells will only be taken into account in the expression in the sum. The log likelihood is given by

$$lnL(\beta) = \sum_{m=1}^{N} \Big(e^{x_m\beta} - log \sum_{n\in R(t_j)}^{N} e^{x_n\beta} \Big). \tag{1.22}$$

1.2.3 Parametric models

As explained in Section 1.2, assumptions about the baseline hazard are necessary in parametric estimation. Some of these models can have both a Proportional Hazards (PH) and an Accelerated Failure Time (AFT) interpretation (i.e. Weibull and Exponential model), others only work in the AFT framework (Loglogistic, Lognormal or Gompertz model). The former follows the assumption that the effect of covariates on the hazard is a multiplicative one. The AFT (or log-time) interpretation assumes that covariates affect the time that an individual takes to follow its way down the survival curve, i.e. it measures how covariates accelerate the time to transition. The Accelerated Failure Time models are given as

$$ln(t_j) = X\beta + \epsilon_j, \tag{1.23}$$

where t_j is the observed time of transition, i.e. the interval that passes until a transition occurs, and ϵ_j represents an error term, which follows the distribution of the assumed model. We then define τ_j, such that

$$\tau_j = e^{-X\beta} t_j, \tag{1.24}$$

where $e^{-X\beta}$ is the parameter measuring the acceleration of time through the effect of the covariates. Time is defined to be accelerated if $e^{-X\beta} > 1$ and decelerated if $e^{-X\beta} < 1$. Through taking the logarithm of Equation (1.24), we get

$$ln(t_j) = X\beta + ln(\tau). \tag{1.25}$$

The likelihood function is given by

$$L(\beta, \Theta) = \frac{S(t_j|X\beta, \Theta)}{S(t_{0j}|X\beta, \Theta)} \lambda(t_j|X\beta, \Theta)^{c_j}, \tag{1.26}$$

where the fraction represents the probability of surviving for the interval t_{0j} to t_j. The exponent c_j is a parameter that is equal to zero if the observation is censored and one if there is a transition. Thus, if censoring is present the likelihood function is equal to the survivor function. The parameter Θ represents the set of the shape and scale parameter of the corresponding distribution.

The following subsections concentrate on the AFT interpretation of the Weibull model and the Loglogistic model. These sections are based on Cleves et al. (2010).

Weibull Model

The Weibull distribution model depends on two parameters, namely the coefficient β_0 (scale parameter) and a shape parameter p. The variable τ given in Equation (1.24) is assumed to be Weibull distributed with parameters p and β_0. Its cumulative distribution function $F(\tau)$ is

$$F(\tau) = 1 - exp(-(e^{-\beta_0}\tau)^p). \tag{1.27}$$

The baseline survivor function –the survivor function with all covariates equal to zero– at the end of interval t_j is given by

$$S_0(t_j) = exp[-(e^{-\beta_0}t_j)^p] \tag{1.28}$$

and the survivor function is

$$S(t_j) = exp[-(e^{-\beta_0}e^{-X\beta}t_j)^p]. \tag{1.29}$$

Loglogistic model

In the Loglogistic model the variable τ_j is distributed as loglogistic with parameters β_0 and γ. The distribution function of τ looks as follows:

$$F(\tau) = 1 - \frac{1}{1 + (e^{-\beta_0}\tau)^{\frac{1}{\gamma}}} \tag{1.30}$$

The baseline survivor function at time t_j is given by

$$S_0(t_j) = \frac{1}{1 + (e^{-\beta_0}t_j)^{\frac{1}{\gamma}}} \tag{1.31}$$

and the general survivor function at time t_j is given by

$$S(t_j) = \frac{1}{1 + (e^{-\beta_0}e^{-X\beta}t_j)^{\frac{1}{\gamma}}}. \tag{1.32}$$

1.3 The unemployment system in Austria

This section is a general overview of the unemployment system in Austria. For more detailed information see Public Employment Service Austria (AMS) (2013).

The benefits of filing for unemployment are the access to various programs –such as retraining and vocational training– or to the placement process. Individuals have to fill out a form at the Public Employment Service Austria (AMS). It contains questions about personal data, such as age, gender, education level[3] or questions about preferred

[3]The documentation of educational levels from the Employment Service Austria allows to include educational data with great coverage.

job alternatives. While unemployed an individual is still allowed to be marginally employed. Marginal employment is defined as an employment relation which leads to earning less than 376.26 Euro per month for an employment longer than one month and 28.89 Euro per day for an employment shorter than one month in 2012 (Austrian Social Security, 2012). These jobs do not contribute to the unemployment insurance. Data show that marginal employment is often active when looking at unemployed individuals.

1.3.1 Unemployment benefits

The Austrian unemployment system allows individuals to receive unemployment payments according to their previous work and the duration of unemployment. In order to be entitled, individuals have to file for unemployment, i.e. be willing to work, be employable and should obviously be unemployed. Additional criteria, which depend on personal characteristics, have to be fulfilled. When first filing for unemployment benefits individuals have to have 52 weeks of dependent contributory employment (which will from now on be referred to as *employment*) in the last two years. Within these 52 weeks sick leaves which are matched to an employer –sick leaves on the job– and employment outside of Austria –if the country has an agreement with Austria– can be included. For the first claim the required duration of previous employment for under 25-year-olds amounts to 26 weeks within the past year. Having already received unemployment benefits, the minimum duration of employment has to be 28 weeks within the past year.

Benefits are in abeyance during spells of sickness, if the individual is incapable of working or during receipt of unemployment assistance and cease, when employment is found. The amount of unemployment payments depends on the application time over the year and is calculated on the base of the income of a calendar year. If the application is filed before the 30$^{\text{th}}$ of June, the income of the year *before* the last calendar year (including all special payments) is the base. If an individual applies in the second half of the year, the *past calendar year* is taken into account. All unemployment insurance contributions from employment or self insurance of the year of interest are calculated into a monthly gross payment, which cannot be higher than the maximum assessment base. Then a daily net payment is computed from which 55% are paid on a monthly basis (daily net payments · days per month) in the following month. There is also the possibility for receiving additional family payments or other benefits. For individuals who do not possess any yearly basis of assessment in the past year or the year before that, payments from the last six months before application will be taken into account.

It is possible to receive benefits while being marginally employed, though the individual needs to have been without employment for at least one month. With the unemployment registration the individual declares him- or herself to comply with accepting job offers.[4] Should the individual –for whatever reasons– not be available while being unemployed, there will be no entitlement to unemployment payments. If the individual is available, but does not want to accept a job offer, the individual has the possibility to turn the job offer down. The Public Employment Agency has to penalize a turned down job offer by ceasing the unemployment payments for six weeks. Extension of this penalty can be the case if more jobs are turned down.

In Austria the time period during which unemployment insurance is paid varies with age of the individual and the amount of years spent in employment before the application. If the basic requirements from the first paragraph of this subsection are fulfilled, payments can be received for up to 20 weeks. If a person spent three years in employment during the past five years then unemployment payments can be extended to 30 weeks. The Austrian unemployment system also allows for extended payments for people over the age of 40. If an individual had been employed for at least 312 weeks within the last ten years, unemployment payments can be received for 39 weeks. For individuals older than 50 years unemployment payments are extended to up to one year if the individual had been employed for at least 468 weeks within the last 15 years. Individuals who quit the last job or were laid off for misconduct cannot receive unemployment insurance for four weeks. If a new job is found before the maximum amount of weeks of payment is exhausted, the rest of the unemployment insurance can be received later on if the next unemployment spell starts within five years after the cessation of payment and no new entitlement is complied.

Other payments, which can affect the duration of unemployment spells, are unemployment assistance or temporary allowance. The former can be claimed after unemployment payments cease. The amount of unemployment assistance payments depends on various factors, such as the amount of the preceding unemployment payments, the number of children or the income of partners. Depending on the economic situation of the individual, unemployment assistance can be received for one year at maximum. After that, individuals have to apply again or file for unemployment payments, if the individual had been employed 28 weeks within the past year. Temporary allowance can

[4]The Austrian Public Employment Service only gives job offers that match the abilities of the registered individual.

be received after partial retirement in the Austrian social system. The same conditions as for unemployment insurance and unemployment assistance apply, i.e. employability, willingness to work and availability to the Public Employment Agency for placement. In addition the individual's age should be close to the retirement age and the individual had to be unemployed for 12 months in the last 15 months. Temporary allowance can then be received up to the point where the individual transitions into retirement.

1.4 Unemployment and its influencing factors

Duration analysis has gained popularity in the previous decades. There is literature focussing on its application to the labor market (Flinn and Heckman, 1982; Kiefer, 1988; Lancaster, 1992). A large amount of earlier studies focuses on unemployment duration in Great Britain: Lancaster (1979) focused on unemployment spells of unskilled workers. A dataset from a cohort study, including 2,332 British men unemployed in autumn 1978, has been the foundation of many analyses (Narendranathan et al., 1985; Gamerman and West, 1987; Narendranathan and Stewart, 1993). Gamerman and West (1987) applied bayesian estimation, while Narendranathan and Stewart (1993) used semiparametric analysis and parametric analysis (Weibull) in order to be able to compare their results with previous literature. It has been touched upon the fact that exit to employment or to out of the labor force might have an influence on the estimates of the duration of unemployment spells. This issue has been adressed in analyzing Swedish register data (Bring and Carling, 2000). The authors found that misspecification and attrition lead to similar effects like unobserved heterogeneity, i.e. a downward bias. The effects of measurement errors in survey data on nonparametric, semiparametric and parametric analysis have also been of interest in recent literature (Pyy-Martikainen and Rendtel, 2009).

Recently, duration analysis has been applied in studies for many different countries. A labor force survey has been used for estimation of the duration of unemployment spells in a small study of Lombardy's labor market (Mussida, 2007). Similar to Narendranathan and Stewart (1993), the author used a semiparametric competing risks model that differentiates between finding employment or dropping out of the labor force. Exit into different kinds of employment has also been modeled for other countries. The transition to wage earner or self-employment in Spain (Berenguer, 2001) and to formal or informal employment in Mexico (Calderón-Madrid, 2008) has been assessed in the past. There have been several studies for the Slovenian (Van Ours and Vodopivec, 2006; Boršic and Kavkler, 2009) and Romanian labor market (Dănăcică and Babucea, 2010; Ciucă and Matei, 2011). For Romania there have been studies on the effects of personal characteristics such as gender, age, educational level and regions on unemployment duration. Ciucă and Matei (2011) focus on nonparametric analysis, before applying the Cox Proportional Hazards model for the selected covariate *age*. Dănăcică and Babucea (2010) analyze unemployment duration in a Romanian region named Gorj. For Slovenia nonparametric analysis has been used in order to show differences between regions (Boršic and Kavkler, 2009). The authors Van Ours and Vodopivec (2006) assessed the effect of changes in unemployment benefits on unemployment duration.

The effect of unemployment insurance on unemployment duration has been of interest in many papers concerning the Austrian labor market. Similar to this thesis, the estimations in this literature are based on a combination of the Austrian Social Security database and Austrian unemployment register. The literature for Austria focuses on law changes affecting the duration of unemployment benefit payments (Winter-Ebmer, 1998; Lalive and Zweimüller, 2004; Lalive et al., 2006; Lalive, 2008). A paper assessing unemployment insurance and unemployment duration in Austria found that looking at the exit from registered unemployment only can bias the results (Card et al., 2007). The authors show that in the years from 1980-2001 the hazard of exiting registered unemployment after expiration of unemployment benefits rises by 200%, whereas there is only an increase of 20% when looking at the time from one employment to the next. The negative effect of unemployment benefits on duration has also been established for Spain (Bover et al., 2002). Compared with the large amount of data used for the studies of Austria above, the Austrian part of a study comparing personal characteristics and their influence on unemployment duration in Romania, Austria, Slovenia, Croatia and Macedonia (Kavkler et al., 2009) was based on a very small sample of survey data (Microcensus).

For West Germany there is also literature focusing on changes in unemployment duration due to changes in unemployment payments in the 1980s (Hunt, 1995). It has been found that an increase in unemployment compensation affects the age group of 44 to 48-year-olds the most and that women react stronger to these changes. Another interesting approach has been to take personality traits, according to the classifications of the *Big Five*, into account (Uysal and Pohlmeier, 2011). It has been observed that personality traits do influence the duration of unemployment spells in Germany (and also employment spells, but on a smaller level).

Other interesting dimensions of unemployment have been taken into account. For Spain, Bover et al. (2002) include business cycles in their analysis. Kong (2012) found that in China the reputation of the university that individuals graduate from has an influence on the hazard. For a closed down pulp in Sweden health issues have been taken into account (Bergström and Edin, 1992), which has also found application in Narendranathan and Stewart (1993). An interesting dimension of regional unemployment in Austria has been determined in including the level of public housing (Badinger and Url, 2002). The integration of informal employment in Mexico gave insights on an otherwise "invisible" labor market (Calderón-Madrid, 2008).

Most of the time results follow what theory suggests. It has been observed that especially old individuals suffer low employment hazards (Berenguer, 2001). Also married men have been found to face shorter stays in unemployment, which is even more pronounced if they have children (Calderón-Madrid, 2008). Another finding is the advantage of men in the labor market (Ciucă and Matei, 2011), but there is also literature concluding the opposite (Kavkler et al., 2009; Dănăcică and Babucea, 2010). The question of education has been taken into account in most analyses, but there is no agreement on the effect. Finally, unemployment benefits are found to have a negative impact on unemployment duration, i.e. they lengthen the spells (Van Ours and Vodopivec, 2006; Lalive et al., 2006; Lalive, 2008) and especially affect older individuals (Hunt, 1995).

1.5 Data

The data used for the analysis is a sample drawn from a database, which is a combination of register data from the Public Employment Agency (AMS) and employment data[5] from the database of the Austrian Social Security. The type of database is also known as linked employer-employee data[6] (LEED), because employment is defined by a link between employer and employee. The dataset depicts formal employment to the day, eliminating measurement errors, which could arise from working with survey data (Corak, 1996; Pyy-Martikainen and Rendtel, 2009). The observation period is chosen to be 2001-2012, but data go back to 1980s and some even farther. The choice of the time frame was due to availability of data for the unemployment rate in Austria.

[5]Only formal employment is included. Employers are obliged to report every employment to the Austrian Social Security.

[6]The linked employer-employee database is called AMDB and is provided by the Federal Ministry of Labor, Social Affairs and Consumer Protection (bmask) and the Public Employment Service Austria (AMS).

The structure of LEED data is as follows: The database links two entities –i.e. individuals (PENR) and firms/employers/Public Employment Service (BENR) to a relation, i.e. a spell. For each entity there are various information in the data. Table 1.2 gives an example of possible characteristics of the entities.

Table 1.2: Possible attributes of the entities (PENR and BENR) in the database

employee	**employer**
anonymous personal number (PENR)	anonymous employer number (BENR)
gender	region
age (through date of birth)	industry
citizenship	subsidized / not subsidized
number of children (only for women)	...
...	

The link of the two entities gives further information on the relation, e.g. the employment or unemployment spell. The relation is characterized by

- duration, through
    - spell start
    - spell end
- status of the spell (unemployment, employment, parental leave, retirement and others)
- spell location relative to the selected time frame, which is similar to the types of censoring in Figure 1.1:
    - A - starts before 2001 and reaches into the observation period
    - B - the whole spell lies within the time frame
    - C - the spell starts between 2001 and 2012 and is not completed at the end of 2012
    - D - the spell started before 2001 and is censored at the end of the observation period, i.e. it ends after 2012

For the empirical analysis only unemployment spells of the location B (completed) and C (right censored) will be of importance, since the focus is on individuals who transitioned to unemployment after the beginning of 2001.

The linked employer-employee database as a whole depicts the Austrian labor force very well and is thus a great tool for analyzing developments in the labor market. The labor market history of an individual (PENR) in the database is depicted as a number of consecutive, but not necessarily disjoint spells. These spells can either correspond to a link of an employer and an employee or to a match with the Austrian Public Employment Service. The latter is the case when either the individual is registered as unemployed or there is an administrative booking.

In reality there can be a vast amount of spells describing different statuses and there might be overlaps. The aspiration is to construct consecutive, non overlapping spells for the analysis. In this thesis an overlap of marginal employment and registered unemployment will be treated as unemployment. The idea behind this decision is, that individuals who were content with their job choice would not file for unemployment. There can also be overlaps of registered unemployment and administrative statuses, which are assigned to an employer, such as vacation benefits or resignation benefits. In this case the hierarchy design is harder to determine. It has been chosen to count these statuses as labor market active, because these benefits result from employment. Vacation benefits are a compensation for vacation that has not been taken during employment, i.e. a compensation for time matched to employment. When receiving resignation benefits the individual still receives his salary and is insured at the Austrian Social Security.

1.5.1 Data construction

The first step –before starting the empirical analysis– was to group spells into labor market active and non-labor market active spells. An example would be the match of sick leaves on the job to the corresponding employment spell. With a grouping algorithm it was possible to create continuous employment spells. In the second step, a spell that is the complement of employment –namely an *out-of-work* (OOW) spell has been created. Similar construction has been done for the analysis of the labor market in Spain (Berenguer, 2001). Up to that point, overlaps are still possible, since individuals can have two or more jobs or ongoing registered unemployment while being marginally employed. If there is an overlap of marginal employment and registered unemployment, it was decided that the spell will be counted as unemployment. Intuitively this makes sense, because marginal employment during registered unemployment is most likely not an individual's desired target job. Further information on the construction of out-of-work spells for data preparation will be given in the next subsection.

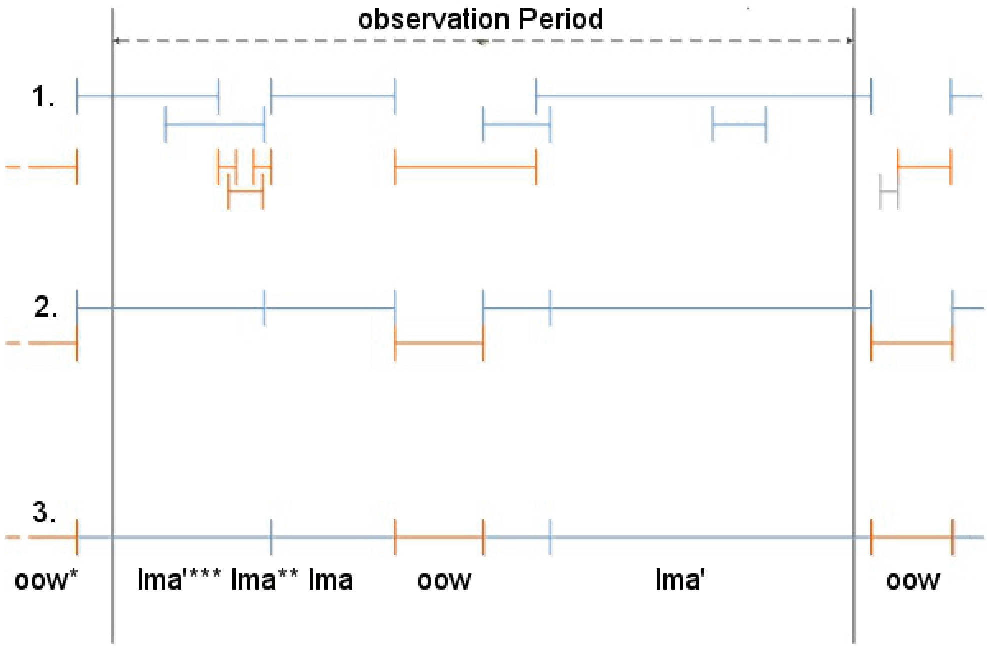
1.5.2 Out-of-work spells

Similar to Berenguer (2001), the focus lies on unemployment spells which follow an employment spell and end with an exit into employment. When looking at the data in their initial form, it becomes obvious that unemployment can be an on-off process without resulting into employment. There can be times where an individual is neither labor market active nor in the unemployment system. It has been found that the exit rate from the Austrian unemployment system differs vastly from the exit rate into employment, due to individuals leaving registration after cessation of benefits (Card et al., 2007). To solve this problem, the *out-of-work spell* is characterized by being the complement of employment with starting point one day after the end of the previous employment spell. Through overlap of employment spells, there also exist overlapping, identical OOW-spells. In the next step, the corresponding source episode (the previous employment) and target episode (the future employment) were assigned. For every source and target episode there is corresponding information on the two entities (see attributes of PENR, BENR) and also information resulting from the relation, e.g. size of the firm, number of employees. The target spell is unknown for censored spells.

Still, up to this point, the data can contain transitions from birth to first employment or from employment to retirement or death. For the construction of the sample this will be controlled for by two restrictions. First, the age group is restricted to individuals to individuals of working age (at the beginning of the unemployment spell), i.e. between 15 and 64 years old. Then, spells which do not lie in between two employments are dropped. Further restrictions can be found in Table A.1 in Appendix A.

All OOW-spells were created through an algorithm, which assignes a *begin date* one day after the end of an employment and the *end date* one day before the start of the future employment. Through this definition of dates there can also be nonreal out-of-work spells with negative duration, which represent side jobs or multiple employments.[7] These spells are not taken into account in the analysis. For this thesis, the out-of-work spells need to fulfill the criterion of containing at least one day of registered unemployment.

[7] An example: If an individual works two jobs, starting the same day, and one lasts longer, the algorithm defines the longer spell as target employment, since its end date is past the end date of the shorter employment spell. This leads to an out-of-work spell starting one day after the end of the shorter employment spell. The end date of the out-of-work spell would be day before the starting date of the longer employment spell, which is exactly one day before the starting point of the short employment spell. This leads to a spell with a negative duration equal to the length of the short employment spell.



*out of work **labor market active (single spell) ***labor market active (multiple spells)

Figure 1.3: Simplified figure of the construction process of out-of-work spells

Figure 1.3 should help in visualizing the spells in the data before, during and after editing the data. Followed by a more technical approach to the design just explained. The spell file is characterized by many different spells. The position of the spells is defined through the starting point and the end point recorded in the linked employer-employee database. In the first step in Figure 1.3, the spells are not disjoint and both employment spells (blue spells), unemployment spells and other out-of-work spells –e.g. parental leave, vocational training and sick leaves– (orange spells) are contained. In the editing process transitions from one spell to another are defined. For each PENR–BENR relation there is a source spell (former occupation) and a target spell (future job). The time period in between these employments is characterized as OOW-spell. The length of the time out-of-work is computed by

$$(begin\_date_{\text{target spell}}) - (end\_date_{\text{source spell}}) - 1.$$

Further, the employment spells are summarized, such that transitions from one employment to another are apparent –a labor market career of consecutive spells is created. The shape of the employment data is not important for the core analysis of this paper, but might be interesting for further analysis.

