Unemployed in Germany: Factors Influencing the Risk of Losing the Job

Nina Westerheide[1] & Göran Kauermann[2]

[1] Department of Business Administration and Economics, Bielefeld University, Germany

[2] Department of Statistics, Ludwig-Maximilians-University Munich, Germany

Correspondence: Nina Westerheide, Department of Business Administration and Economics, Bielefeld University, Postfach 100131, 33501 Bielefeld, Germany. E-mail: nwesterheide@wiwi.uni-bielefeld.de

Received: June 15, 2014 Accepted: July 21, 2014 Online Published: July 31, 2014

doi:10.5430/rwe.v5n2p43 URL: http://dx.doi.org/10.5430/rwe.v5n2p43

Abstract

Unemployment is a central issue in modern economies and its analysis and investigation consists of two aspects. First, what is the risk of getting unemployed based on economic, local and individual characteristics and second, what is the chance of getting reemployed. In this paper we focus on the first question by making use of the massive database from the German Federal Employment Agency (IABS Scientific Use File 'Regional File 1975 – 2004') to model the risk of an individual to become unemployed between 2000 and 2004 in Germany. As individual covariates we include gender, age and education as fixed effects in our model. Beside these individual characteristics, regional as well as calendrical and economic information is considered and included as smooth functional effects in the model. As result of our data analysis we uncover strong educational and age specific effects as well as dominating calendrical and spatial effects on the individual's risk of getting unemployed.

Keywords: unemployment, log-linear poisson model, additive poisson model, P-spline smoothing

1. Introduction

A well-known problem in economies and a focal point in economic research is unemployment, see for example Layard, Nickell, and Jackman (2009), Blanchard (2006) or Ljungqvist and Sargent (1998). Often the unemployment rate is used as a macroeconomic measure to explain changes in regional and national labour markets, as done, for instance, in official statistics in European Commission (2009), Eurostat (2009) or OECD (2009). The rate of unemployment is determined by two factors, first the risk of getting unemployed and second the chances for reemployment after losing the job. The latter can be measured by the the duration of unemployment which is of utmost interest to explain the unemployment behaviour of individuals for different points of focus, see for example Narendranathan and Stewart (1993), Böheim and Taylor (2000), Bover, Arellano, and Bentolila (2002), Røed and Zhang (2003), Lauer (2003), Tatsiramos (2009) or Westerheide and Kauermann (2012). Analyses regarding only the unemployment duration in Germany include Hunt (1995), Steiner (1997, 2001) or Fahrmeir, Lang, Wolff, and Bender (2003). Beside the duration of unemployment, the risk of getting unemployed is also of high interest to better grasp the reasons of the unemployment rate's hight. Unemployment risk is defined and analysed in different ways and in different contexts. Galiani and Hopenhayn (2003), for instance, analysed the risk of unemployment in Argentina between 1989 and 1998 making use of hazard models. Covizzi (2008) determined the unemployment risk of Swiss individuals concerning union dissolution, health, and gender with Cox proportional hazard models. Thapa (2004) and Arai and Vilhelmsson (2004) explored the unemployment risk of immigrants to natives in Australia and Sweden, respectively. Both used a logistic regression model. Hammer (1997) utilised logistic and Poisson regression models to investigate the unemployment risk of young Norwegian individuals. Fieldhouse (1996) looked at social and geographical factors to investigate the unemployment risk in Great Britain using logistic regression models after looking at factor-specific unemployment rates. Regarding the different papers analysing the risk of unemployment in Germany, we refer to Reinberg and Hummel (2002, 2003, 2005) who used qualification-specific unemployment rates to analyse the unemployment risk in different educational groups in Germany. Arrow (1996) analysed the impact of health on the unemployment risk by using, amongst others, a Cox's proportional hazard model. Wilke (2004) analysed -beside the unemployment duration- the risk of unemployment given employment in Germany, that means he looked at the ratio of the number of individuals getting unemployed and the number of employed individuals in a defined period and compared the results with the unemployment rate. Lauer (2003) analysed the influence of education on the risk of getting unemployed and reemployed in a cross-national study with a discrete time competing risks hazard rate model based on the data of the German Socio-Economic Panel Study for Germany and the Emploi survey for France. Lurweg (2010) used a pooled logistic regression to analyse amongst others the impact of international trade on the risk of getting unemployed. A more recent article on this topic is from Wichert and Wilke (2012). Beside an extensive analysis of the misclassification of the individual's educational background and citizenship in German administrative data, they analyse the transition from employment into unemployment making use of data from the original administrative data set IABS and a corrected version adjusted for education and citizenship misclassification. Some of the papers mentioned above include regional information in their analysis as, for example, Fieldhouse (1996) who used geographical factors concerning different British regions or Thapa (2004) whose analysis contains Australian regions. However, none of the papers above that analyse the unemployment risk in Germany include regional or spatial information apart from differentiating between the old West German states and the newly formed German states, see Reinberg and Hummel (2002, 2003, 2005) or Lurweg (2010). In addition, often only data of the Old Länder is used for analysing the risk of unemployment in Germany, see for instance Wilke (2004), Lauer (2003) or Wichert and Wilke (2012).