1.5.3 Description of covariates

This subsection is designed to give a short description of the covariates used for the empirical analysis. For some individuals there are missing values in one or more variables. Since the product limit estimator (Subsection 1.2.1) only compares the survival curves for categorical variables, the estimation could be done with a larger dataset. In order to keep the connection between the sample for nonparametric, semiparametric and parametric analysis, the observations which have at least one missing value are excluded. Through the LEED it is possible to assign certain characteristics, which may influence unemployment duration, to an individual.

gender: This variable is assigned to the PENR and gives the gender of an individual.

age: A categorical variable derived from the date of birth of an individual (PENR). It gives the age group an individual was in at the beginning of the spell. Categorization follows the grouping of STATISTICS AUSTRIA: *15-24, 25-34, 35-44, 45-54, 55-64.*

educational level: A categorical variable assigned to PENR summarized into five levels: *low edu* (compulsory education, no education), *appr* - apprenticeship, *med edu* - medium educational level (secondary school), *high edu* (higher education entrance qualification) and *uni* - university degree.

paid: A dummy variable equal to 1, if the individual receives benefit payments during the out-of-work spell and 0 otherwise.

child: A dummy variable, which is assigned to PENR and equal to 1, if the woman has at least one child at the beginning of the out-of-work spell, 0 otherwise. The number of children can only be assigned to women in the Social Security database through the number of live births. Nonparametric analysis includes a covariate indicating the numbers of children at the beginning of the spell, top-coded to two or more children.

urate: A covariate that controls for macroeconomic effects. It gives the unemployment rate for each age group. From 2000-2009 there only are annual data, whereas from 2010-2012 there are quarterly data. Figure A.1 in Appendix A gives a visualization of the unemployment rate in the observation period.

marginal: A dummy variable assigned to a relation of PENR and BENR. It equals 1 if the individual is marginally employed for at least one day and 0 otherwise.

prevep: Gives the length (in days) of the previous employment spell (assigned to a relation of PENR and BENR)

quarter: A categorical variable giving the quarter the OOW-spell started in. (assigned to a relation of PENR and BENR)

federal state: A categorical variable indicating the federal state the previous employer is registered in. *Bgld* (Burgendland), *Carinthia*, *Low Aus* (Lower Austria), *Up Aus* (Upper Austria), *Sbg* (Salzburg), *Styria*, *Tyrol*, *Vbg* (Vorarlberg) and *Vienna*. (assigned to BENR)

federal state-NUTS-1 level: A categorical variable used for nonparametric analysis, which is characterized following the NUTS-1 level of federal states with exclusion of Vienna. Values of the covariate are *East* (Burgenland, Lower Austria), *South* (Carinthia, Styria), *West* (Upper Austria, Salzburg, Tyrol, Vorarlberg) and *Vienna*. (assigned to BENR)

sector: A categorical variable used in nonparametric analysis describing the sector of economy the individual was previously working in. Its values are *Primary sector*, *Secondary sector* and *Tertiary sector*. (assigned to BENR)

industries: A categorical variable that includes a selection of industries and a residual category: *farm* - agriculture; forestry and fishing (primary sector); *manu* - manufacturing and *constr* - construction (secondary sector); *trade* - wholesale and retail trade, *tour* - accommodation and food service activities, *pub* - public administration and defense; compulsory social security and *soc* human work and social work activities (tertiary sector) and the residual category*other*. (assigned to BENR)

labor turnover: A categorical variable designed according to a measure of labor turnover. It was categorized by the ratio of different employed individuals per year-round employment. *low turn* - low labor turnover of less than 4 individuals per year-round employment (cumulated for the years 1997 to 2011) (B, C, D, F, G, H, K, O, Q, U), *med turn* - medium turnover between 4 and 6 individuals per year-round employment (E, J, L, M, S, T) and *high turn* - high labor turnover of over 6 individuals per cumulated year-round employment (A, I, N, P, R). The capital letters reflect industries according to the OENACE-classification (Table A.4).

industry size: A categorical variable giving the size of the previous industry. As the size of industries is very similar over the years, the industry size was determined by the number of different individuals in an industry in the year 2010. The categorization of industry size happened according to the number of employed individuals. (assigned to PENR and BENR).
Women: *small ind* (A, B, D, E, T, U), *med ind* (F, H, J, K, L, P, R, S), *big ind* (C, G, I, M, N, O, Q).
Men: *small ind* (A, B, D, E, L, R, T, U), *med ind* (I, J, K, M, P, Q, S), *big ind* (C, F, G, H, N, O).

number of seasonal peaks: A categorical variable which indicates the number of employment peaks per year in the previous industry of employment, not taking the single peak caused by internships during the summer months into account. The categorization has been done according to the course of employment throughout a year. Since the behavior is approximately the same over the years, the grouping was based on employment in the year 2010. Categories are *no peak* –a case when there seems to be no seasonal peak (B, C, D, E, G, J, L, M, N, O, P, S, T), but upward trends are still allowed, one peak (*1 peak*), which is usually during summer months (A, F, K, R) and two peaks per year (*2 peaks*) –one during winter and one during summer months (H, I, Q, U).

1.5.4 The problem of unobserved heterogeneity

There are many factors which can influence the out-of-work spell duration and cannot be observed when using linked employer-employee data. Some have been part of previous research, i.e. formal or informal employment, reason for the end of the previous employment, search strategies and marital status (Calderón-Madrid, 2008) or a more detailed covariate for educational level or school reputation (Kong, 2012). It has also been observed that personality traits may have an influence on unemployment duration (Uysal and Pohlmeier, 2011). Other unobservable factors influencing unemployment duration, such as attractiveness (Biddle and Hamermesh, 1998) or the name of an individual (Bursell, 2007; Brenner et al., 2008; Arai and Skogman Thoursie, 2009) could be of use in analyzing processes in the labor market.

Due to data privacy the name of an individual is not observed. Also attractiveness cannot be measured. In my opinion, not having information about marital status or number of children for men is the major shortcoming of the administrative data, as it has been shown in previous literature that these factors do matter (Lancaster, 1979; Lalive and Zweimüller, 2004; Calderón-Madrid, 2008). Additional important information, which cannot not be obtained is the type of employment (full or part time employment).

Taking these factors into account it can be assumed that there might still be unobserved heterogeneity present, leading to a possible downward bias or a negative bias on time dependence (Vermunt, 1996). Therefore results should be interpreted accordingly.

Chapter 2

Empirical analysis

In this chapter all empirical evidences are summarized. It is structured as follows: Section 2.1 deals with descriptive statistics of the population, followed by the construction and description of the sample used for the analysis. Section 2.2 will continue with descriptive analysis using the Kaplan-Meier estimator and conclude with a comparison of semiparametric and parametric analysis.

2.1 Descriptive statistics

This section is used to describe the population of out-of-work spells in the observation period prior to the application of sample restrictions. The steps to create the sample are explained and a comparable descriptive sample analysis is given.

2.1.1 Summary statistics for the population

The population contains all spells from January 1st of 2001 to December 31st of 2012, which are at least one day long and contain at least one day of registered unemployment. The following numbers in this subsection were extracted through a query on the database of all transition spells.

Before applying any restrictions –just taking all possible transitions[1] into account– there are 8,789,835 spells which belong to 2,428,411 individuals in the population. Out of these unemployment spells 39.9% belong to female individuals, whereas 60.1% are generated by male labor market participants.[2] Women create less unemployment spells

[1]Excluding marginal employment - as it is counted as unemployment.

[2]In the data there are four individuals for whom it was not possible to assign the gender. These individuals will be eliminated when further restrictions are applied to the data.

than men, i.e. 46.8% of the unemployment spells belong to women, respectively 53.2% belong to men. Table 2.1 gives the shares and absolute values of unemployment spells and unemployed individuals by grouped citizenships.

Table 2.1: Absolute and relative values for the distribution of out-of-work spells and unemployed individuals by grouped citizenship

	Unemployment spells		**Unemployed individuals**	
Citizenship	**absolute**	**percent**	**absolute**	**percent**
Austrian	6,848,948	77.9%	1,952,559	78.2%
Third Country	1,411,385	16.1%	365,803	14.7%
EU before 2004	255,177	2.9%	86,188	3.5%
EU 2004	274,311	3.1%	92,194	3.7%
Total**	**8,789,821**	**100%**	**2,496,744***	**100%**

*Some individuals appear in more than one group, because of a change of citizenship. In order to appear in two categories at least two spells must have been observed.
**For 14 spells/nine individuals it was not possible to assign a citizenship.

Table 2.1 shows that more than three quarters of the individuals unemployed within the observation period had the Austrian citizenship at the beginning of their unemployment spell. Table 2.2 gives the number of spells and individuals per starting year.

Table 2.2: Absolute and relative values for the distribution of out-of-work spells and unemployed individuals by starting year

	Unemployment spells		**Unemployed individuals**	
Year	**absolute**	**percent**	**absolute**	**percent**
2001	694,649	7.9%	528,738	8.0%
2002	724,981	8.2%	551,837	8.4%
2003	731,809	8.3%	553,190	8.4%
2004	706,591	8.0%	537,583	8.2%
2005	720,543	8.2%	543,458	8.2%
2006	709,954	8.1%	536,130	8.1%
2007	705,278	8.0%	528,931	8.0%
2008	745,823	8.5%	553,424	8.4%
2009	799,077	9.1%	596,802	9.1%
2010	765,996	8.7%	565,924	8.6%
2011	746,616	8.5%	550,804	8.4%
2012	738,518	8.4%	545,895	8.3%
Total	**8,789,835**	**100%**	**6,592,716**	**100%**

It can be observed that the share and the number of unemployed individuals and out-of-work spells were approximately the same over the years. There is a peak in 2009 (starting in 2008), which might be due to the financial crisis in 2008. The Annual Economic Report of the Federal Ministry of Economy, Family and Youth (2010, p.18) stated that about 50,000 jobs were cut back until mid-2009 and that the number of jobs grew slowly in the second half of the year.

2.1.2 Editing the data – Sample construction

A subset of the population described above is the foundation of the analysis in this thesis. Before applying restrictions the drawn sample counts 105,757 transition spells, which are created by 101,143 individuals. Following the structure of Subsection 2.1.1, the share of women corresponds to 39.8% (recall 39.9% for the whole population) and to 60.2% (60.1%) for men. Table 2.3 depicts the number and shares of individuals with different citizenships in the sample. It shows that the distribution of citizenship is also very similar to the population sample from 2001-2012 in Table 2.1.

Table 2.3: Absolute and relative values for the distribution of out-of-work spells and unemployed individuals in the sample by grouped citizenship

	Unemployment spells		**Unemployed individuals**	
Citizenship	**absolute**	**percent**	**absolute**	**percent**
Austrian	82,532	78.0%	78,907	78.0%
Third Country	16,867	16.0%	16,093	15.9%
EU before 2004	3,099	2.9%	2,991	3.0%
EU 2004	3,258	3.1%	3,151	3.1%
Total*	**105,756**	**100.0%**	**101,142**	**100.0%**

*For one spell/individual it was not possible to assign a citizenship.

The first step is the elimination of individuals who exit into retirement. 3,769 spells (about 4%) in the sample correspond to leaving the labor market for retirement. 130 individuals were dropped from the sample, because unfortunately they died within the observed time frame. This leads to 101,858 spells, which represent individuals who participated in the labor market prior to their unemployment spell and participated again later or were censored at the end of the observation period. In the database there is an administrative process, which assigns a labor market active status to individuals who undergo a rehabilitation process. These spells are also excluded, since administrative transitions are not of interest for the analysis.

To prevent a bias from incomparable spells, the following restrictions have been made: Spells during which individuals exit into training or into parental leave will be ruled out, because it could prolong the stay out of work. During these statuses individuals can be assumed not to be looking (as) actively for employment. Elimination leaves us with 101,173 spells.

For the question of what defines a successful transition to work, I imposed the restriction that the target employment must be active for at least 30 days.[3] The same restriction counts for source employment. 34,543 spells (about 34%) of the remaining sample do not fullfil this criterion.

Second to last step is to select the cases which fit the description of the desired sample. For this thesis I decided to focus on *first spells*[4] of unemployment (OOW) which belong to Austrian individuals of working age whose spells are longer than one week.[5] Up to this point there are 49,775 spells left in the sample, which now is per definition equal to the number of individuals.

The final restriction is the elimination of spells with missing values, leading to a sample of 39,208 spells/individuals. A share of 41.5% belongs to women and 58.5% to male unemployed labor market participants. Table A.1 in Appendix A summarizes the elimination process.

2.1.3 Descriptive analysis of the sample

To get a grasp of the distribution of characteristics in the two samples, this subsection will focus on descriptive statistics. Table 2.4 gives the absolute and relative numbers of educational level in the sample. It shows that the biggest shares of unemployment spells belong to individuals who completed an apprenticeship or the lowest level of education (completed or not completed compulsory education). For both women and men the fewest spells are generated by individuals having completed the highest level of education (university, academy or university of applied sciences). Compared to the population the share of the apprenticeship has increased, whereas the share of men

[3]Censored spells do not have to fulfill this criterion.

[4]The first spell occurring in the observation period. Taking the first spells within an observation period is a mechanism of reducing multiple spells, which would be equal to choosing one spell per individual at random. Since most of the individuals (97.36%) do not experience more than one OOW-spell I decided to take the first spell. The first spell does *not* mean there need to have been no transition spells before.

[5]This restriction was chosen, because unemployment shorter than one week is hard to interpret and is also an unrealistic case.

and women with compulsory education has decreased. A description of all possible educational levels can be found in Table A.3 in Appendix A.

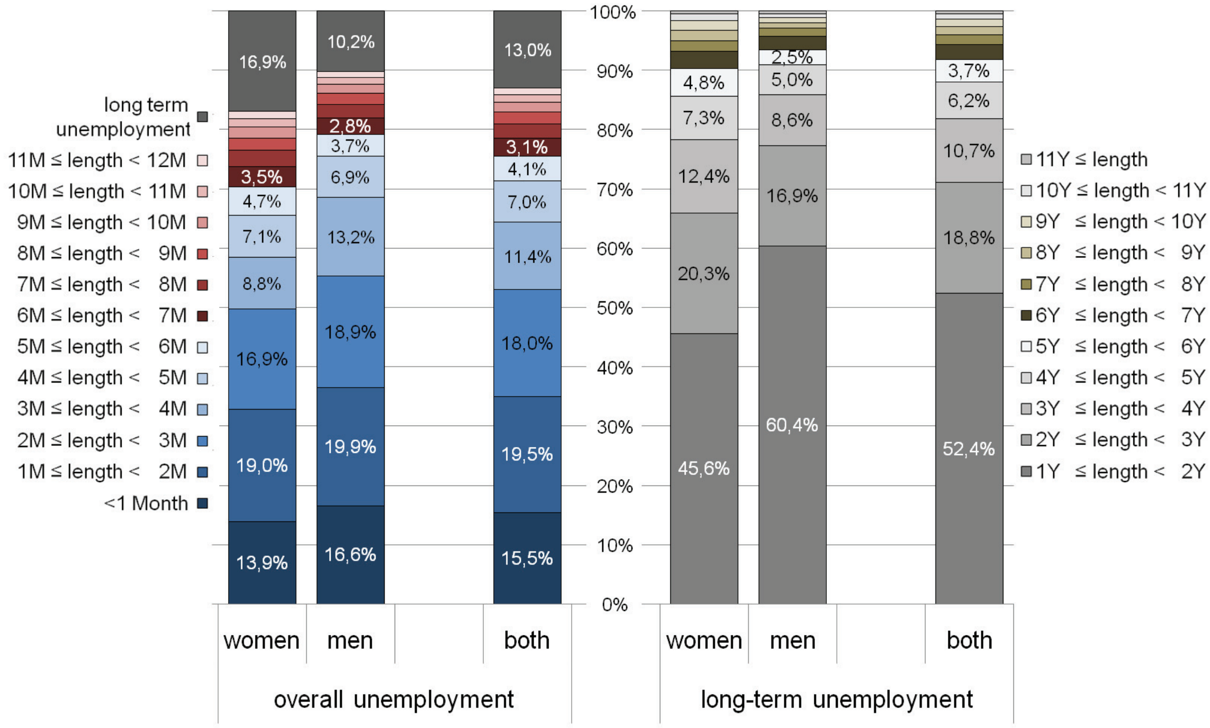
Table 2.4: Absolute and relative values for the distribution of educational level in the final samples for women and men

	Women			Men		
Education	**Freq.**	**Percent**	**Cum.**	**Freq.**	**Percent**	**Cum.**
low level	6,338	38.99%	38.99%	7,343	31.99%	31.99%
apprenticeship	5,575	34.30%	73.29%	12,760	55.59%	87.58%
medium level	1,710	10.52%	83.81%	822	3.58%	91.16%
high level	1,893	11.65%	95.45%	1,497	6.52%	97.69%
university	739	4.55%	100.00%	531	2.31%	100.00%
Total	**16,255**	**100.00%**		**22,953**	**100.00%**	

Since unemployment duration is the variable of interest, Figures 2.1 and 2.2 depict the share of unemployment durations. Figure 2.1 will give an overview of the spell durations by gender. Afterwards the difference between censored and completed spells is visualized in Figure 2.2.

The left hand side of Figure 2.1 gives the shares of durations of out-of-work spells in months (colored) and long-term unemployment, which is defined as unemployment longer than 365 days (gray). It can be observed that men in the sample have a bigger share of unemployment spells which are shorter than one year. Especially the unemployment spells shorter than four months are more important in the sample of male labor market participants. The different shades of gray on the right hand side of Figure 2.1 correspond to the gray area in the bars on the left hand side (long-term unemployment). The figure shows that among long-term OOW-spells the share of spells between one and two years is the most prominent.

Figure 2.1 counts censored spells as completed –which has been found to cause a bias (Corak, 1996). Therefore Figure 2.2 describes the differences in spell duration of completed and censored spells for men and women. For men and women combined completed spells shorter than one year comprise 89.7% of the sample, whereas for censored spells the share is just 50.3%. Similar to the findings of Figure 2.1, men have bigger shares of spells shorter than four months compared to women. It can also be obtained that the (minimum) duration of censored spells is on average longer than the duration of completed spells. This makes sense, because for censored spells it is not possible to tell the individuals who dropped out of the labor market from those still

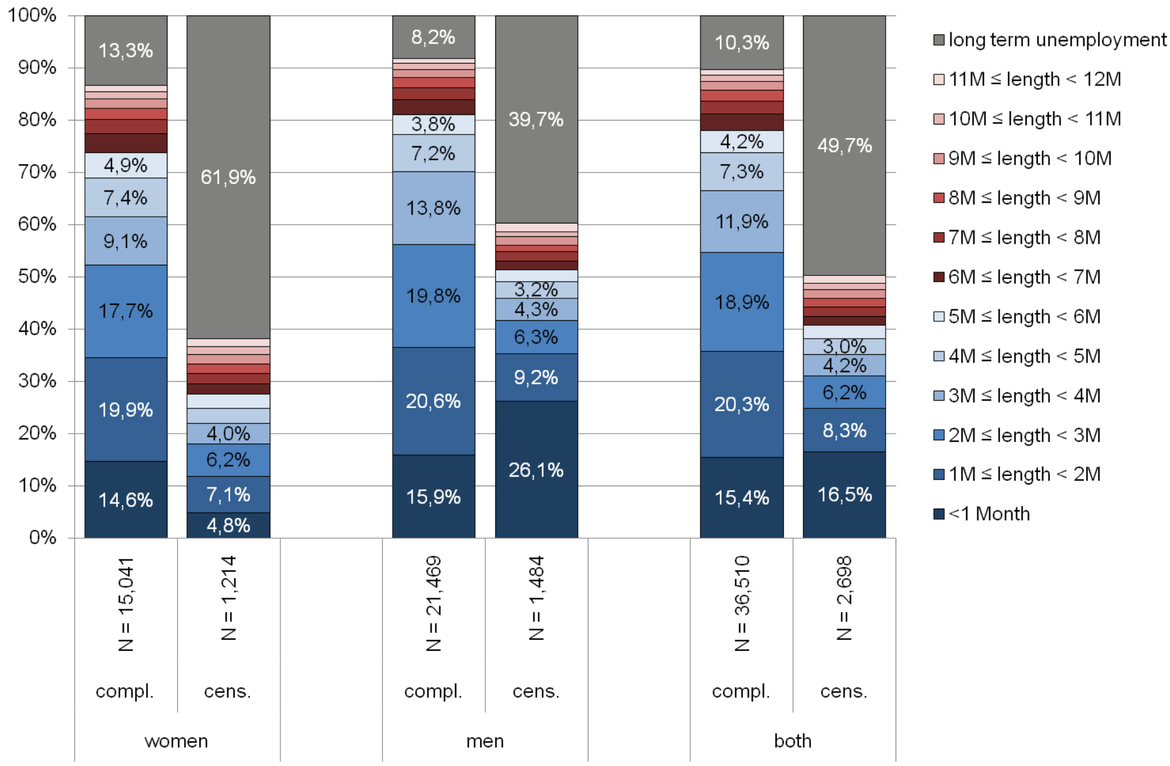


The bar chart on the left gives the overall distribution of spells (one-month intervals). The blue areas depict the spells where a job has been found in less than six months and the red area pictures spells that ended after six months to under one year.
The bar chart on the right gives the distribution of spell lengths for long-term unemployment –i.e. spells that are longer than 12 months.
The transformation into months has been made by dividing the spell length in days by 30.4.

Figure 2.1: Overall distribution of spell lengths (left hand side) and distribution of long-term unemployment spells (right hand side)

going to be employed in the future. The combination of Tables 2.1 and 2.2 shows that counting censored spells as completed leads to an increase in the share of long-term unemployment.

As my research questions focus on industy-specific characteristics, it is important to know about the "dynamics" in the labor market. Tables 2.5 and 2.6 give the numbers of the individuals, who make a transition from one source industry (rows) to the target industry (columns), according to the NACE structure (see Table A.4 in Appendix A). The gray boxes give individuals who stay in the industry and the orange boxes give the number of censored spells within an industry. It is shown that the majority of the individuals stay in the source industry. Especially the individuals in 'construction' (F), 'mining and quarrying' (B) and 'accommodation and food service activities' (I) have a huge amount of inter-industry transitions. Reasons for this can be the recall by



Unemployment duration for censored spells only depict a *minimum* duration.

Figure 2.2: Overall distribution of the length of completed and censored OOW-spells by gender

the previous employer and the low employment requirements. It can also be obtained that most of the individuals without employment have previously worked in the industries C –'manufacturing', G –'wholesale and retail trade', I –'accommodation and food service activities' and N –'administrative and support service activities'. These industries are also among the biggest industries according to the number of employed individuals. The numbers for men and women can be found in Tables A.5, A.6, A.7 and A.8 in Appendix A. Appendix D includes Kaplan-Meier estimates for the survivor function of individuals who changed industry or employer. These outcomes will not be part of parametric analysis, since values can only be determined for individuals with completed spells.

Another interesting aspect is to know how many individuals received benefit payments for at least one day during their out-of-work spell. Data in Table 2.7 show that this is the case for about 95% of individuals. Table 2.8 shows that marginal employment of at least one day during the out-of-work spell is only present for about 13% of individuals.

Table 2.5: Numbers of changes from source to target industries (OENACE classification), Part 1

	Target industry												
Source industry	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**Total**
A	**233**	0	14	0	3	12	21	21	13	2	0	0	**400**
B	0	**117**	5	0	1	8	6	7	1	0	0	1	**152**
C	16	7	**1,911**	6	13	189	414	111	156	28	30	17	**4,214**
D	0	0	2	**9**	0	3	2	3	0	2	1	0	**43**
E	1	0	8	0	**32**	11	8	5	2	0	0	1	**96**
F	9	6	249	1	12	**5,969**	192	156	65	9	11	17	**7,596**
G	27	5	417	5	11	162	**2,534**	158	292	74	58	36	**5,575**
H	2	5	56	1	2	66	144	**1,025**	56	8	7	12	**1,774**
I	17	3	181	2	3	64	396	147	**5,817**	16	23	24	**7,819**
J	1	0	27	1	1	4	55	11	13	**157**	5	4	**556**
K	0	0	21	2	0	8	45	12	16	8	**76**	6	**363**
L	1	0	17	0	0	17	27	10	22	2	7	**80**	**293**
M	2	1	81	1	1	60	102	25	36	37	30	20	**1,170**
N	10	2	270	2	5	190	285	118	154	36	22	21	**3,564**
O	5	1	73	0	3	43	107	25	60	15	23	13	**1,671**
P	0	0	28	1	1	30	48	10	54	1	4	4	**611**
Q	7	0	77	0	3	44	141	26	65	14	9	8	**1,721**
R	2	0	19	1	0	5	34	17	42	6	6	1	**516**
S	3	0	34	0	4	25	114	16	82	8	5	3	**1,034**
T	0	0	2	0	0	0	2	0	5	0	0	1	**35**
U	0	0	0	0	0	0	0	0	0	0	0	0	**5**
Total	**336**	**147**	**3,492**	**32**	**95**	**6,910**	**4,677**	**1,903**	**6,951**	**423**	**317**	**269**	**39,208**

Table 2.6: Numbers of changes from source to target industries (OENACE classification), Part 2

	Target industry											
Source industry	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**cens**	**resid**	**Total**
A	6	17	9	3	5	2	5	0	0	**25**	9	**400**
B	1	0	1	1	1	1	0	0	0	**1**	0	**152**
C	103	444	133	30	158	22	43	3	0	**242**	138	**4,214**
D	4	6	4	1	1	0	0	0	0	**5**	0	**43**
E	4	13	2	0	1	0	1	0	0	**5**	2	**96**
F	50	294	53	29	55	11	17	1	0	**273**	117	**7,596**
G	111	469	178	59	219	42	108	1	0	**403**	206	**5,575**
H	26	120	31	11	23	5	15	0	0	**119**	40	**1,774**
I	60	237	90	76	135	62	78	3	0	**256**	129	**7,819**
J	50	55	20	9	18	12	21	0	0	**44**	48	**556**
K	15	35	18	3	14	6	9	0	0	**31**	38	**363**
L	9	28	12	2	14	1	7	0	0	**20**	17	**293**
M	**374**	81	49	30	48	12	31	2	1	**72**	74	**1,170**
N	75	**1,658**	84	31	127	27	43	3	0	**338**	63	**3,564**
O	57	103	**725**	43	183	11	52	2	0	**88**	39	**1,671**
P	16	42	32	**207**	40	12	16	0	0	**39**	26	**611**
Q	32	142	125	29	**683**	10	50	3	0	**208**	45	**1,721**
R	17	20	16	14	8	**249**	9	0	0	**35**	15	**516**
S	28	47	66	19	59	12	**403**	0	0	**66**	40	**1,034**
T	0	1	0	0	5	0	0	**13**	0	**3**	3	**35**
U	0	0	1	0	1	0	1	0	**1**	**0**	1	**5**
Total	**1,038**	**3,812**	**1,649**	**597**	**1,798**	**497**	**909**	**31**	**2**	**2,273**	**1,050**	**39,208**

Table 2.7: Absolute and relative values for the distribution of receipt of benefit payments by gender

	Women		Men		Total	
paid	**absolute**	**percent**	**absolute**	**percent**	**absolute**	**percent**
0	994	6.12	953	4.15	1,947	4.97
1	15,261	93.88	22,000	95.85	37,261	95.03
Total	**16,255**	**100.00**	**29,534**	**100.00**	**39,208**	**100.00**

Table 2.8: Absolute and relative values for the distribution of marginal employment by gender

	Women		Men		Total	
marginal	**absolute**	**percent**	**absolute**	**percent**	**absolute**	**percent**
0	13,525	83.21	20,630	89.88	34,155	87.11
1	2,730	16.79	2,323	10.12	5,053	12,89
Total	**16,255**	**100.00**	**22,953**	**100.00**	**39,208**	**100.00**

2.2 Results

In this thesis male and female individuals will be analyzed separately. There are 16,255 women out of work, of which 15,041 exit into employment. For men 21,469 out of 22,953 unemployment spells result in employment. The maximum spell duration is 4,383 days (the whole time period from 2001-2012). Only 7.5% of the spells for women and 6.5% of the spells for men are censored.

The first subsection gives the results of the Kaplan-Meier estimator for the categorical variables. In Subsection 2.2.2 the results of the semiparametric and parametric analysis will be compared.

2.2.1 Nonparametric estimates

The following subsection provides the Kaplan-Meier survival estimates by the categorical variables specified in Subsection 1.5.3. Logrank tests for equality of survival curves were used to determine whether the depicted differences are significant. For all covariates the tests were significant, rejecting the null hypothesis of equality (see Appendix B).

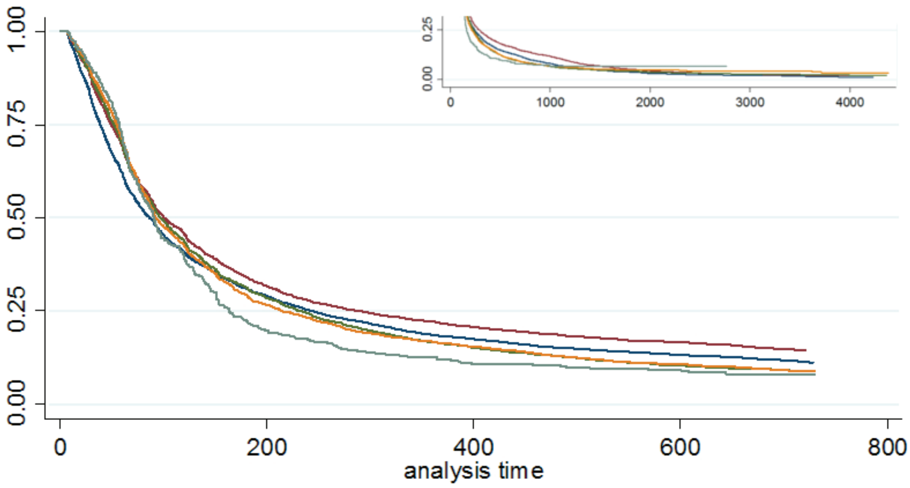
Recall that the survivor rate function is a nonincreasing monotonic function. All figures (except for Figure 2.5, which shows the Nelson-Aalen smoothed hazard estimate) show the estimated survival curves, zoomed in on the time frame of two years (730 days). In the top right corner the estimated survival curves from zero to 25% for the whole observation period of 4,383 days are included. The short horizontal line at the beginning of the survival estimates is caused by the sample selection of spells which are longer than seven days. Individuals, whose completed spell length is not known, i.e. censored spells, do not have to fulfill this restriction and are thus included in the the horizontal line. The y-axis gives the probability of surviving beyond a specific spell length and the x-axis gives the duration of the OOW-spell. In Figure 2.5 the y-axis indicates the hazard rate and the x-axis the duration of the OOW-spell.

Differences between age groups

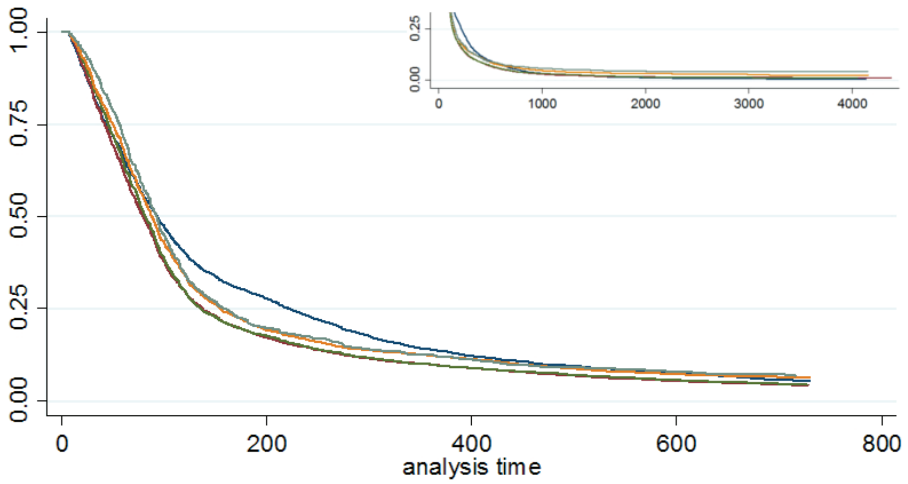
Figure 2.3 describes the differences among age groups for women (Figure 2.3a) and men (Figure 2.3b). The age groups to be compared are *15-24* (blue), *25-34* (red), *35-44* (green), *45-54* (yellow) and *55-64*-year-olds (turquoise). The age group *55-64* represents only 2,7% (4,7%) of the spells of women (men). For women 11.5% of the spells of the age group 55-64 are censored at the end of the observation period. Over all age groups the censored spells equal 6.6%. Within the male sample the share of censored spells belonging to individuals over 55 is equal to 12.5% compared to 5.1% for the whole sample. Interpretation can therefore be a bit troublesome.

First, it can be observed is that men tend to leave unemployment faster, i.e. men's survival curves drop faster and are closer to zero. After the first month, 82% (84%) of women (men) of the age group 15-24 were still out of work. The survivor rate was 87% (83%) for individuals aged 25-34, respectively 88% (86%) for individuals aged 35-44 and over 89% (87%) for individuals older than 45.

Furthermore, the ordering of survival curves differ among gender. Figures 2.3a and 2.3b show that especially individuals at the age of 25 to 34 (red) have different probabilities of staying out of work. In the top right corner of Figure 2.3a the red curve is situated above all other curves from about 80 to 1,400 days. When looking at men (Figure 2.3b), the age group of 25-34-year-olds behaves approximately the same as the age group of 35-44-year-olds (green). After one year 22% of women of age group 25-34 would still be unemployed, compared to only about 10% of men. The survival curve for men aged 15-24 (blue) shows that after 95 days to about 500 days the probability of staying out of work is higher than for all other age groups of men. The youngest age



(a) women, 2 years



(b) men, 2 years

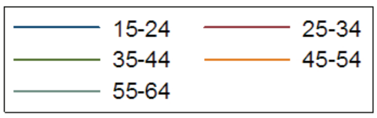


Figure 2.3: Kaplan-Meier survival estimates by age groups

group of women does not face this problem. Within the first 90 days of unemployment (approximately), young women tend to exit unemployment faster than all other age groups.

The visualization of survival curves shows that men over 55 have the highest probability of staying out of work within the first 13 weeks, while young men aged 25-44 have a low probability. Women older than 55 have the highest probability of staying unemployed at first, then this age group reaches the lowest probability of survival (at about 17 weeks) in the time frame of two years.

Most literature find that age has a negative effect on finding employment (Narendranathan and Stewart, 1993; Berenguer, 2001; Calderón-Madrid, 2008; Boršic and Kavkler, 2009).

Differences in the receipt of benefits

Figure 2.4 compares the differences in survival for individuals who *receive benefit payments* (red) with the individuals who *do not receive payments* (blue). These payments can also be interpreted as some sort of additional income. The covariate *paid* only indicates that some kind of benefit payment was received and there is no information on how long the benefits have been paid.

Theory suggests that individuals are utility maximizers and thus weigh the benefits and costs of having a job to being out of work. Individuals receiving unemployment benefits –an additional income– are therefore assumed to have lower incentives to search for a job (higher reservation wage, lower costs of unemployment), but the incentive should increase closer to the date of unemployment benefit expiration (Lalive et al., 2006). Taking theory into account individuals are expected to have a higher probability of survival before expiration eligibility to payments.

In Figure 2.4 the course of survivor functions for men and women during the entire observation period (top right corner) reflects the overall picture that women stay longer out of work than men. In the limited time frame it can be seen that curves are steeper for men at the beginning. Apart from these differences, the behavior of individuals with (red curve) and without benefit receipt (blue curve) are very similar for men and women. At first, the blue curve is steeper, i.e. for unemployed workers who do not receive benefits the probability of staying out of work drops faster than for benefit recipients. After one month of unemployment the probability of staying out of work is twelve (16) percentage points lower for women (men) who do not receive benefit payments. At around 12 weeks (women), respectively around 12,5 weeks (men), the curves intersect. At that point in time 55% of the women in the sample are still out of work, compared to only about 48% of men. As theory suggests, this finding might indicate that individuals who have an income during unemployment have lower incentives to actively search for a job as long as benefits are received. Recall Section 1.3, where it has been described that individuals who fulfill the criteria are entitled to up to 20 weeks of unemployment benefits until they find employment or refuse to take a job.

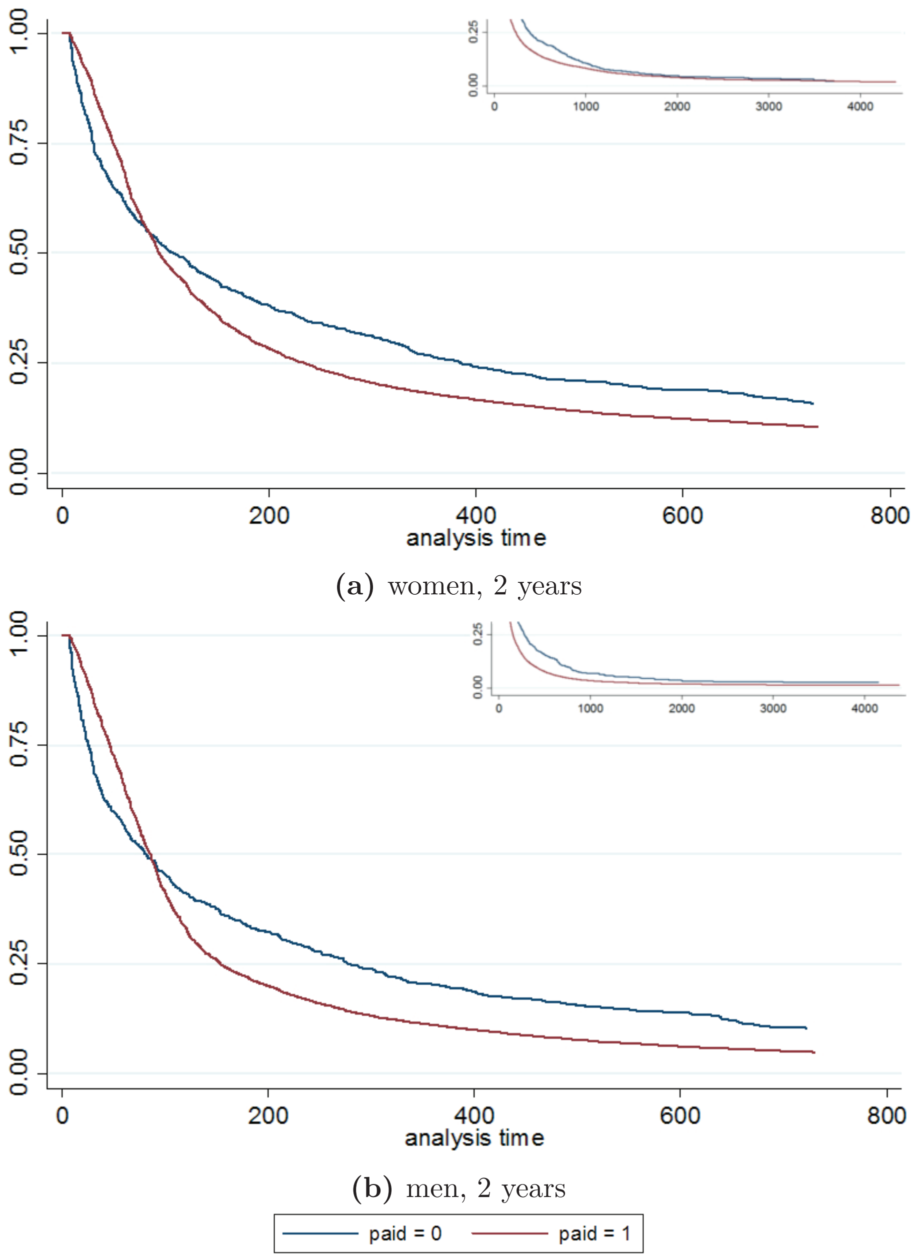
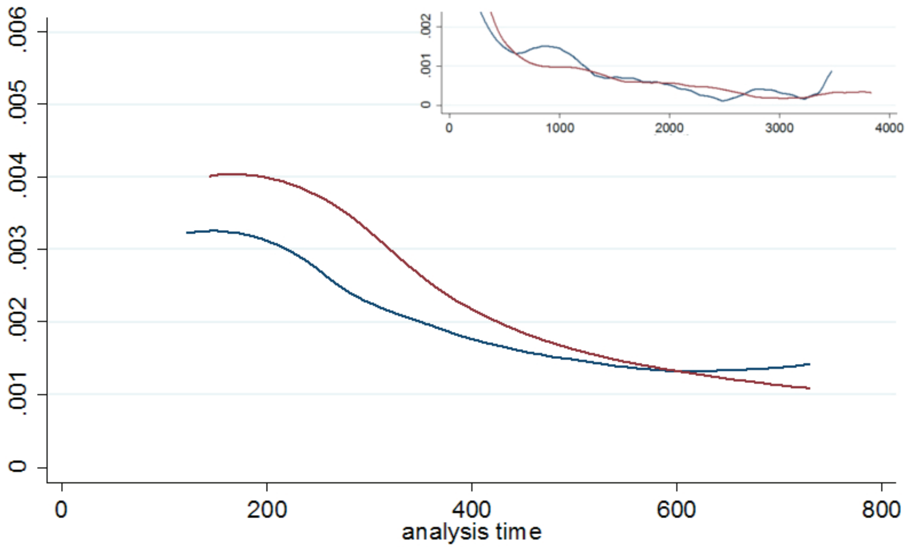
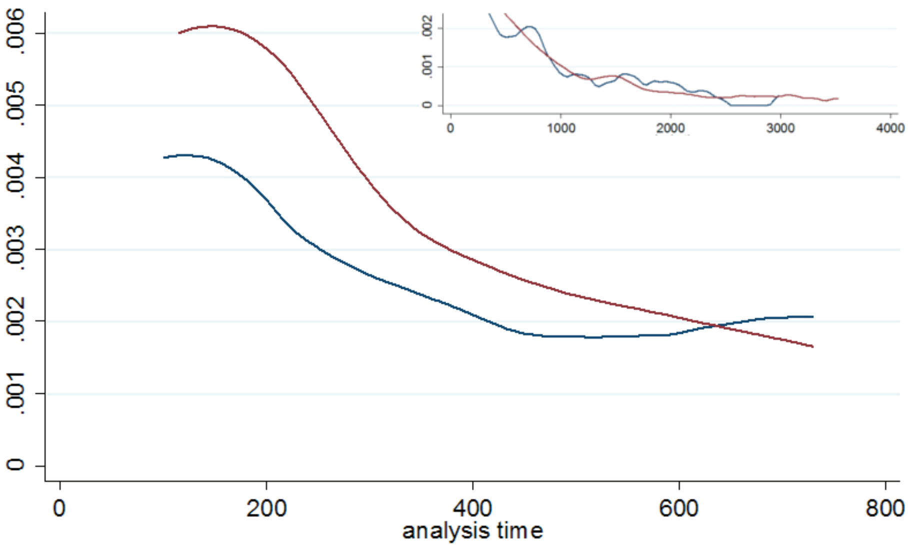


Figure 2.4: Kaplan-Meier survival estimates by receipt of benefits

It is necessary to take the hazard rate into account to examine whether individuals face different probabilities of finding a job . The Nelson-Aalen estimator, introduced in Subsection 1.2.1, is used to estimate the hazard rate function for individuals who receive benefit payments and for individuals who do not. When taking a look at the smoothed hazard estimates in Figure 2.5, it can be obtained that the hazard rate is still slightly higher for benefit recipients for about 600 days for women and 650 days for men. Within this time frame the hazard rate seems to differ only by at most 0.1% for women and by 0.2% for men.



(a) women, 2 years



(b) men, 2 years

paid = 0   paid = 1

Figure 2.5: Smoothed hazard estimates by receipt of benefits

The findings from this subsection lead to the conclusion that unemployment benefit payments or unemployment assistance during out-of-work spells might have a negative impact on job search, leading to longer spells.

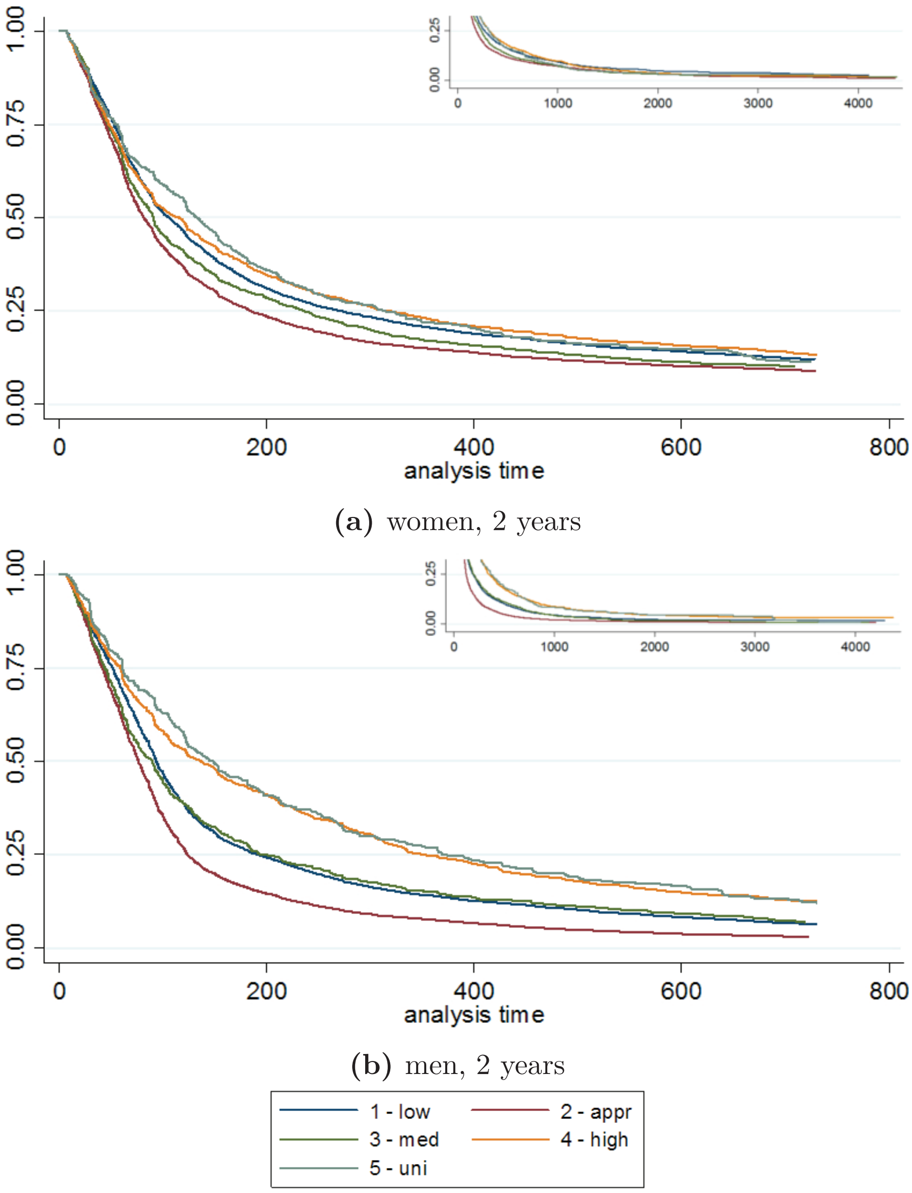
The effect of benefit payments on unemployment duration has been discussed in economic literature (Katz, 1986; Bover et al., 2002; Van Ours and Vodopivec, 2006). The literature with focus on the effects of Austrian REBP (Lalive and Zweimüller, 2004; Lalive, 2008) found prolonging effects of increased duration of eligibility.

Differences among educational levels

Figure 2.6 compares survival estimates across different educational levels. The blue curve represents the *lowest educational level* (compulsory education and no education), the red curve depicts a completed *apprenticeship*. The survival curve for *medium educational level* (secondary school) is colored green, for *high educational attainment* (higher education entrance qualification) the color yellow and for *university degrees* the color turquoise is chosen.

It becomes obvious that the level of highest completed educational level causes more differences in out-of-work spells for men, i.e. the estimeted survival curves are farther apart. Additionally, in the sample of male labor market participants, the survival curves of medium and low educational attainment and high level of education and university degrees (or equivalents) are clustered. This indicates that the difference of having completed a university degree or a higher education does not make visible difference in the duration of unemployment for men, as it does not for low and medium educational attainment. This phenomenon does not exist for women and might imply that educational levels are more important for men when it comes to finding employment. Reasons for this could be that women might not look for jobs which require certain educational levels as often as men do. Considering the Tables A.5 and A.6 in Appendix A, which list the source and target industries for women, one can see that about 31% of the spells are in the industry 'accommodation and food service activities' (tourism). This industry is known to have low educational requirements.

For both men and women, having completed an apprenticeship (red curve) is an advantage in finding employment. This is also the most common educational level across unemployment spells (Table 2.4). After one year of unemployment only about 15% (7.5%) of the women (men) who completed an apprenticeship are still without employment. High educational levels (yellow and turquoise) seem to have a high probability of staying out of work, i.e. the survival curve is above all other curves. After six months of unemployment about 44% of highly educated men are still out of work compared to 35% of women in the sample. After one year about 23% (25%) of highly educated women (men) are still looking for a job. The fact that higher educational levels involve no improvements in the probability of staying out of work could be due to a selection bias, i.e. highly educated individuals do not enter unemployment as often as individuals with other educational attainments do. The highest educational level only represents less than 5% of the sample.



(a) women, 2 years

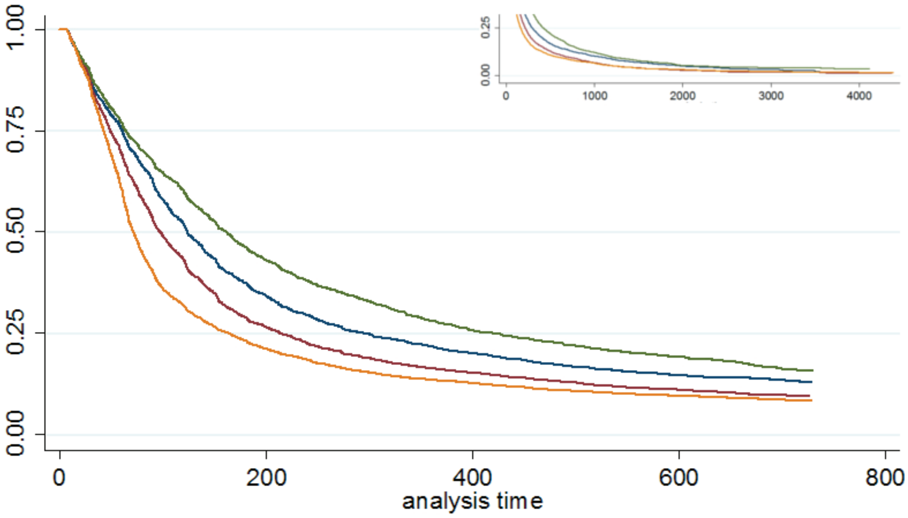
(b) men, 2 years

Figure 2.6: Kaplan-Meier survival estimates by highest level of completed education

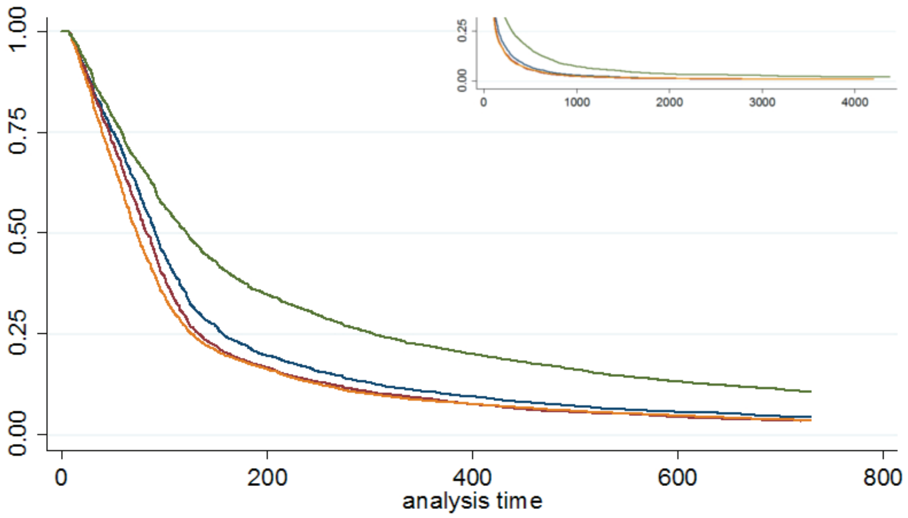
Similar results have been found for the Romanian labor market (Ciucă and Matei, 2011). For the REBP recipients in Austria (Lalive and Zweimüller, 2004) an analogous effect of educational level on unemployment duration has been observed. The authors claim it could be due to the fact that highly educated individuals are more specialized and have therefore a lower job offer probability.

Differences across federal state of previous employment

In this subsection focus lies on federal states of previous employment. Figure D.3 in Appendix D gives a comparison of all federal states. In the following, the federal state of previous employment –grouped according to the first-level NUTS classification with



(a) women, 2 years



(b) men, 2 years

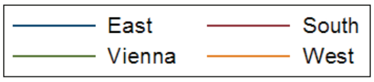
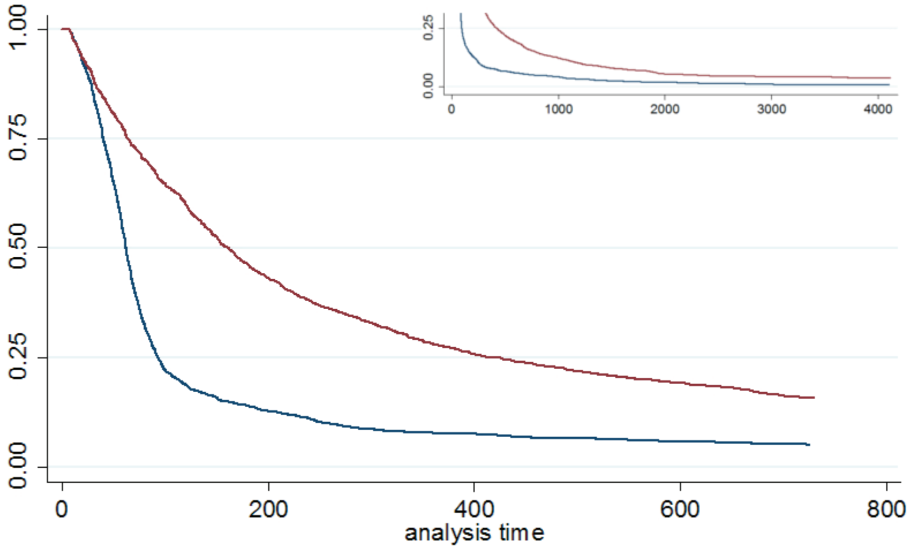


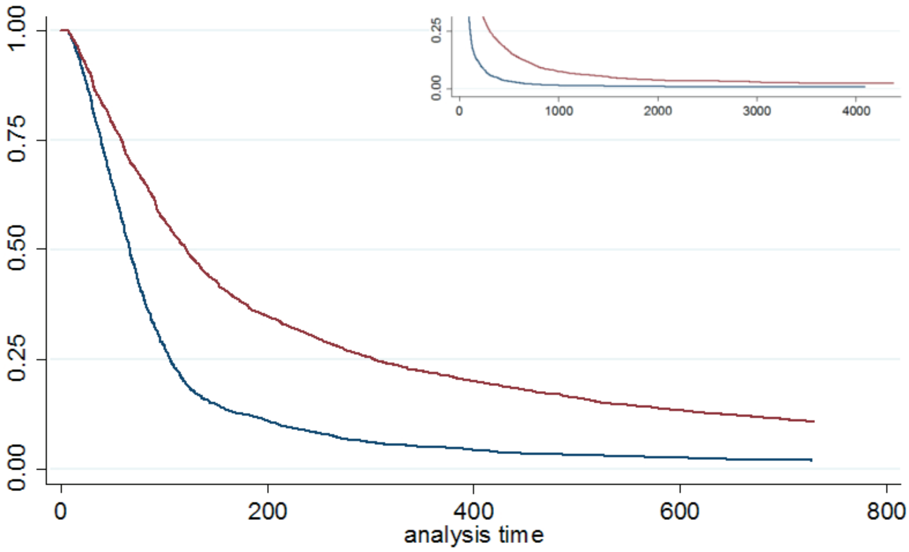
Figure 2.7: Kaplan-Meier survival estimates by federal state according to NUTS-1 classification

exclusion of Vienna– is analyzed. The federal states are divided into three groups. *East Austria* (Burgenland, Lower Austria), *Vienna* , *South Austria* (Carinthia, Styria) and *West Austria* (Upper Austria, Salzburg, Tyrol, Vorarlberg).

First thing to be noticed is that for men the survivor curves of the federal states (except from Vienna) lie closer together, which might be due to differences in employment choices by gender. Excluding Vienna, the other first-level NUTS groups differ in the importance of tourism. Since a large share women previously worked in the



(a) women, 2 years



(b) men, 2 years

Figure 2.8: Kaplan-Meier survival estimates by federal states Tyrol and Vienna

industry ‘accommodation and food service activities’ (which is linked to tourism), i.e. 31%, the lines might lie farther apart.

Both Figures 2.7a and 2.7b indicate that West Austria has the lowest survival curves, followed by South and East Austria. Vienna has by far the highest probability of staying out of work. After one year about 14% (8%) of women (men) in West Austria are still unemployed, respectively 16% (9%) of women (men) having previously worked in South Austria, 22% (11%) of East Austria and 28% (22%) from Vienna.

When comparing the survival curves for every single federal state (see D.3 in Appendix D), the most interesting finding is that for both men and women OOW-spells assigned to Tyrol have the lowest survival curves, whereas it is highest for Vienna. Figure 2.8 gives the survival curves for these two extremes, where the blue curve depicts Tyrol and the red curve corresponds to Vienna.

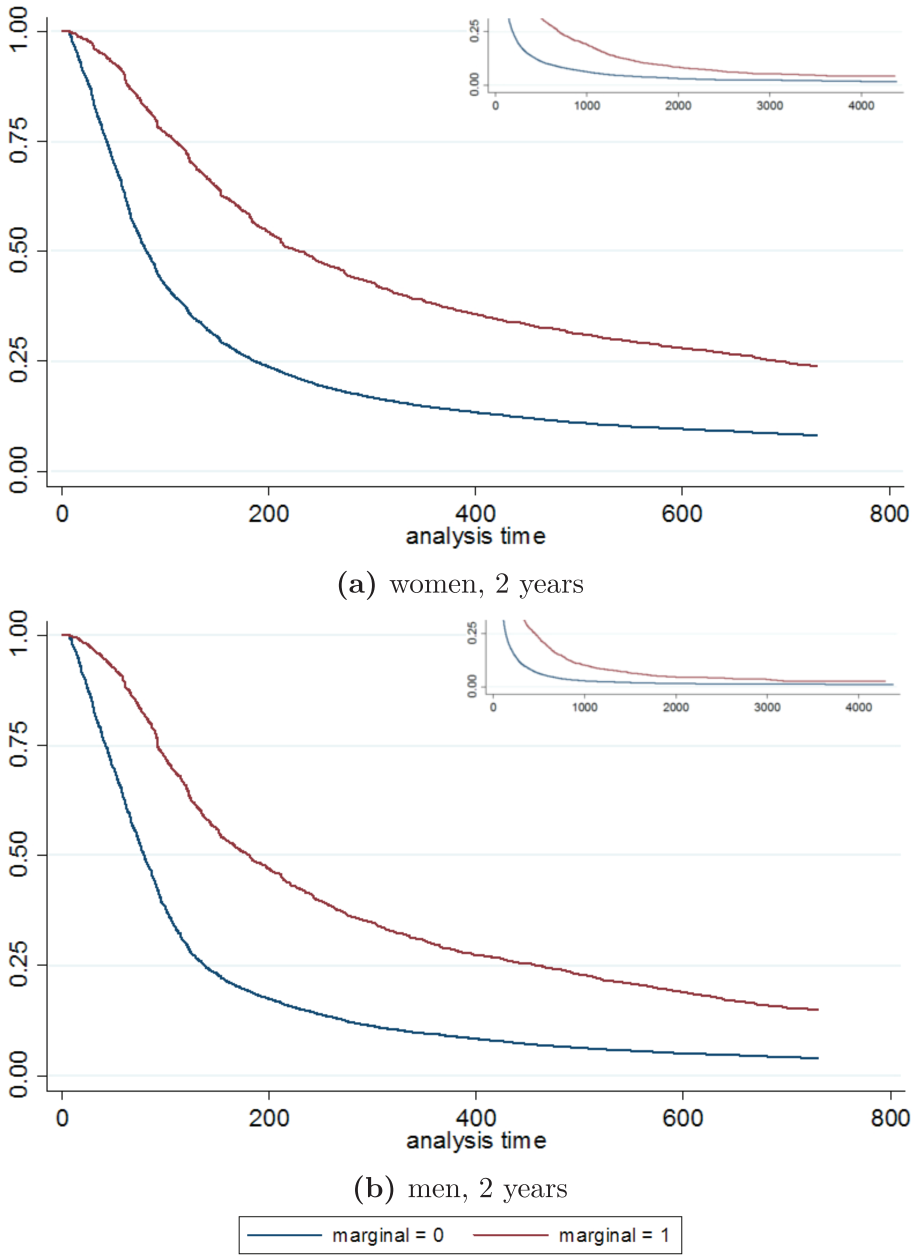
For women (men) who previously worked in Vienna the probability of staying unemployed beyond one month is equal to 89% (89%), respectively 28% (22%) after one year. Out of all women (men) unemployed within the observation period, 85% (82%) of those who previously worked in Tyrol were still without a job after one month. At the end of one year of unemployment there are only 9% (4%) of women (men) still out of work. These findings are of particular interest, because the Kaplan-Meier survival estimates for the federal states with a two seasonal peaks and (like Tyrol) seem to be lower. This finding might be connected to the spikes in seasonal employment within the industry 'accommodation and food service activities'. The higher estimated survival curve for Vienna does not come as a surprise, since Vienna is found to be "*[...] a very dominant urban community characterized by long unemployment duration*" (Lalive, 2008, p.792).

Differences by marginal employment

In Section 1.5 it was explained that marginal employment during the out-of-work spell is counted as unemployment. The reason for this decision was that individuals who are still registered as unemployed might not see the marginal employment as target employment.[6] Figure 2.9 compares the survival curves of individuals who are *marginally employed* during their out-of-work spell (red) to those who are *without employment* during their unemployment spell (blue).

Figure 2.9 shows that for both women and men the differences among being and not being marginally employed are vast in the first two years. The Kaplan-Meier survival estimates indicate that around 38% (30%) of marginally employed women (men) count as long-term unemployed, compared to approximately 15% (10%) of women (men) who are not marginally employed. These outcomes were expected, because marginal

[6] The question might arise, whether counting marginal employment as unemployed is valid. The covariate only indicates that there has been at least one day of marginal employment. Individuals who were marginally employed throughout their unemployment spell could have worked only weekends or only for some weeks (with interruptions). The fact that some of these spells would have been eliminated in the data preparation process leads to the conclusion that counting marginal employment as unemployed has its point.



(a) women, 2 years

(b) men, 2 years

Figure 2.9: Kaplan-Meier survival estimates by marginal employment

employment lowers the pressure of finding new employment through the availability of an income, thus leading to longer OOW-spells.

Differences by the number of children

In this subsection the focus lies on the influence of having children on the duration of unemployment. Recall that out-of-work spells which contain parental leave were eliminated from the sample. The number of children is given for the beginning of the unemployment spell. For the nonparametric analysis I top-coded the number of

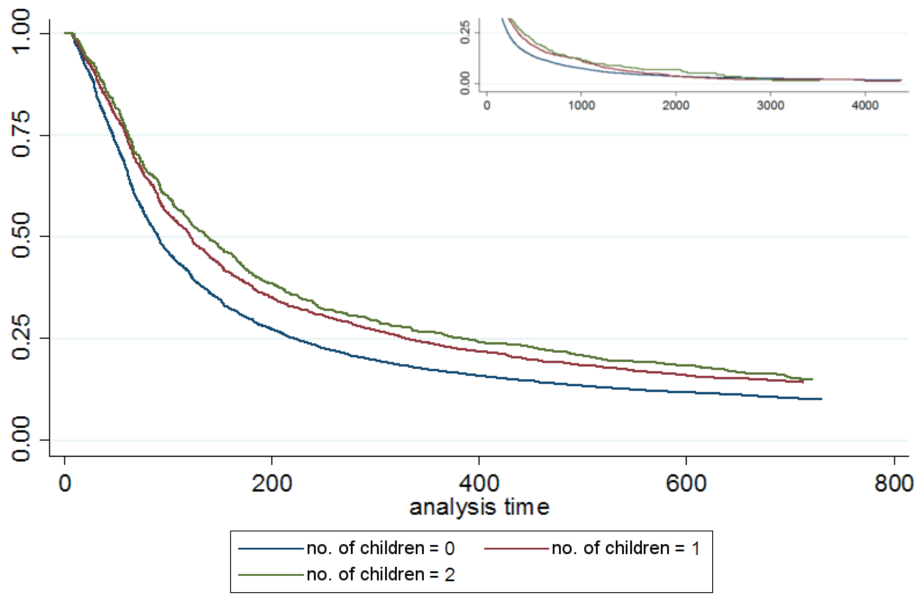


Figure 2.10: Kaplan-Meier survival estimates by number of children, women

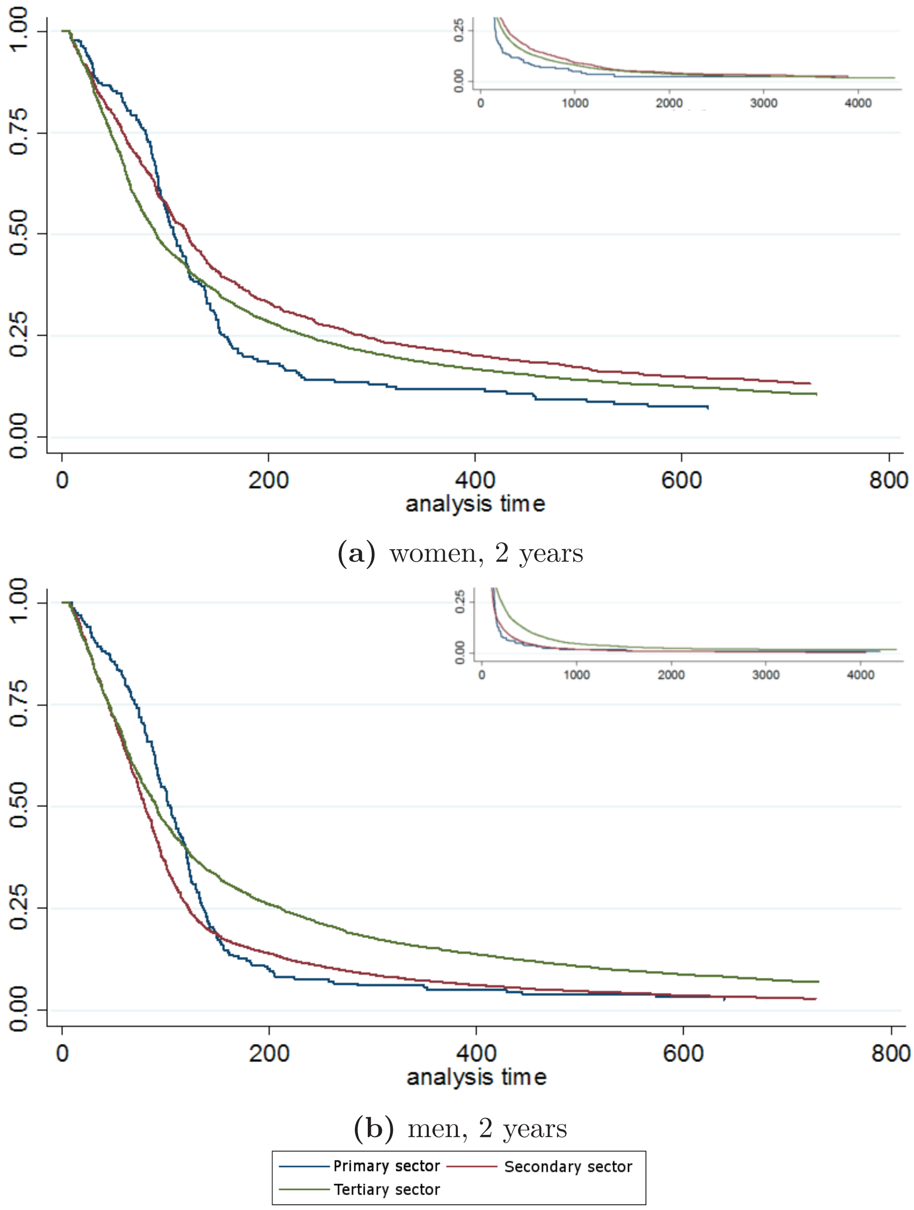
children to two or more children. The blue curve represents women *without children*, whereas the red curve is the survival curve for women with *one child*. The estimated survivor function of women having at least *two children* is depicted by the green curve.

The estimates show that women who do not have children at the beginning of their unemployment spell have an advantage in finding a job. After about five years the difference of individuals left in the sample is almost non existant. Figure 2.10 also indicates that the more children a woman has the longer the stay out of employment. This outcome can be based on two aspects. First, women having children might have different priorities and thus make the choice to stay out of work. The other aspect is, that women might face difficulties in finding a job, because of their children.

Motivating – Differences across (source) industries

Figure 2.11 gives a comparison of survival rates across sectors. Further information on differences in survival among source industries will be illustrated in Figure 2.12. The *primary sector* (blue) represents agriculture, forestry and fishing, the *secondary sector* (red) manufacturing and the *Tertiary sector* (green) summarizes service industries.

The survival estimates in Figure 2.11 show that women's probability of remaining unemployed is lower in the tertiary sector, whereas men seem to have an advantage in the secondary sector. The primary sector's survival rate lies above the other curves at first and then stays below.

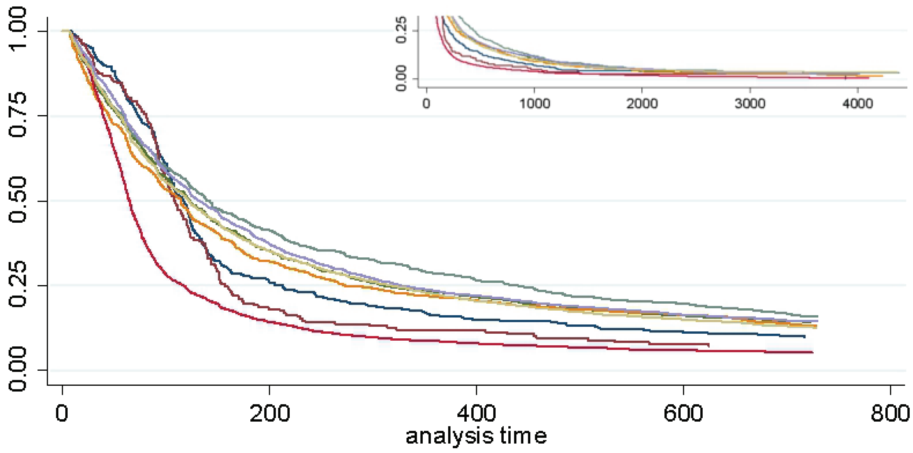


(a) women, 2 years

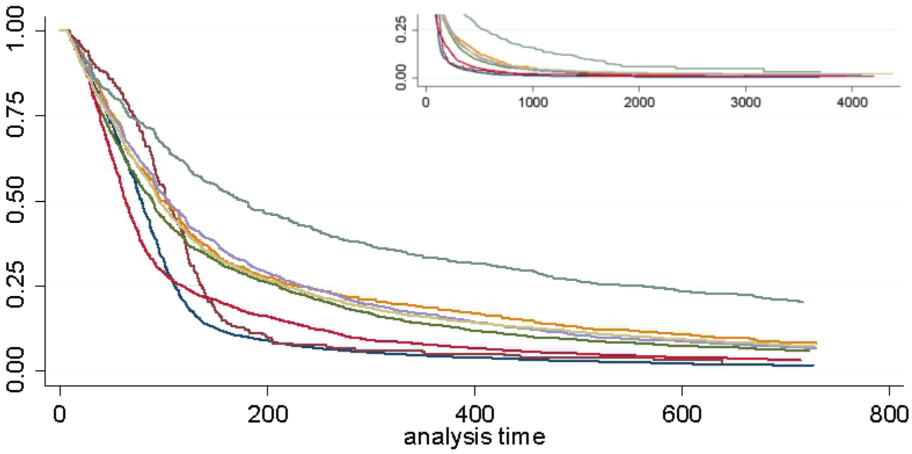
(b) men, 2 years

Figure 2.11: Kaplan-Meier survival estimates by sector

It is expected that the differences do not only depend on the economic sector, but also on the source industries. Therefore a comparison of the most important source industries and the farming industry is included in the nonparametric analysis. Figure 2.12 gives the Kaplan-Meier estimates for a the industries *construction* (constr), *agriculture, forestry and fishing* (farm), *manufacturing* (manu), *public administration and defense* (pub), *human health and social work activities* (soc), *accommodation and food service activities* (tour), *wholesale and retail trade* (trade) and the residual category *others*. The industry 'agriculture, forestry and fishing' represents the primary sector. Through its inclusion Figure 2.11 and 2.12 can be easily compared.



(a) women, 2 years



(b) men, 2 years

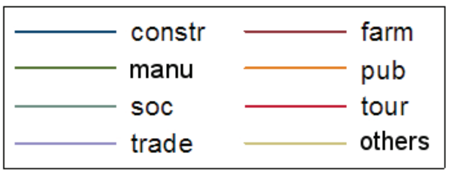


Figure 2.12: Kaplan-Meier survival estimates by industry of last employment

It is interesting that, while the Kaplan-Meier estimates for sectors showed that the primary sector (*farm*) had a lower survival curve than both secondary and tertiary sector (after 14-16 weeks), the division into industries shows that women from the industry 'accommodation and food service activities' (*tour*) have a lower probability of remaining unemployed than the sector it belongs to would indicate.[7] The advantage of finding a job in this industry arises from certain characteristics (Shaw and Williams, 1988). The industry is characterized by:

[7]This outcome does not capture the effects of the week-end employments, since short unemployment spells –less than seven days – were previously eliminated.

1. A very low educational level required for many occupations in this sector.
2. Individuals, who work in this sector (assuming it is not just an internship) are very likely to wait until they get a recall during seasonal peaks.
3. Because of the two aspects mentioned above, jobs in these industries can be used as transitional solutions

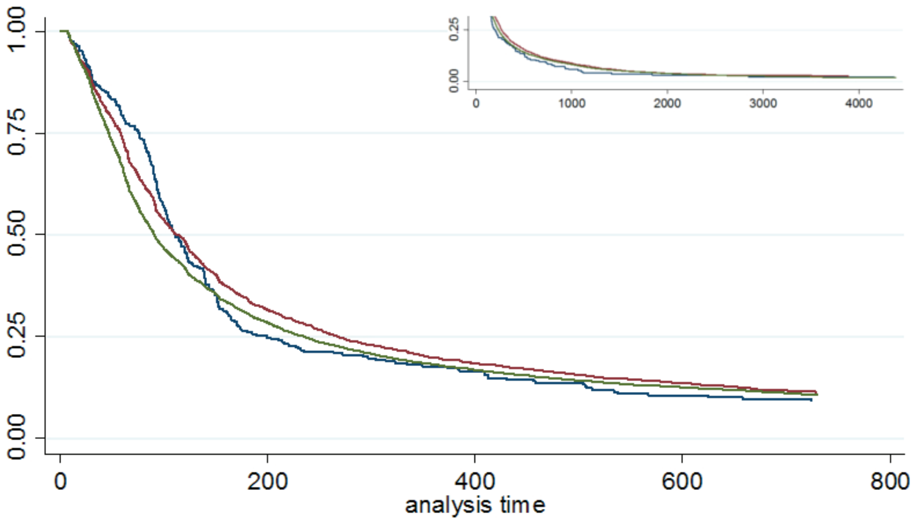
For men, the industries ‘construction’ and ‘agriculture, forestry and fishing’ have the lowest survival estimates. Reasons for short OOW-spells in these industries are the same as point 1. and 2. above. The survival estimates for the industries ‘manufacturing’, ‘wholesale and retail trade’ and ‘public administration’ lie close to the curve of the residual category. It can be observed that individuals who worked in ‘human health and social work activities’ have the highest probability of staying unemployed in the long run and it is even higher for men. Similar results have been found for unemployed individuals in the industries of ‘construction’ and tourism, who were subject to the Regional Extended Benefit Program for Austria (Lalive and Zweimüller, 2004).

As industry-specific characteristics might generate differences across industries, the following subsections focus on the characteristics *industry size*, *labor turnover* and *seasonal peaks*

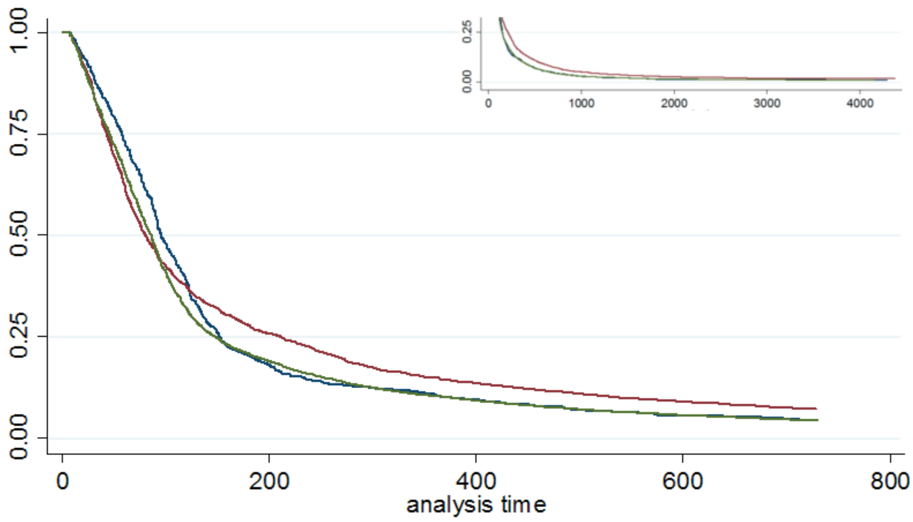
Differences across the size of industry of previous employment

Figure 2.13 compares the Kaplan-Meier estimates for *small* (blue), *medium sized* (red) and *big* industries (green).

The Kaplan-Meier survival estimates by size of the industry differ among men (Figure 2.13b), but are closer together for women (Figure 2.13a). It is interesting that the blue curve depicting small industries and the green curve (big industries) are close together after about 150 days, but the red curve (medium sized) is situated above. Within the first few days the picture is different. After one month of unemployment the probability of staying out of work is lowest for big industries at about 86% (85%) for women (men). In the medium sized industries 88% (85%) of women (men) are still without a job. The highest survival curve after one month belongs to small industries, about 89% of women and men are still unemployed. This result can be interpreted in the following way: For individuals who previously worked in a big industry the possibility of job openings is higher, especially in the first few months. For small industries it might be that industry-specific knowledge improves the job prospects.



(a) women, 2 years



(b) men, 2 years

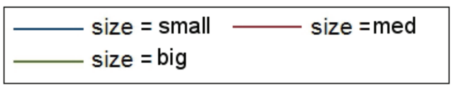
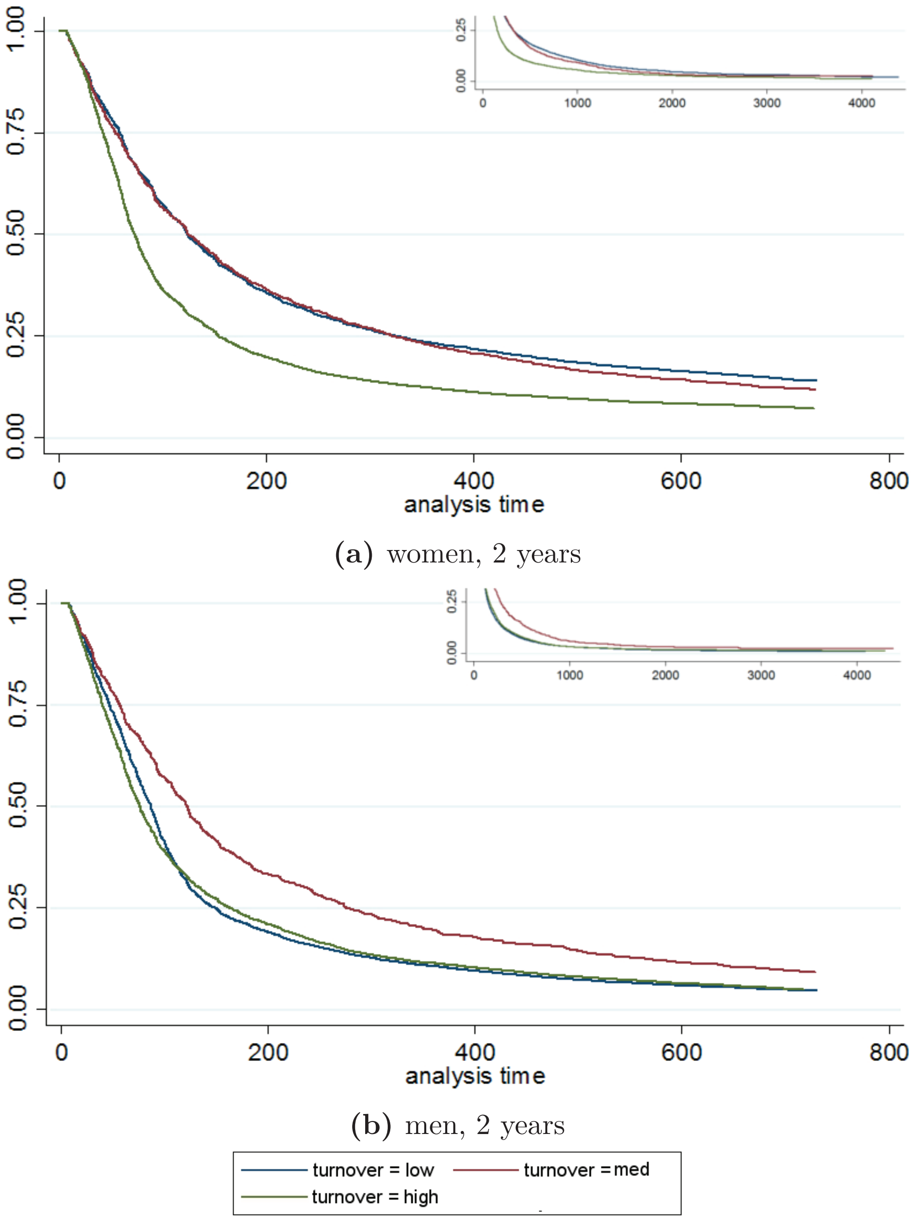


Figure 2.13: Kaplan-Meier survival estimates by size of the industry

Differences across labor turnover

Figure 2.14 represents the probabilities of staying unemployed for industries with different labor turnover, a measure which is defined as the number of employees per year-round employment. The survival curve of industries with *low labor turnover* is colored blue, *medium labor turnover* is depicted red and industries with *high labor turnover* are represented by the green curve.

First, it can be observed that the survival rates of industries with medium and high labor turnover behave approximately the same for men and women. For women the



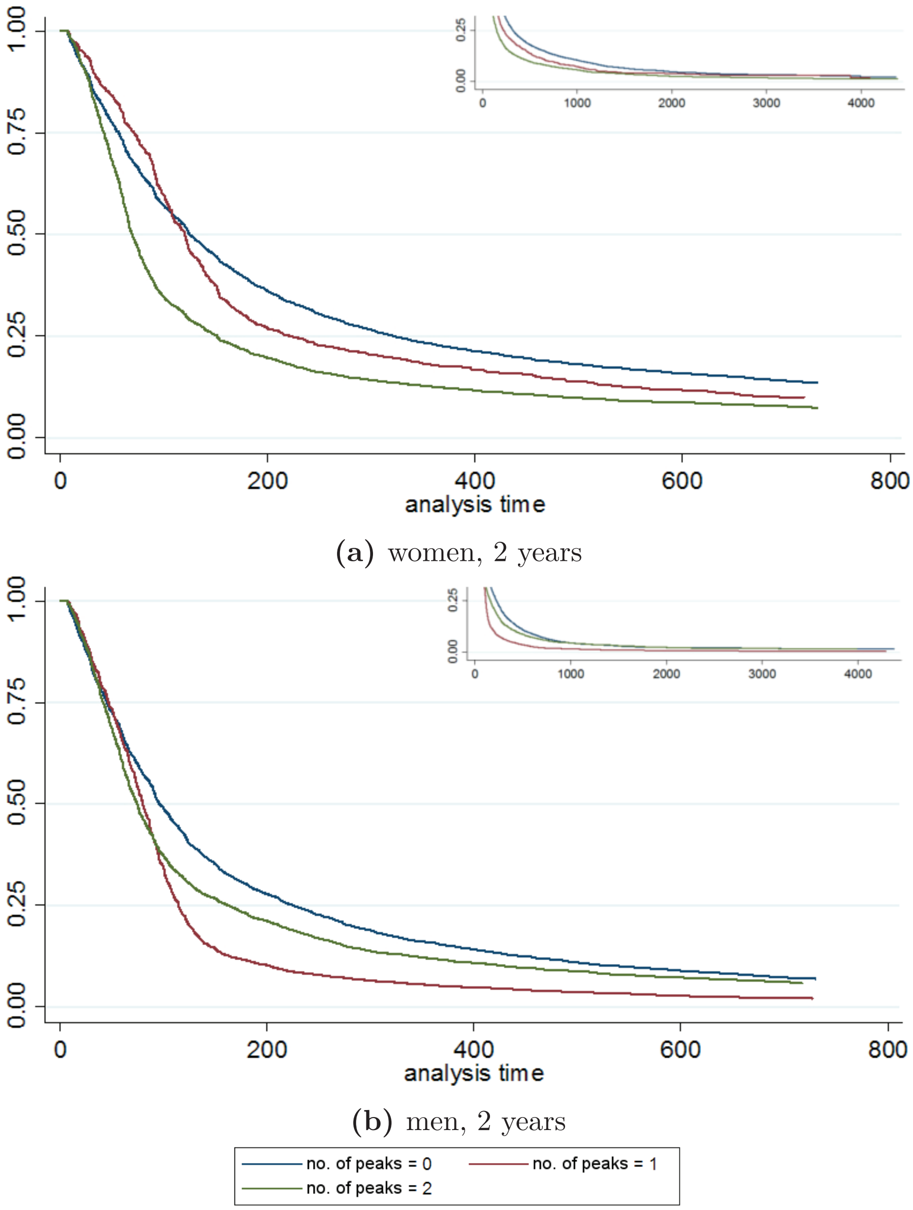
(a) women, 2 years

(b) men, 2 years

Figure 2.14: Kaplan-Meier survival estimates by labor turnover

survival curve of having worked in an industry with low labor turnover behaves similar to the survival rate from industries with medium turnover. For men the picture is the opposite: men who previously worked in industries with low turnover have approximately the same probability of leaving unemployment as in industries with high labor turnover.

Out of all women (men) in the sample about 23% (10%) with previous employment in industries with low labor turnover face long-term unemployment, respectively 23% (19%) for medium turnover and about 12% (11%) for high turnover. The low survival



(a) women, 2 years

(b) men, 2 years

Figure 2.15: Kaplan-Meier survival estimates by number of seasonal peaks

estimates for industries with high labor turnover result from the job openings throughout the year. The similarities of industries with low turnover and high turnover for men can be explained by specialization: It might be that industries with low turnover require specialized employees, which makes it easier to find employment.

Differences across number of seasonal peaks of the source industry

Unemployment duration might also be influenced by seasonal effects. In industries where seasonal peaks, which require a lot of employees, exist, the survival estimates might be lower than the estimates for industries without seasonal peaks. In order to be able to analyze seasonal effects I designed a categorical variable, which lists the number of employment peaks per year. In Figure 2.15 the blue curve represents industries that experience *no seasonal effects*, industries with *one seasonal peak* are colored red, *two seasonal peaks* are represented by the green curve.

The results derived from the Kaplan-Meier estimator are different across gender. While for women having previously been employed in industries with two employment peaks increases the probability of finding a job, men's job finding probability benefits from previous employment in an industry with one seasonal peak. This is related to the dominant industries, i.e. 'accommodation and food service activities' (two seasonal peaks) for women and 'construction' (one seasonal peak) for men. The lower probability of employment in industries which are not subject to seasonality is common for both men and women. After one year about 23% of the women and 15% of the men who worked in these industries are still searching for a job. Especially in this case the knowledge of full-time or part-time employment would give further insights. Previous literature suggests that individuals who are part-time employed in tourism are more often laid off than full time-employed individuals (Shaw and Williams, 1988).

2.2.2 Comparison of the results: Cox Proportional Hazards, Weibull (AFT) and Loglogistic model

In this subsection the influence of covariates on unemployment duration is analyzed. The Cox Proportional Hazards model (semiparametric), the Weibull (AFT) model and the Loglogistic model (parametric models) were applied.

The first and fourth column of Tables 2.9 and 2.10 give the regression output for the Cox Proportional Hazards model, which estimates the effect of the covariates on the hazard rate, i.e. hazard ratio (HR). The other four columns correspond to the Weibull Model (AFT metric) and the Loglogistic model (AFT metric), which model the accelerating or decelerating effect of covariates on the time spent in unemployment, i.e. the time ratio (TR). As the presented coefficients estimate different effects, the relationships of the PH coefficients on out-of-work duration have opposite directions than the AFT coefficients. While hazard ratios greater than one equal a higher probability of finding employment, a time ratio greater than one mean a longer unemployment duration, i.e. survival is prolonged. Both hazard and time ratio are different from the common coefficients. In order to obtain coefficients instead of hazard or time ratios one has to take the logarithm of the hazard or time ratio.

Table 2.9: Estimation results: Cox Proportional Hazards, Weibull and Loglogistic model, Part 1

	Women						Men					
	Cox PH		**Weibull AFT**		**Loglogistic AFT**		**Cox PH**		**Weibull AFT**		**Loglogistic AFT**	
Variable	**HR**	**(SE)**	**TR**	**(SE)**	**TR**	**(SE)**	**HR**	**(SE)**	**TR**	**(SE)**	**TR**	**(SE)**
			Equation 1 : _t						Equation 1 : _t			
35-44	.	.	.	.	.	.	.	.	.	.	.	.
15-24	0.965	(0.043)	1.196**	(0.068)	0.935	(0.048)	0.925†	(0.042)	1.042	(0.053)	1.084†	(0.050)
25-34	0.897**	(0.022)	1.268**	(0.040)	1.037	(0.029)	1.057**	(0.022)	0.919**	(0.022)	0.937**	(0.019)
45-54	0.949*	(0.024)	1.050	(0.034)	1.070*	(0.030)	0.884**	(0.019)	1.179**	(0.028)	1.098**	(0.023)
55-64	0.960	(0.050)	0.944	(0.064)	1.070	(0.060)	0.863**	(0.030)	1.161**	(0.045)	1.159**	(0.038)
high edu	.	.	.	.	.	.	.	.	.	.	.	.
low edu	0.941*	(0.026)	1.068†	(0.038)	1.080*	(0.035)	1.194**	(0.036)	0.761**	(0.026)	0.789**	(0.025)
appr	1.157**	(0.033)	0.816**	(0.030)	0.857**	(0.028)	1.491**	(0.044)	0.572**	(0.019)	0.653**	(0.020)
med edu	1.094**	(0.038)	0.875**	(0.039)	0.917*	(0.037)	1.281**	(0.058)	0.700**	(0.035)	0.751**	(0.035)
uni	1.066	(0.048)	0.885*	(0.052)	0.948	(0.051)	0.962	(0.051)	1.048	(0.062)	1.044	(0.059)
paid	1.112**	(0.039)	0.806**	(0.036)	1.098*	(0.050)	0.877**	(0.032)	0.983	(0.039)	1.384**	(0.055)
child	0.835**	(0.019)	1.221**	(0.036)	1.272**	(0.033)	-	-	-	-	-	-
urate	1.006	(0.008)	0.975*	(0.010)	1.008	(0.010)	0.993	(0.008)	1.007	(0.009)	1.006	(0.008)
marginal	0.543**	(0.012)	2.433**	(0.070)	2.463**	(0.064)	0.541**	(0.013)	2.179**	(0.057)	2.194**	(0.051)
prevep	1.000**	(0.000)	1.000**	(0.000)	1.000**	(0.000)	1.000**	(0.000)	1.000**	(0.000)	1.000**	(0.000)
1.quart	.	.	.	.	.	.	.	.	.	.	.	.
2.quart	1.094**	(0.027)	0.895**	(0.028)	0.917**	(0.026)	0.835**	(0.018)	1.218**	(0.029)	1.091**	(0.024)
3.quart	1.011	(0.025)	0.945†	(0.030)	1.005	(0.029)	0.701**	(0.015)	1.485**	(0.036)	1.318**	(0.030)
4.quart	1.115**	(0.026)	0.848**	(0.026)	0.951†	(0.026)	0.819**	(0.014)	1.141**	(0.022)	1.274**	(0.021)

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

The covariates *35-44* and *1.quart* represent the base groups of *age* and *quarter* the unemployment spell started in. The age base group has been chosen, because this age group is assumed to have stable labor market careers.

Table 2.10: Estimation results: Cox Proportional Hazards, Weibull and Loglogistic model, Part 2

	Women						Men					
	Cox PH		**Weibull AFT**		**Loglogistic AFT**		**Cox PH**		**Weibull AFT**		**Loglogistic AFT**	
Variable	**HR**	**(SE)**	**TR**	**(SE)**	**TR**	**(SE)**	**HR**	**(SE)**	**TR**	**(SE)**	**TR**	**(SE)**
			Equation 1 : _t						Equation 1 : _t			
Vienna	.	.	.	.	.	.	.	.	.	.	.	.
Bgld	1.149*	(0.070)	0.754**	(0.059)	0.832**	(0.057)	1.200**	(0.053)	0.765**	(0.038)	0.858**	(0.036)
Carinthia	1.397**	(0.046)	0.614**	(0.026)	0.666**	(0.025)	1.489**	(0.041)	0.586**	(0.018)	0.704**	(0.019)
Low Aus	1.140**	(0.034)	0.833**	(0.032)	0.812**	(0.028)	1.271**	(0.031)	0.715**	(0.019)	0.788**	(0.019)
Up Aus	1.216**	(0.036)	0.772**	(0.029)	0.736**	(0.025)	1.403**	(0.034)	0.641**	(0.017)	0.709**	(0.017)
Sbg	1.567**	(0.052)	0.553**	(0.023)	0.593**	(0.022)	1.486**	(0.042)	0.612**	(0.019)	0.672**	(0.019)
Styria	1.249**	(0.037)	0.720**	(0.027)	0.746**	(0.026)	1.282**	(0.032)	0.711**	(0.019)	0.777**	(0.019)
Tyrol	1.739**	(0.051)	0.495**	(0.019)	0.560**	(0.019)	1.703**	(0.045)	0.519**	(0.015)	0.622**	(0.017)
Vbg	1.197**	(0.054)	0.818**	(0.048)	0.788**	(0.041)	1.181**	(0.048)	0.883**	(0.040)	0.789**	(0.032)
low turn	.	.	.	.	.	.	.	.	.	.	.	.
med turn	1.092**	(0.031)	0.865**	(0.032)	0.918**	(0.030)	1.178**	(0.051)	0.839**	(0.040)	0.883**	(0.039)
high turn	1.213**	(0.025)	0.772**	(0.020)	0.808**	(0.019)	1.193**	(0.024)	0.805**	(0.018)	0.839**	(0.018)
no peak	.	.	.	.	.	.	.	.	.	.	.	.
1 peaks	0.999	(0.036)	0.953	(0.044)	1.060	(0.042)	1.362**	(0.025)	0.667**	(0.013)	0.838**	(0.014)
2 peaks	1.202**	(0.025)	0.790**	(0.021)	0.852**	(0.021)	1.246**	(0.033)	0.777**	(0.023)	0.861**	(0.023)
small ind	-	-	-	-	-	-	.	.	.	.	.	.
med ind	-	-	-	-	-	-	1.026	(0.043)	1.041	(0.049)	0.957	(0.039)
big ind	-	-	-	-	-	-	1.213**	(0.049)	0.819**	(0.037)	0.833**	(0.032)
Intercept	-	-	419.6**	(31.65)	127.6**	(8.946)	-	-	366.1**	(27.90)	114.1**	(7.990)
			ln_p		ln_γ				ln_p		ln_γ	
Intercept	-	-	-0.254**	(0.006)	-0.386**	(0.007)		-	-0.108**	(0.005)	-0.547**	(0.006)

Significance levels: † : 10% $*$: 5% $**$: 1%

Vienna, *low turn no peak* and *small ind* represent the base groups of *federal states*, *labor turnover number of seasonal peaks* and *industry size*.

It can be observed that, within the sample of men, more covariates are significant. Similar to the findings of the nonparametric analysis, it is shown that women aged 25-34 have a lower probability of finding employment than the base group of 35 to 44-year-olds, i.e. the hazard rate is ceteris paribus 10.3% lower (Cox PH model). When looking at how the process is accelerated, the Weibull model tells us that these women leave out-of-work periods about 26.8% slower, respectively 3.7% slower for the Loglogistic model. For men the outcome is a bit different: Compared to the base group only the individuals of age group 25-34 seem to have an advantage on the labor market, i.e. according to the Cox PH model the hazard rate is 5.7% higher.

Compared to having a high educational level, having completed an apprenticeship increases the hazard rate of women immensely, i.e. ceteris paribus by 15.7% (Cox PH model). Their unemployment duration is 18.4% (14.3%) shorter in the Weibull (Loglogistic) model. For men this effect is even more pronounced: The hazard rate rises by 49.1%, the stay out of work is 42.8% (34.7%) shorter. The effect of having completed a medium educational level is found to influence the time out of work in the same direction. Only individuals with a low educational level perform poorer than highly educated individuals in the labor market. Findings in previous literature have been controversial in this aspect. For this sample of the Austrian labor market, the results of highly educated individuals are comparable to those of REBP recipients in Austria (Lalive and Zweimüller, 2004).

Though in the nonparametric estimation the survival curves of the individuals by benefit receipt seemed to behave very similar for women and men, semiparametric and parametric estimation find different impacts on unemployment duration. While for women the hazard rate of benefit recipients is about 11% higher, it is 12.3% lower for men. The time ratios of the Weibull (Loglogistic) model state that women spend 19.4% less (9.8% more) time in unemployment while receiving benefits. In the Loglogistic model men are found to have 38.4% longer out-of-work spells. This comparison of results from different models indicates that the findings are not robust. In the literature there are many papers studying unemployment duration in Austria (Winter-Ebmer, 1998; Lalive et al., 2006; Lalive, 2008), where it has been found that an extension of benefit eligibility increases the duration of unemployment. Taking previous literature into account, the outcomes of the Loglogistic model seems to be the most appropriate.

Taking a look at both the values of the time ratio of the Weibull and Loglogistic model of the covariate *marginal*, it can be seen that, compared to the unemployment duration of individuals without marginal employment during their unemployment spells, the spell length marginally employed individuals is found to be more than twice as long. Quarterly effects on employment were found to be significant, but reverse for men and women. Women who started their unemployment spell in the first quarter have a at least 9.4% higher hazard rate (Cox PH model), respectively their spell is at least 10.5% shorter in the Weibull model. For men the hazard rate is at least 16.5% lower compared to men who became unemployed in the first quarter of the year. The time spent in unemployment is found to be at least 14.1% (9.1%) longer in the Weibull (Loglogistic) model.

The reverse effect of unemployment rate on the hazard rate –the higher the unemployment rate, the lower the hazard rate– which had been found for Austria from 1988-1993 (Lalive and Zweimüller, 2004) could not be supported in this selected sample, which might be due to either the empirical design of the data or the selected time frame.

The effects of the federal state the individual has worked in before show similar results to the nonparametric analysis. In the urban state of Vienna unemployment spells seem to be at least 20% longer than in any other federal state. Especially in Tyrol, Salzburg and Carinthia –all federall states that experience touristic peaks in summer and winter– individuals have a higher probability of finding employment and respectively a shorter stay out of work. When having previously worked in Tyrol women (men) have a 73.9% (70.3%) higher hazard rate. When looking at the Loglogistic model women (men) are found to spend 44% (37.8%) less time out of work.

Tables 2.9 and 2.10 show that most hazard and time ratios for the constructed industry-specific covariates are highly significant. The size of the previous industry was not a significant predictor for unemployment duration of women, so it had been excluded from the analysis. For men only coming from a big industry is found to have an impact on out-of-work spells: Compared to small industries the hazard rate is 21.3% higher. Depending on the regression method, men's unemployment duration is reduced by about 16% (Loglogistic model) to 18% (Weibull model).

The impact of labor turnover is similar for women and men. Having previously been employed in industries with high labor turnover the hazard rate is 21.3% (women), respectively 19.3% (men) higher compared to the ones with low turnover. The Accelerated Failure Time models found that among women who have worked in industries with medium labor turnover the spell length should be 13.5% (8.2%) shorter for the

Weibull (Loglogistic model), respectively 16.1% (11.7%) shorter for men. When having been employed in industries with high labor turnover the spells have been found to be the shortest, i.e. 22.8% (19.2%) in the Weibull (Loglogistic) model for women and 19.5% (16.1%) for men.

When focusing on seasonal effects, the results state that presence of seasonal peaks increases the probability of leaving OOW-spells. The importance of industries can also be considered when interpreting the results. In Subsection 2.2.1 it has been shown that women having previously worked in the industry 'accommodation and food service activities' experience the lowest survival estimate. Tables A.5 and A.6 show that a lot of source spells belong to this industry. For men the same has been discovered for the industry 'construction' (Tables A.7 and A.8). The fact that out-of-work spells of individuals who have worked in industries with two seasonal peaks tend to be 14.8% (13.9%) shorter than OOW-spells of industries with no seasonal peak for women (men) might therefore correlate with the large amount of 'accommodation and food service activities'-industry in the sample. For men originating from a industry with one seasonal peak –i.e. construction– compared to industries with no seasonal peak increases the hazard rate by 36.2% in the Cox Proportional Hazards model. The spells have found to be 33.3% shorter in the Weibull model, respectively 16.2% shorter in the Loglogistic model.

Another interesting aspect of duration analysis is duration dependence. Parametric analysis includes an estimate for the shape parameter indicating the shape of the hazard rate function. The negative coefficient for the intercept for ln_p in the columns *Weibull AFT* in Table 2.10 indicates negative duration dependence in the Weibull Model, i.e. the longer an individual is unemployed, the lower is the probability of exiting unemployment. This finding is well established in the literature, but it has been stated that negative duration dependence can have two causes: One is negative duration dependence itself and the other is unobserved heterogeneity (Calderón-Madrid, 2008). Though in my opinion unobserved heterogeneity is present in most studies (i.e. an individual's abilities, personality traits, name of the individual, motivation and other factors) the fact that finding a job is harder after experiencing longer unemployment spells seems intuitive.

The negative intercept for ln_γ in the columns *Loglogistic AFT* in Table 2.10 indicates that at first the employment hazard is rising and then declining. The assumption of a nonmonotonic hazard applies even better to out-of-work spells in this sample.

Recalling that the minimum duration is seven days, it becomes obvious that finding a job in a time span longer than one week becomes even more probable. Similar to the Weibull model, the negative duration dependence for longer spell durations is realistic.

Finally, a comparison of the Akaike information criterion in Table 2.11 shows that the Loglogistic model, which allows for a nonmonotonic baseline hazard, seems to be the best fit for the data, i.e. it has the smallest value. The same result can be obtained from the Bayesian information criterion (BIC). Comparing the Log likelihood or applying the Wald test would not lead to reliable results, because the models are not nested. The model with the poorest fit is the Cox Proportional Hazards model. Despite the advantages that arise from not having to specify a baseline hazard, the proportionality assumption is a really strong assumption. Tables C.1 and C.2 in Appendix C show that the proportionality assumption does not hold.

Table 2.11: AIC and BIC

	Women			Men		
Model	**df**	**AIC**	**BIC**	**df**	**AIC**	**BIC**
Cox PH	28	262,728.1	262,943.6	29	388,190.1	388,423.3
Weibull	30	55,793.4	56,024.3	31	71,666.5	71,915.8
Loglogistic	30	**50,823.7**	51,054.6	31	**64,680.9**	64,930.2

Chapter 3

Conclusion

The analysis of out-of-work spells of men and women with Austrian citizenship in the Austrian labor market reached similar results to other European and non-European countries. The nonparametric analysis of age groups showed that there are differences in the course of survival estimates between men and women. An interesting, but expected finding was, that women aged 25-34 (the years during which the probability that women have children is the highest) tend to have longer unemployment spells than men and all other age groups of women for spell lengths to up to a bit over three years. This result, which can be due to personal choices or difficulties for these individuals to find employment, is also supported by semiparametric and parametric estimation. It was possible to control for children in the sample of women. Nonparametric analysis showed that the more children a woman had at the beginning of her unemployment spell, the higher was the probability of staying unemployed. In the Loglogistic model it could be obtained that out-of-work spells are about 27.2% longer.

Another interesting finding concerns the receipt of unemployment payments. Nonparametric analysis shows higher probabilities of staying out of work in the first few weeks of unemployment when receiving benefits. Semiparametric and parametric estimation confirm previous findings that benefit payments prolong the stay in unemployment, i.e. women receiving benefit payments have 9.8% longer unemployment spells than individuals who do not receive payments. For men the spells were even 38.4% longer, according to the Loglogistic model.

Differences can also be found across educational levels. Nonparametric analysis showed that the effect of educational level on unemployment duration is more pronounced within the sample of men, which could be explained by the job choices men and women in the sample make. It has been shown that especially in the sample of women, 31% percent of the spells belong to the industry 'accommodation and food service activities', which is also the biggest target industry, i.e. 28% of the spells end with employment in this industry. This sector is known to have low educational requirements. Furthermore, the individuals who have completed an apprenticeship have a shorter stay in unemployment, while individuals with a high educational level stay unemployed longer compared to almost all other groups.

It has been found that unemployment spells differ across industries. Especially individuals who previously worked in industries with seasonal employment peaks and low requirements regarding the educational level ('construction' for men, 'accommodation and food service activities' for women) are faster in finding new employment. These industries also comprise a large share of unemployment spells within their sample, i.e. 31% of all OOW-spells. It has to be taken into account that the shorter stays in unemployment do not give any information on how the stay in employment will look like, i.e. the tourism industry does often face the prejudice that it is a lesser job (Shaw and Williams, 1988), which could be due to non-stable employment-careers.

Concerning industry-specific characteristics the Kaplan-Meier estimates showed that for big industries the probability of leaving the out-of-work spell is higher than for small sized industries, i.e. the survival rates are lower. There were no significant differences across sizes detected for the sample of women. For men parametric estimation with the assumption of a Loglogistic distribution of the baseline hazard gives that spells are 16.7% shorter for big industries compared to small ones. The estimates for the labor turnover indicate that finding employment for individuals who worked in industries with low labor turnover is the hardest. As expected industries with high labor turnover, i.e. industries in which more individuals are hired per year-round employment, shorten the duration of the out-of-work spell by 19.2% (16.1%) for women (men). When looking at seasonal employment peaks, the nonparametric, semiparametric and parametric analyses found that having previously worked in an industry with employment peaks shortens the out-of-work stay. It is interesting that men were found to endure shorter spells in industries with one seasonal peak, while women find employment quicker if they previously worked in industries with two seasonal peaks. This finding might be closely related to the fact that 31% (31%) of the spells in the sample correspond to women (men) who previously worked in the industry 'accommodation and food service activities' ('construction'). Both industries are known to have low entrance requirements for some professions. The main finding of the parametric estimation concerning the shape of the baseline hazard was that there seems to be negative duration dependence in the long run, which is usually the case for the analysis of unemployment duration.

The analysis of the Austrian labor market –with the use of a large sample of administrative data (linked employer-employee data), without having to face the problems, which may arise from surveys– showed that differences in unemployment duration can have many causes. Apart from the advantages that can be named, there are also some disadvantages: Interesting characteristics i.e. marital status, number of children for men, number of individuals per household, and others could not be taken into account. Also there is no information on the type of employment (part-time or full-time job). For the future it would therefore be interesting to include survey data or develop proxies for full/part-time employment (e.g. through yearly income). Another aspect, which might be interesting for further research is the inclusion of individuals of other citizenships, the extension of the sample and integration of a more detailed level of industries to determine whether differences across industries are driven by a certain subgroup, i.e. if the subgroup 'food and beverage service activities' determines the characteristic course of unemployment for the industry 'accommodation and food service activities'. This inclusion has to be done carefully as administrative changes in the classification are more likely at this granularity.

Bibliography

Aalen, O. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–726, 1978.

Arai, M. and Skogman Thoursie, P. Renouncing personal names: An empirical examination of surname change and earnings. *Journal of Labor Economics*, 27(1):127–147, 2009.

Austrian Social Security, 2012. URL `http://www.sozialversicherung.at/portal27/portal/esvportal/channel_content/cmsWindow?action=2&p_menuid=72962&p_tabid=5&p_pubid=657127`.

Badinger, H. and Url, T. Determinants of regional unemployment: some evidence from Austria. *Regional studies*, 36(9):977–988, 2002.

Berenguer, A. E. *Statistical and Empirical Issues in the Analysis of Duration data.* PhD thesis, Departament d'Economia i Emprese, University Pompeu Fabra, October 2001.

Bergström, R. and Edin, P.-A. Time aggregation and the distributional shape of unemployment duration. *Journal of Applied Econometrics*, 7(1):5–30, 1992.

Biddle, J. E. and Hamermesh, D. S. Beauty, productivity and discrimination: lawyers' looks and lucre. *National Bureau of Economic Research Working Paper No. 5366*, 1998.

Boršic, D. and Kavkler, A. Duration of regional unemployment spells in Slovenia. *Managing Global Transitions*, 7(2):123–146, 2009.

Bover, O., Arellano, M., and Bentolila, S. Unemployment duration, benefit duration and the business cycle. *The Economic Journal*, 112(479):223–265, 2002.

Brenner, C., Schütz, K., Zorn, J., and Mierke, K. *Aktuelle Themen der Wirtschaftspsychologie-Beiträge und Studien*, volume 1, chapter Diskriminierung von

Stellenbewerbern ausländischer Abstammung: Ein Experiment mit Laien und Experten. Kölner Wissenschaftsverlag, 2008.

Bring, J. and Carling, K. Attrition and misclassification of drop-outs in the analysis of unemployment duration. *Journal of Official Statistics-Stockholm*, 16(4):321–330, 2000.

Bursell, M. What's in a name? A field experiment test for the existence of ethnic discrimination in the hiring process. *SULCIS WP*, 7(7), 2007.

Calderón-Madrid, A. Unemployment Duration in Mexico: Its Determinants and Implications for the Labor Market Segmentation Controversy and Public Policy Design. *El Colegio de México, Mimeo*, 2008.

Calvó-Armengol, A. and Jackson, M. O. The effects of social networks on employment and inequality. *The American Economic Review*, 94(3):426–454, 2004.

Card, D., Chetty, R., and Weber, A. The spike at benefit exhaustion: Leaving the unemployment system or starting a new job? *National Bureau of Economic Research Working Paper No. 12893*, 2007.

Ciucă, V. and Matei, M. Survival rates in unemployment. *International Journal of Mathematical Models and Methods in Applied Sciences*, 5(2):362–370, 2011.

Cleves, M. A., Gutierrez, R. G., Gould, W. W., and Marchenko, Y. V. *An Introduction to Survival Analysis Using Stata*. Stata Press, 2010.

Corak, M. Measuring the duration of unemployment spells. *The Canadian Journal of Economics/Revue canadienne d'Economique*, 29:S43–S49, 1996.

Cox, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

Cox, D. R. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

Dănăcică, D.-E. and Babucea, A.-G. Using Survival Analysis In Economics. *Analele Stiintifice ale Universitatii "Alexandru Ioan Cuza" din Iasi - Stiinte Economice*, 57:439–450, November 2010. URL `http://anale.feaa.uaic.ro/anale/resurse/sta_danacica_babucea.pdf`.

Federal Ministry of Economy, Family and Youth. Wirtschaftsbericht Österreich 2010, July 2010. URL `http://www.bmwfj.gv.at/Wirtschaftspolitik/Wirtschaftspolitik/Documents/Wirtschaftsbericht%20%C3%96sterreich%202010.pdf`. Accessed: 2013-05-13.

Federal Ministry of Economy, Family and Youth. Wirtschaftsbericht Österreich 2012, June 2012. URL http://www.bmwfj.gv.at/Wirtschaftspolitik/Wirtschaftspolitik/Documents/Wirtschaftsbericht%20%C3%96sterreich%202012.pdf. Accessed: 2013-05-16.

Federal Ministry of Labor, Social Affairs and Consumer Protection. Arbeitsmarktdaten März 2013, 2013. URL http://www.bmask.gv.at/site/Arbeit/News/Arbeitsmarktdaten_Maerz_2013. Accessed: 2013-05-17.

Flinn, C. and Heckman, J. New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics*, 18(1):115–168, 1982.

Gamerman, D. and West, M. An application of dynamic survival models in unemployment studies. *The Statistician*, 36(2/3):269–274, 1987.

Governatori, M., Grzegorzewska, M., Maier, C., Medeiros, J., Meyermans, E., Minty, P., Slebinger, M., Valk, J. v. d., et al. Employment in Europe 2009. *Office for Official Publications of the European Communities*, 2009.

Hunt, J. The effect of unemployment compensation on unemployment duration in Germany. *Journal of Labor Economics*, 13(1):88–120, 1995.

Kaplan, E. L. and Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

Katz, L. F. Layoffs, recall and the duration of unemployment. *National Bureau of Economic Research Cambridge Working Paper No. 1825*, 1986.

Kavkler, A., Dănăcică, A. G. Daniela-Emanuela Babucea, Bicanic, I., Böhm, B., Tevdovski, D., Toševska, K., and Boršič, D. Cox Regression Models for Unemployment Duration in Romania, Austria, Slovenia, Croatia, and Macedonia. *Romanian Journal of Economic Forecasting*, 10(2):81–105, 2009.

Kiefer, N. M. Economic duration data and hazard functions. *Journal of economic literature*, 26(2):646–679, 1988.

Kong, J. The Analysis of Graduates' Job Search by Nonparametric Survival Analysis. *2012 International Conference on Technology and Management. Lecture notes in Information Technology*, 21, 2012.

Lalive, R. How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics*, 142(2):785–806, 2008.

Lalive, R. and Zweimüller, J. Benefit entitlement and unemployment duration: The role of policy endogeneity. *Journal of Public Economics*, 88(12):2587–2616, 2004.

Lalive, R., Van Ours, J., and Zweimüller, J. How changes in financial incentives affect the duration of unemployment. *The Review of Economic Studies*, 73(4):1009–1038, 2006.

Lancaster, T. Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society*, 47(4):939–956, 1979.

Lancaster, T. *The econometric analysis of transition data.* Economic Society Monographs 17. Cambridge University Press, 1992.

Mussida, C. Unemployment duration and competing risks: A regional investigation. *Università Cattolica del Sacro Cuore, Dipartimenti e Istituti di Scienze Economiche (DISCE)*, No. dises0749, 2007.

Narendranathan, W. and Stewart, M. B. Modelling the probability of leaving unemployment: competing risks models with flexible base-line hazards. *Applied Statistics*, 42(1):63–83, 1993.

Narendranathan, W., Nickell, S., and Stern, J. Unemployment benefits revisited. *The Economic Journal*, 95(378):307–329, 1985.

Nelson, W. Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966, 1972.

Public Employment Service Austria (AMS), 2013. URL `http://www.ams.at/sfa/14080.html`.

Pyy-Martikainen, M. and Rendtel, U. Measurement errors in retrospective reports of event histories. A validation study with Finnish register data. *Survey Research Methods*, 3(3):139–155, 2009.

Schoenfeld, D. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

Shaw, G. and Williams, A. Tourism and employment: reflections on a pilot study of Looe, Cornwall. *Area*, 20(1):23–34, 1988.

STATISTICS AUSTRIA. Arbeitsmarktstatistik 4.Quartal 2011. Mikrozensus-Arbeitskräfteerhebung, 2012. URL `http://www.statistik.at/web_de/static/arbeitsmarktstatistik_-_4._quartal_2011_schnellbericht_063304.pdf`. Accessed: 2013-05-17.

STATISTICS AUSTRIA. Formale Struktur des Österreichischen Bildungswesens, 2013a. URL `https://www.statistik.at/web_de/statistiken/bildung_und_kultur/formales_bildungswesen/index.html`.

STATISTICS AUSTRIA. Arbeitsmarktstatistik 4.Quartal 2012. Mikrozensus-Arbeitskräfteerhebung, 2013b. URL `http://www.statistik.at/web_de/static/arbeitsmarktstatistik_-_4._quartal_2012_schnellbericht_070351.pdf`. Accessed: 2013-05-17.

Uysal, S. D. and Pohlmeier, W. Unemployment duration and personality. *Journal of Economic Psychology*, 32(6):980–992, 2011.

Van Ours, J. C. and Vodopivec, M. How shortening the potential duration of unemployment benefits affects the duration of unemployment: Evidence from a natural experiment. *Journal of Labor Economics*, 24(2):351–378, 2006.

Vermunt, J. K. *Log-linear event history analysis: A general approach with missing data, latent variables, and unobserved heterogeneity*, volume 8. Tilburg University Press Tilburg., 1996.

Winter-Ebmer, R. Potential Unemployment Benefit Duration and Spell Length: Lessons from a Quasi-Experiment in Austria. *Oxford Bulletin of Economics and Statistics*, 60(1):33–45, 1998.

Appendix A

Data and data construction

Table A.1: Detailed information about the elimination process

Task	spells prior to task	eliminated observations	spells after task
elimination of retiremt.	105,757	3,769	101,988
el. of dead	101,988	130	101,858
el. of admin. transitions	101,858	174	101,684
rehabilitation	101,858	52	101,806
temporary allowance	101,806	122	101,684
el. of non comparable	101,684	511	101,173
training/education	101,684	289	101,395
parental leave	101,395	222	101,173
el. of short employment	101,173	34,543	66,630
short target-employment	101,173	20,979	80,194
short source-employment	80,194	13,564	66,630
last restrictions	66,630	27,422	39,208
working age (15-64)	66,630	11	66,619
Austrian citizenship	66,619	14,334	52,285
1^{st} spell	52,285	1,379	50,906
el. shorter than 7 days	50,906	1,131	49,775
el. of missing values	49,775	10,567	39,208
Total	**105,756**	**66,548**	**39,208**

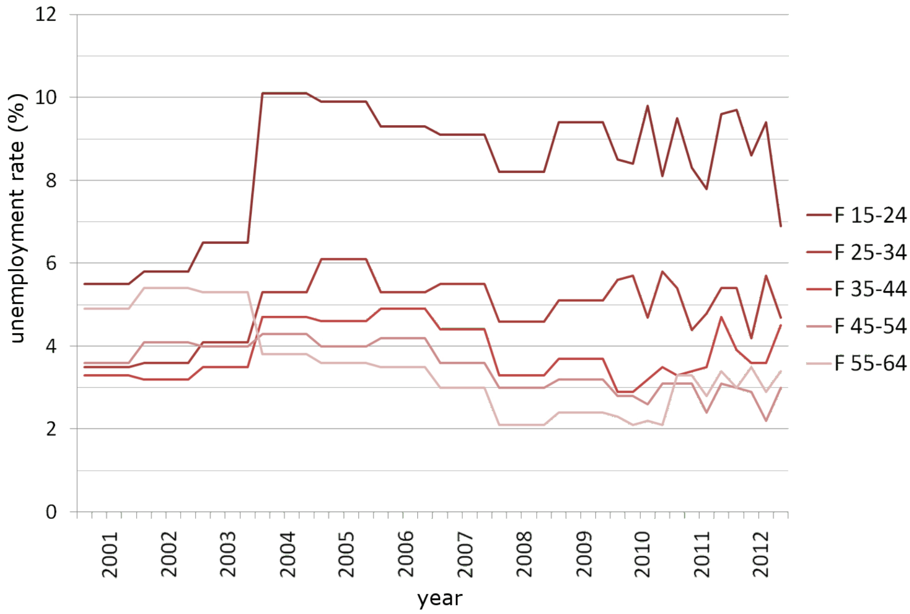
Table A.2: Construction of the covariate *educational level*

ISCED	classification	attribute
ISCED level 2	low	PS - compulsory education PO - no completed educational level
ISCED Level 3	med	MK - secondary business school MS - other secondary school MT - secondary technical school
ISCED Level 3 ISCED level 4	appr	LE - apprenticeship LT - integrated apprenticeship LM - master craftsman
ISCED level 3 ISCED level 4	high	HA - upper secondary school HS - upper secondary school (others) HK - upper secondary school (business) HT - upper secondary school (technical)
ISCED level 5	uni	UB - university degree (Bachelor) UV - university AK - academy FH - university of applied sciences

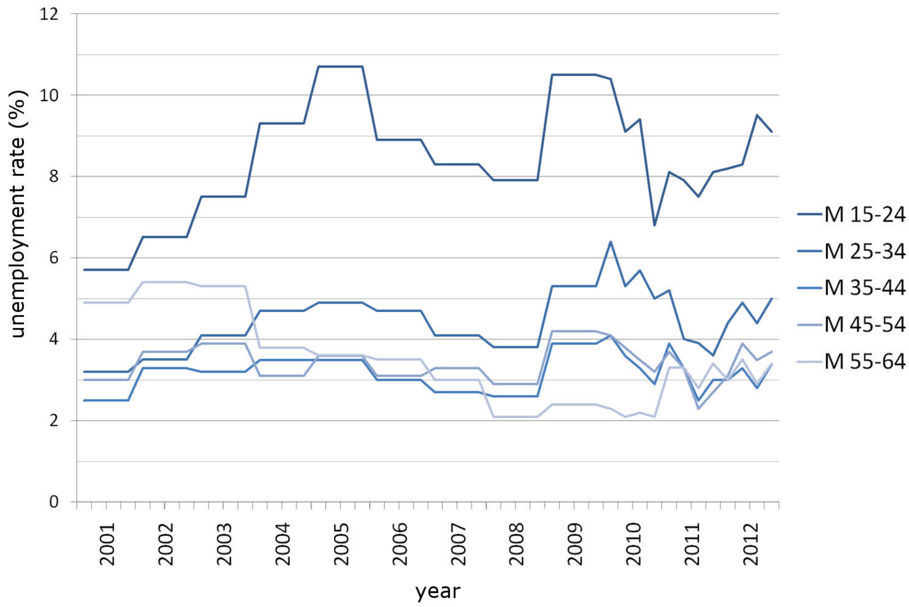
Source:STATISTICS AUSTRIA (2013a)

Table A.3: Numbers and shares of educational levels in the final sample

	Women			Men		
Education	**Freq.**	**Percent**	**Cum.**	**Freq.**	**Percent**	**Cum.**
AK	128	0.79%	0.79%	30	0.13%	0.13%
FH	64	0.39%	1.18%	74	0.32%	0.45%
HA	538	3.31%	4.49%	430	1.87%	2.33%
HK	443	2.73%	7.22%	221	0.96%	3.29%
HS	835	5.14%	12.35%	266	1.16%	4.45%
HT	77	0.47%	12.83%	580	2.53%	6.98%
LE	5,518	33.95%	46.77%	12,509	54.5%	61.47%
LM	57	0.35%	47.12%	250	1.09%	62.56%
LT	-	-	-	1	0.00%	62.57%
MK	683	4.20%	51.33%	246	1.07%	63.64%
MS	980	6.03%	57.35%	400	1.74%	65.38%
MT	47	0.29%	57.64%	176	0.77%	66.15%
PO	276	1.70%	59.34%	444	1.93%	68.08%
PS	6,062	37.29%	96.63%	6,899	30.06%	98.14%
UB	9	0.06%	96.69%	1	0.00%	98.14%
UV	538	3.31%	100.00%	426	1.86%	100.00%
Total	**16,255**	**100.00%**		**22,953**	**100.00%**	



(a) women



(b) men

Figure A.1: Unemployment rates during the observation period

Table A.4: OENACE 2008 - industries

Code	Element
A	agriculture, forestry and fishing
B	mining and quarrying
C	manufacturing
D	electricity, gas, steam and air conditioning supply
E	water supply, sewerage, waste management and remediation activities
F	construction
G	wholesale and retail trade; repair of motor vehicles and motorcycles
H	transportation and storage
I	accommodation and food service activities
J	information and communication
K	financial and insurance activities
L	real estate activities
M	professional, scientific and technical activities
N	administrative and support service activities
O	public administration and defense; compulsory social security
P	education
Q	human health and social work activities
R	arts, entertainment and recreation
S	other service activities
T	activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	activities of extraterritorial organisations and bodies

Source: STATISTICS AUSTRIA

Table A.5: Numbers of changes from source to target industries (OENACE classification), women, Part 1

	Target industry												
Source industry	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**Total**
A	**115**	0	7	0	2	3	14	2	8	1	0	0	**192**
B	0	**6**	0	0	0	0	5	0	0	0	0	0	**11**
C	5	0	**439**	0	5	25	172	25	96	11	9	6	**1,232**
D	0	0	1	**3**	0	2	1	1	0	1	0	0	**17**
E	0	0	3	0	**5**	1	2	1	0	0	0	0	**23**
F	1	0	15	0	0	**216**	28	7	20	1	4	5	**390**
G	14	2	186	4	3	35	**1,438**	48	206	37	29	24	**3,045**
H	0	1	13	0	0	7	48	**230**	25	3	4	2	**447**
I	10	3	124	1	2	22	298	73	**3,772**	8	16	13	**5,060**
J	1	0	4	0	1	1	30	3	10	**62**	1	1	**235**
K	0	0	8	2	0	3	31	4	9	4	**33**	3	**172**
L	0	0	7	0	0	0	16	5	13	1	4	**32**	**145**
M	1	0	41	1	0	19	67	9	22	26	16	15	**648**
N	5	1	74	0	1	12	137	20	88	19	13	9	**1,259**
O	1	0	29	0	0	12	69	10	42	7	13	6	**896**
P	0	0	12	1	1	3	30	4	29	0	3	2	**342**
Q	5	0	26	0	0	8	94	14	37	5	6	5	**1,106**
R	1	0	9	0	0	3	18	0	28	4	3	1	**260**
S	2	0	18	0	3	9	86	5	66	5	2	3	**745**
T	0	0	2	0	0	0	2	0	3	0	0	0	**26**
U	0	0	0	0	0	0	0	0	0	0	0	0	**4**
Total	**161**	**13**	**1,018**	**12**	**23**	**381**	**2,586**	**461**	**4,474**	**195**	**156**	**127**	**16,255**

Table A.6: Numbers of changes from source to target industries (OENACE classification), women, Part 2

	Target industry											
Source industry	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**cens**	**resid**	**Total**
A	2	6	6	1	5	1	2	0	0	**13**	4	**192**
B	0	0	0	0	0	0	0	0	0	**0**	0	**11**
C	35	99	47	9	90	8	23	2	0	**97**	29	**1,232**
D	2	2	1	0	1	0	0	0	0	**2**	0	**17**
E	3	4	0	0	0	0	1	0	0	**2**	1	**23**
F	14	19	9	4	8	1	3	1	0	**22**	12	**390**
G	75	213	110	38	155	21	74	0	0	**257**	76	**3,045**
H	13	26	7	5	15	2	7	0	0	**33**	6	**447**
I	43	134	67	51	101	34	58	3	0	**168**	59	**5,060**
J	26	22	11	8	12	2	13	0	0	**14**	13	**235**
K	3	18	9	2	6	2	7	0	0	**13**	15	**172**
L	5	11	11	2	10	1	6	0	0	**12**	9	**145**
M	**202**	40	34	18	34	8	19	2	1	**44**	29	**648**
N	40	**497**	45	13	81	12	22	2	0	**150**	18	**1,259**
O	32	45	**369**	25	127	7	37	1	0	**43**	21	**896**
P	11	21	20	**131**	24	6	11	0	0	**24**	9	**342**
Q	19	61	88	20	**532**	6	33	3	0	**114**	30	**1,106**
R	13	8	10	6	6	**118**	5	0	0	**22**	5	**260**
S	19	35	44	16	39	10	**307**	0	0	**47**	29	**745**
T	0	0	0	0	5	0	0	**9**	0	**3**	2	**26**
U	0	0	1	0	1	0	1	0	**0**	**0**	1	**4**
Total	**557**	**1,261**	**889**	**349**	**1,252**	**239**	**629**	**23**	**1**	**1,080**	**368**	**16,255**

Table A.7: Numbers of changes from source to target industries (OENACE classification), men, Part 1

	Target industry												
Source industry	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**Total**
A	**118**	0	7	0	1	9	7	19	5	1	0	0	**208**
B	0	**111**	5	0	1	8	1	7	1	0	0	1	**141**
C	11	7	**1,472**	6	8	164	242	86	60	17	21	11	**2,982**
D	0	0	1	**6**	0	1	1	2	0	1	1	0	**26**
E	1	0	5	0	**27**	10	6	4	2	0	0	1	**73**
F	8	6	234	1	12	**5,753**	164	149	45	8	7	12	**7,206**
G	13	3	231	1	8	127	**1,096**	110	86	37	29	12	**2,530**
H	2	4	43	1	2	59	96	**795**	31	5	3	10	**1,327**
I	7	0	57	1	1	42	98	74	**2,045**	8	7	11	**2,759**
J	0	0	23	1	0	3	25	8	3	**95**	4	3	**321**
K	0	0	13	0	0	5	14	8	7	4	**43**	3	**191**
L	1	0	10	0	0	17	11	5	9	1	3	**48**	**148**
M	1	1	40	0	1	41	35	16	14	11	14	5	**522**
N	5	1	196	2	4	178	148	98	66	17	9	12	**2,305**
O	4	1	44	0	3	31	38	15	18	8	10	7	**775**
P	0	0	16	0	0	27	18	6	25	1	1	2	**269**
Q	2	0	51	0	3	36	47	12	28	9	3	3	**615**
R	1	0	10	1	0	2	16	17	14	2	3	0	**256**
S	1	0	16	0	1	16	28	11	16	3	3	0	**289**
T	0	0	0	0	0	0	0	0	2	0	0	1	**9**
U	0	0	0	0	0	0	0	0	0	0	0	0	**1**
Total	**175**	**134**	**2,474**	**20**	**72**	**6,529**	**2,091**	**1,442**	**2,477**	**228**	**161**	**142**	**22,953**

Table A.8: Numbers of changes from source to target industries (OENACE classification), men Part 2

	Target industry											
Source industry	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**cens**	**resid**	**Total**
A	4	11	3	2	0	1	3	0	0	**12**	5	**208**
B	1	0	1	1	1	1	0	0	0	**1**	0	**141**
C	68	345	86	21	68	14	20	1	0	**145**	109	**2,982**
D	2	4	3	1	0	0	0	0	0	**3**	0	**26**
E	1	9	2	0	1	0	0	0	0	**3**	1	**73**
F	36	275	44	25	47	10	14	0	0	**251**	105	**7,206**
G	36	256	68	21	64	21	34	1	0	**146**	130	**2,530**
H	13	94	24	6	8	3	8	0	0	**86**	34	**1,327**
I	17	103	23	25	34	28	20	0	0	**88**	70	**2,759**
J	24	33	9	1	6	10	8	0	0	**30**	35	**321**
K	12	17	9	1	8	4	2	0	0	**18**	23	**191**
L	4	17	1	0	4	0	1	0	0	**8**	8	**148**
M	**172**	41	15	12	14	4	12	0	0	**28**	45	**522**
N	35	**1,161**	39	18	46	15	21	1	0	**188**	45	**2,305**
O	25	58	**356**	18	56	4	15	1	0	**45**	18	**775**
P	5	21	12	**76**	16	6	5	0	0	**15**	17	**269**
Q	13	81	37	9	**151**	4	17	0	0	**94**	15	**615**
R	4	12	6	8	2	**131**	4	0	0	**13**	10	**256**
S	9	12	22	3	20	2	**96**	0	0	**19**	11	**289**
T	0	1	0	0	0	0	0	**4**	0	**0**	1	**9**
U	0	0	0	0	0	0	0	0	**1**	**0**	0	**1**
Total	**481**	**2,551**	**760**	**248**	**546**	**258**	**280**	**8**	**1**	**1,193**	**682**	**22,953**

Appendix B

Logrank tests

Appendix B includes the logrank tests for equality of survivor functions for categorical variables used in the analysis.

Table B.1: Logrank test, *age group*

Variable	**Women**		**Men**	
age group	Events observed	Events expected	Events observed	Events expected
15-24	3,967	3,768.01	5,348	5,862.14
25-34	3,799	4,135.69	6,008	5,525.29
35-44	4,037	3,967.52	5,473	5,168.72
45-54	2,824	2,794.29	3,688	3,863.99
55-64	414	375.49	952	1,048.86
Total	**15,041**	**15,041**	**21,469**	**21,469**
	chi2(4) = 43.76		chi2(4) = 123.68	
	Pr>chi2 = 0.0000		Pr>chi2 = 0.0000	

Table B.2: Logrank test, *receipt of benefits*

Variable	**Women**		**Men**	
paid	Events observed	Events expected	Events observed	Events expected
0	929	1,022.43	903	1,060.60
1	14,112	14,018.57	20,566	20,408.40
Total	**15,041**	**15,041**	**21,469**	**21,469**
	chi2(1) = 9.25		chi2(1) = 25.00	
	Pr>chi2 = 0.0024		Pr>chi2 = 0.0000	

Table B.3: Logrank test, *educational level*

Variable	Women		Men	
edu	Events observed	Events expected	Events observed	Events expected
low edu	5,715	6,141.29	6,718	7,523.50
appr	5,253	4,653.08	12,122	10,293.40
med edu	1,609	1,540.38	774	846.95
high edu	1,771	1,934.92	1,366	2045.47
uni	693	771.34	489	759.67
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(4) = 133.09 chi2(4) = 764.30
Pr>chi2 = 0.0000 Pr>chi2 = 0.0000

Table B.4: Logrank test, *marginal employment*

Variable	Women		Men	
marginal	Events observed	Events expected	Events observed	Events expected
0	12,632	11,031.88	19,395	17,742.68
1	2,409	4,009.12	2,074	3,726.32
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(1) = 898.66 chi2(1) = 915.46
Pr>chi2 = 0.0000 Pr>chi2 = 0.0000

Table B.5: Logrank test, *number of children*

Variable	Women	
children	Events observed	Events expected
0	12,454	11,971.61
1	1,871	62,168.18
2	716	901.20
Total	**15,041**	**15,041**

chi2(2) = 99.01
Pr>chi2 = 0.0000

Table B.6: Logrank test, *federal state of previous employment*

Variable state	Women Events observed	Women Events expected	Men Events observed	Men Events expected
Bgld	301	312.71	611	616.18
Car	1,471	1,256.73	2,259	1,868.55
Lo Aus	1,991	2,318.95	3,336	3,403.08
Up Aus	1,943	2,082.19	3,150	2,926.72
Sbg	1,442	1,089.78	1,922	1,617.81
Styria	1,940	2,015.20	3,133	3,073.20
Tyrol	2,445	1,532.37	2,530	1,824.31
Vbg	600	591.20	744	818.44
Vienna	2,908	3,841.88	3,784	5,320.72
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(8) = 1009.28 chi2(8) = 911.00
Pr>chi2 = 0.0000 Pr>chi2 = 0.0000

Table B.7: Logrank test, *quarter*

Variable quarter	Women Events observed	Women Events expected	Men Events observed	Men Events expected
1.quarter	3,026	3,338.81	6,126	5,238.86
2.quarter	3,815	3,582.93	3,442	3,645.68
3.quarter	3,610	3,855.00	3,522	4,520.46
4.quarter	4,590	4,264.26	8,379	8,064.01
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(3) = 85.58 chi2(3) = 403.92
Pr>chi2 = 0.0000 Pr>chi2 = 0.0000

Table B.8: Logrank test, *sector of previous employment*

Variable sector	Women Events observed	Women Events expected	Men Events observed	Men Events expected
Primary	176	169.41	194	192.95
Secondary	1,534	1,733.48	9,868	8,443.49
Tertiary	13,327	13,134.12	11,406	12,831.57
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(2) = 26.22 chi2(2) = 409.88
Pr>chi2 = 0.0000 Pr>chi2 = 0.0000

Table B.9: Logrank test, *industry of previous employment*

Variable	Women		Men	
industry	Events observed	Events expected	Events observed	Events expected
constr	362	384.27	6,830	5,227.11
farm	176	169.38	194	192.93
manu	1,125	1,291.84	2,808	3,022.88
pub	849	906.04	723	873.64
soc	978	1,238.69	517	953.63
tour	4,872	3,277.25	2,661	2,104.49
trade	2,756	3,304.71	2,354	2,827.62
others	3,923	4,468.82	5,382	6,266.70
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(7) = 1047.38 chi2(7) = 1138.36
Pr>chi2 = 0.0000 Pr>chi2 = 0.0000

Table B.10: Logrank test, *industry size of the industry of previous employment*

Variable	Women		Men	
size	Events observed	Events expected	Events observed	Events expected
small	250	258.96	811	852.06
medium	2,523	2,725.81	4,645	5,065.70
big	12,268	12,056.23	16,013	15,551.23
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(2) = 19.25 chi2(2) = 51.32
Pr>chi2 = 0.0001 Pr>chi2 = 0.0000

Table B.11: Logrank test, *labor turnover (industry of previous employment)*

Variable	Women		Men	
turnover	Events observed	Events expected	Events observed	Events expected
low turn	6,667	7,776.00	14,793	14,538.03
med turn	1,687	1,909.96	1,260	1,697.16
high turn	6,687	5,355.05	5,416	5,233.81
Total	**15,041**	**15,041**	**21,469**	**21,469**

chi2(2) = 526.01 chi2(2) = 124.95
Pr>chi2 = 0.0000 Pr>chi2 = 0.0000

Table B.12: Logrank test, *number of seasonal peaks at the industry of previous employment*

Variable	**Women**		**Men**	
peaks	Events observed	Events expected	Events observed	Events expected
0	7,847	9,103.10	9,632	11,186.49
1	931	1,002.85	7,432	5,975.51
2	6,263	4,935.05	4,405	4,307.00
Total	**15,041**	**15,041**	**21,469**	**21,469**

	Women	Men
	chi2(2) = 546.26	chi2(2) = 594.83
	Pr>chi2 = 0.0000	Pr>chi2 = 0.0000

Appendix C

Proportionality assumption

The test for proportionality of hazards has been done via the Schoenfeld residuals (Schoenfeld, 1982). The aim is to compute a residual for each covariate for each individual, which is defined as the difference between the observed value of X and its conditional expectation given the risk set $R(t)$. The proportionality assumption is said to be fulfilled if the test comes back insignificant. Both Tables C.1 and C.2 show significance for the overall model, meaning that the assumption of proportional hazards is *not* fulfilled.

Table C.1: Test of the proportionality assumption: women

Variable	rho	chi2	df
35-44	.	.	.
15-24	0.005	0.34	1
25-34	0.003	0.10	1
45-54	-0.004	0.23	1
55-64	0.003	0.14	1
high edu	.	.	1
low edu	-0.008	0.99	1
appr	-0.015†	3.58	1
med edu	-0.008	0.97	1
uni	-0.006	0.62	1
paid	0.001	0.02	1
child	0.044**	28.89	1
urate	0.002	0.06	1
marginal	0.134**	256.12	1
prevep	0.005	0.35	1
1.quart	.	.	1
2.quart	-0.005	0.35	1
3.quart	0.020*	5.84	1
4.quart	0.7	0.79	1
Vienna	.	.	1
Burgenland	0.001	0.02	1
Carinthia	-0.023**	7.63	1
Lower Austria	-0.012	2.17	1
Upper Austria	-0.026**	10.11	1
Salzburg	-0.041**	25.27	1
Styria	-0.009	1.11	1
Tyrol	-0.036**	19.28	1
Vorarlberg	-0.030**	13.66	1
low turnover	.	.	1
med turnover	-0.003	0.10	1
high turnover	-0.035**	16.7	1
nopeak	.	.	1
1 peak	0.003	0.13	1
2 peaks	-0.009	1.18	1
global test	443.17**		28

Table C.2: Test of the proportionality assumption: men

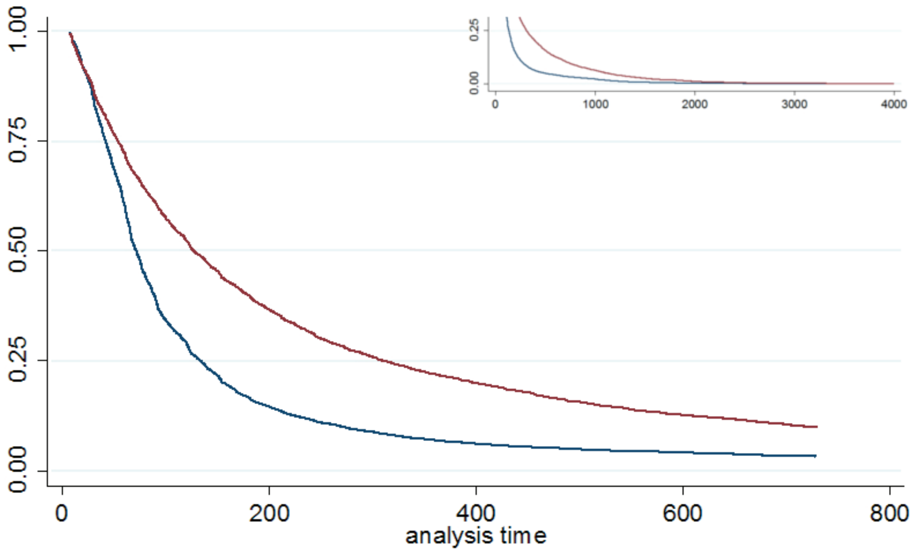
Variable	rho	chi2	df
35-44	.	.	.
15-24	0.024**	11.88	1
25-34	-0.001	0.04	1
45-54	-0.014*	3.99	1
55-64	-0.009	1.81	1
high edu	.	.	1
low edu	-0.028**	16.31	1
appr	-0.028**	17.00	1
med edu	-0.011	2.43	1
uni	0.011	2.65	1
paid	0.037**	30.48	1
urate	-0.003	0.18	1
marginal	0.120**	300.39	1
1.quart	.	.	1
2.quart	0.009	1.74	1
3.quart	0.038**	31.29	1
4.quart	0.050**	53.52	1
Vienna	.	.	1
Burgenland	-0.005	0.47	1
Carinthia	-0.021**	9.42	1
Lower Austria	-0.021**	9.51	1
Upper Austria	-0.032**	21.79	1
Salzburg	-0.031**	20.55	1
Styria	-0.025**	12.87	1
Tyrol	-0.031**	20.08	1
Vorarlberg	-0.035**	26.95	1
low turnover	.	.	1
med turnover	-0.006	0.85	1
high turnover	-0.023**	10.36	1
no peak	.	.	1
1 peak	0.001	0.04	1
2 peaks	-0.017*	5.92	1
small industry	.	.	1
med industry	-0.009	1.78	1
big industry	-0.013†	3.64	1
global test	610.02**		29

Appendix D

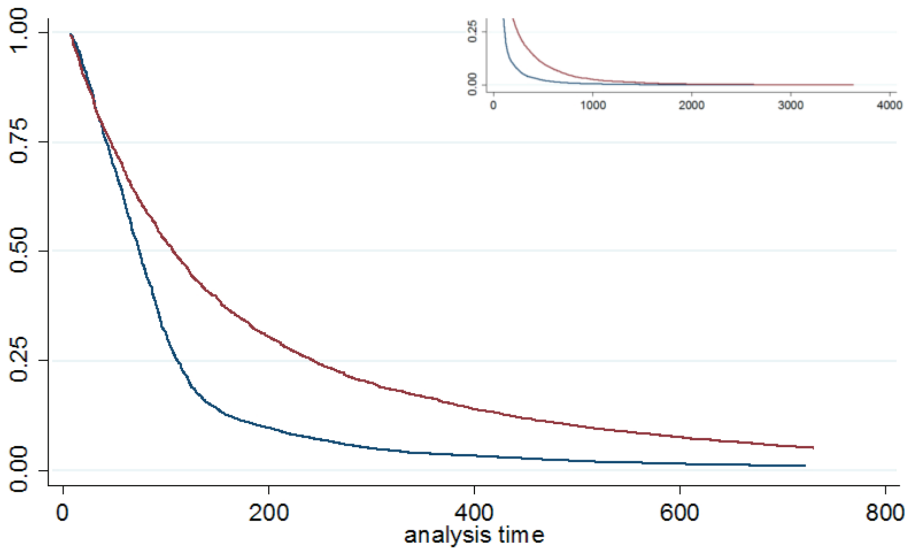
Additional Kaplan-Meier estimates

Appendix D contains the Kaplan-Meier survival estimates for the covariates of inter- and intra-industry changes and change of employer (Figures D.1 and D.2). Only completed spells are included, thus these estimates are not part of the initial analysis. It can be obtained that for both individuals staying in the industry or staying with an employer has a positive effect, i.e. the estimated Kaplan-Meier survival curves are lower.

Figure D.3 gives the Kaplan-Meier survival estimates for the federal states Burgenland, Carinthia, Lower Austria, Upper Austria, Salzburg, Styria, Tyrol, Vorarlberg and Vienna. The most interesting findings are that for both men and women OOW-spells that originate from Tyrol have the lowest survival curve whereas Vienna has the highest. After one year of unemployment about 27.5% (22%) of women (men) in Vienna are still unemployed while for Tyrol only 9% (about 4%) of women (men) are still without employment.



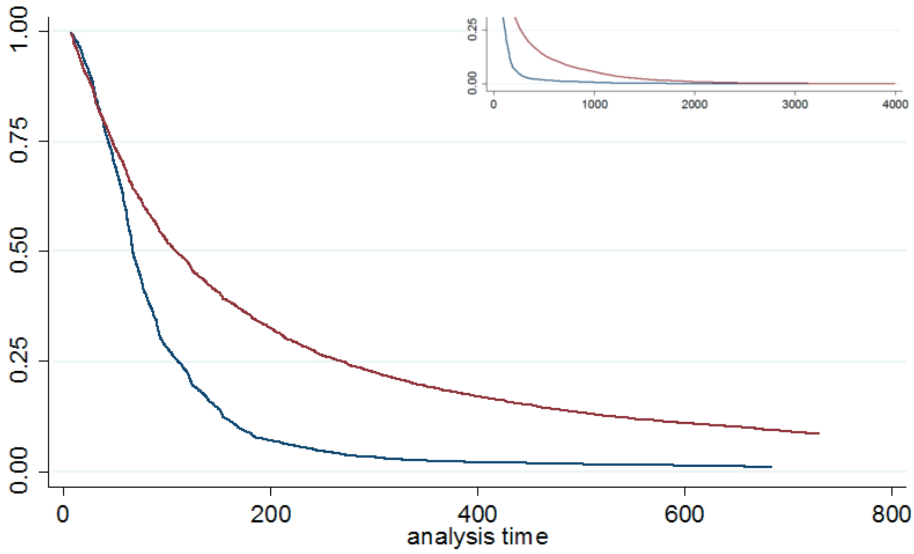
(a) women



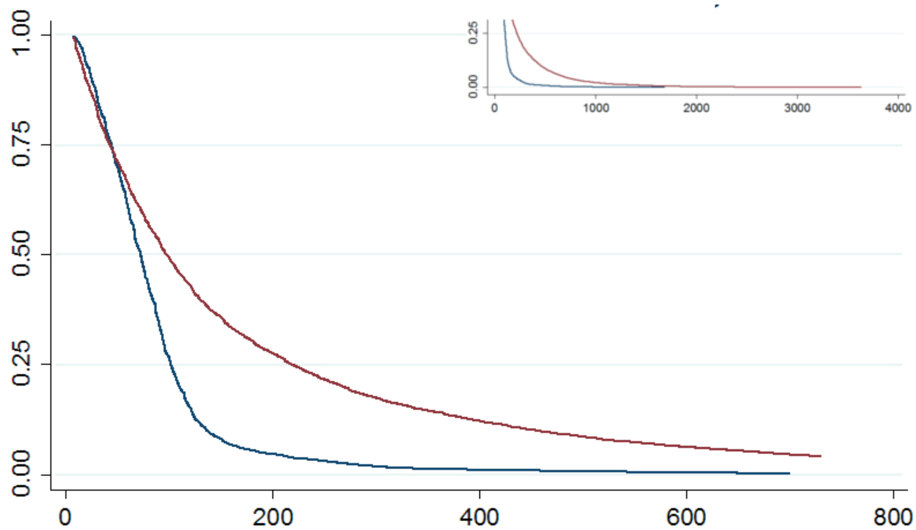
(b) men

inter = 0 inter = 1

Figure D.1: Kaplan-Meier survival estimates by change of industry



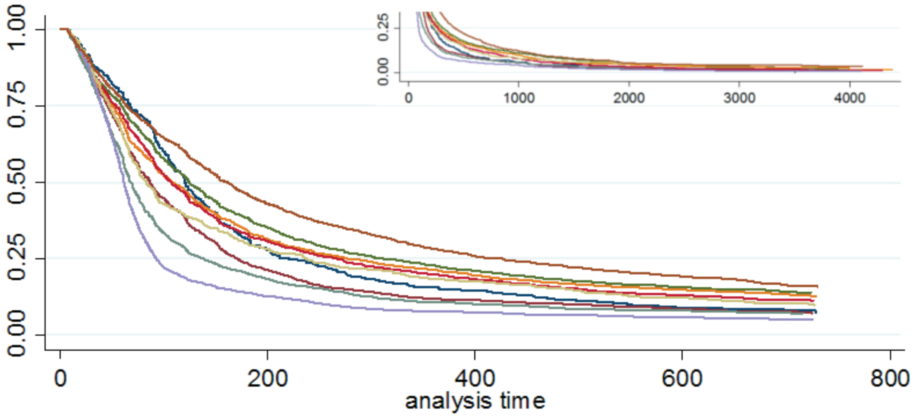
(a) women



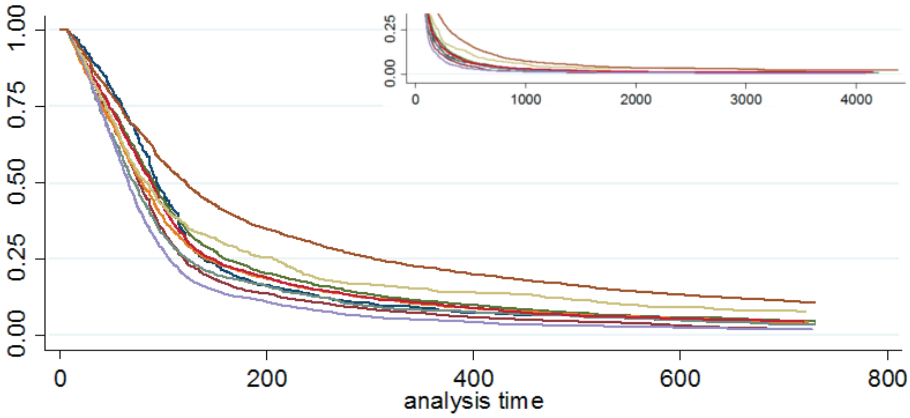
(b) men

new emp = 0 new emp = 1

Figure D.2: Kaplan-Meier survival estimates by change of employer



(a) women, 2 years



(b) men, 2 years

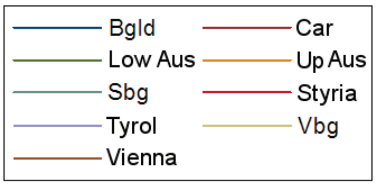


Figure D.3: Kaplan-Meier survival estimates by federal state

Abstract

In this thesis the focus, lies on out-of-work (OOW) spells in a sample of the Austrian labor market. Especially the inclusion of industry-specific characteristics (seasonal peaks, labor turnover and industry size) should shed light on new aspects influencing the duration of non-working spells. In contrast to the majority of studies, the analyses are based on linked employer-employee data. This database, which contains information about labor market active and labor market inactive statuses on a daily basis, is representative of the Austrian population. Duration analysis, which allows for the inclusion of censored spells, is used to measure the impact of certain characteristics on the time spent out of work. To identify differences in the survival curves corresponding to different values of a characteristic, the Kaplan-Meier estimate is used. For a more detailed analysis of the effects of covariates on the duration of the out-of-work stay, semiparametric (Cox Proportional Hazards model) and parametric methods (Weibull and Loglogistic model) were applied.

Nonparametric analysis showed that the time spent out of work seems to be shorter for men compared to women. While the Kaplan-Meier survival estimates by receipt of benefit payments are very similar for men and women, differences were found in the covariates age and educational level. Especially women aged 25 to 44 seem to have longer out-of-work spells than other women.

The inclusion of industry-specific factors such as number of seasonal peaks, labor turnover and industry size of the previous employment showed that women from industries with two seasonal peaks have shorter OOW-spells, while men have shorter spells in industries with one seasonal peak. These effects are also related to the industries in which men and women in the sample were predominantly employed in. In addition, it has been found that in industries with high labor turnover, the probability of finding employment is higher. The survival curves by industry size did not show big differences in their course.

The comparison of the semiparametric Cox Proportional Hazards model and parametric models showed that the assumption of nonmonotonic baseline hazard (Loglogistic model) is preferable; the employment probability rises at first and decreases thereafter with increasing out-of-work duration. Parametric analysis supported the results for the effect of seasonal peaks from the Kaplan-Meier estimates. Concerning the rate of exchange of employees per year-round employment the out-of-work stay has been found to be shorter in industries with high labor turnover.

Zusammenfassung

Diese Arbeit befasst sich mit nicht-erwerbstätigen Episoden (OOW) am österreichischen Arbeitsmarkt. Insbesondere der Einbezug branchenspezifischer Eigenschaften (Personalwechsel, saisonale Schwankungen und Branchengröße) soll einen weiteren Aspekt in den Einflussfaktoren der Dauer nicht-erwerbstätiger Episoden beleuchten. Im Gegensatz zu einem Großteil der Literatur liegen den Analysen Administrativdaten zugrunde. Diese Datenbasis, die die tagegenaue Erfassung von Erwerbskarrieren aller in Österreich Beschäftigter ermöglicht, kann die Bevölkerung gut abbilden. Um den Einfluss bestimmter Charakteristika auf die Episodendauer zu bestimmen, wurde sich der Methode der Ereigniszeitanalyse bedient, die ermöglicht auch zensierte Fälle miteinzubeziehen. Um Unterschiede in den Überlebenskurven nach Ausprägungen verschiedener Variablen darzustellen, wird der Kaplan-Meier Schätzer herangezogen, während zur Schätzung des Ausmaßes des Einflusses aller Variablen das Cox Regressionsmodell sowie das Weibull und das Loglogistic Modell verwendet werden.

Schon in der nicht parametrischen Analyse konnte gezeigt werden, dass die Zeit außerhalb einer Erwerbstätigkeit bei Männern kürzer andauert als bei Frauen. Während der Verlauf der Kaplan-Meier Überlebenskurven nach Leistungsbezug innerhalb der Erwerbslosigkeit nach Geschlecht relativ ähnlich ist, konnten Unterschiede in den Einflussgrößen Alter und Bildung erkannt werden. Vor allem Frauen im Alter von 25 bis 44 Jahren scheinen längere nicht-erwerbstätige Episoden zu haben als andere weibliche Individuen.

Der Einbezug von saisonalen Spitzen, Personalwechsel und Branchengröße der vorangegangenen Tätigkeit konnte zeigen, dass Frauen vor allem in Branchen mit zwei Beschäftigungsspitzen kürzere Arbeitslosendauern aufweisen, respektive Männer in jenen mit einer Spitze. Diese Effekte lassen sich in Zusammenhang mit den Branchen bringen, in denen Frauen und Männer vorwiegend vor Beginn der nicht Arbeitsmarkt-aktiven Episode beschäftigt waren. Darüber hinaus konnte gezeigt werden, dass in Wirtschaftsklassen mit hohem Personalwechsel die Wahrscheinlichkeit eine Erwerbstätigkeit zu finden höher ist. Die Überlebenskurven nach Branchengröße konnten keine großen Unterschiede zeigen.

Beim Vergleich des Cox Regressionsmodells und der parametrischen Modelle konnte festgestellt werden, dass die Annahme einer nicht monotonen Hazardrate (Loglogistic Modell) zu bevorzugen ist; die Wahrscheinlichkeit Arbeit zu finden steigt zuerst an und nimmt später mit zunehmender Erwerbslosigkeitsdauer ab. Ähnlich den nicht parametrischen Schätzern, konnte auch in der parametrischen Analyse eine kürzere Dauer für Branchen mit starkem Personalwechsel gezeigt werden und auch für die Saisonspitzen konnten ähnliche Ergebnisse gefunden werden.

CURRICULUM VITAE

PERSONAL INFORMATION

Name:	Sophie Oberhauser, Bakk.rer.soc.oec.

EDUCATION

Nov 2011 - present	Institute for Advanced Studies (IHS) *Student assistant - Applied Research* *Economics and Finance*
Oct 2011 - present	Vienna University *Master's Programme (Economics)*
Nov 2011	Bachelor's degree
Oct 2008 - Feb 2009	University Paris 1 - Panthéon Sorbonne *Erasmus*
Feb 2008	Scholarship
Oct 2006 - Oct 2011	Vienna University *Bachelor's Programme (Economics)*