With our analysis we aim to contribute to the discussion in two aspects. The first contribution of our paper is to analyse the influence of different covariate effects -including spatial effects comprising the reunified Germany as well as individual, calendrical, and economic effects- on the unemployment risk in all of Germany between 2000 and 2004. Beside the analysis of different unemployment risks, we want to compare our results with other research findings on unemployment risks. The results we want to contrast with the conclusions of studies investigating the duration of unemployment or analyses interpreting unemployment rates which have been analysed far more thoroughly than the risk of getting unemployed. Our second contribution is to show a possible way to illustrate the spatial risk of getting unemployed in Germany by using available software to model and easily fit an additive Poisson model with fixed grouped individual covariate effects and smooth dynamic covariate effects of spatial, calendrical, and economic information.

As database we use the Scientific Use File ‘Regional File 1975-2004’ of the IAB Employment Samples (IABS) which is an administrative data set of the German Federal Employment Agency and provided by the Research Data Centre (Forschungsdatenzentum (FDZ)) at the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung (IAB)). The IAB Employment Samples were already used in context of analysing unemployment risk and durations in Germany, see for instance Wichert and Wilke (2012), Fahrmeir, Lang, Wolff, and Bender (2003) and Kauermann and Westerheide (2012). The Scientitic Use File ‘Regional File 1975-2004’ contains information about the employment biographies of employees covered by social security and of benefit recipients in Germany on a day-to-day basis. Furthermore, it includes spatial information about 343 defined regions. The database is a 2% random sample out of the Employee and Benefit Recipient History of the IAB.

The statistical model being used for our analysis is built upon the log-linear Poisson model, see McCullagh and Nelder (1989). We allow for grouped covariates to simplify the model and to downsize the computational effort. Beside grouped covariates with individual information like gender, age, and education, we include smooth functional effects as proposed in Hastie and Tibshirani (1990) for generalized additive models. The additive Poisson model is fitted with software for generalized additive models in **R**, see R Development Core Team (2009) and Wood (2006).

The paper is organized as follows. Section 2 introduces the statistical model being used. Section 3 gives more detailed information of the database and the utilised covariates. In Section 4, a detailed data analysis is given before we draw our conclusions in Section 5.

2. Statistical Model

We consider aggregated data on a monthly level listing the employment status of individuals. Note that unemployment often results on a monthly basis since contracts are cancelled to the end of a month. We may therefore define the random variables Y_{ti} giving the employment status of the *i*th individual in time interval (month) $t,\ t \in \{1, \dots, T\}$ with $i = 1, \dots, N_t$. With $Y_{ti} = 1$ we denote an individual who is unemployed in month *t*, but has been working in the previous month, otherwise we set $Y_{ti} = 0$. In other words, $Y_{ti} = 1$ indicates individuals getting unemployed from period $t-1$ to period t. We model the occurence of unemployment as Poisson process. Though this assumes independence of the employment history of the individuals, we favour the Poisson process model since the data base does not cover detailed information about the employment history of each individual. In fact, besides some individual covariates in education, age and gender the data do not provide resilient information how long individuals have been employed uninterruptedly. Therefore, the Poisson process as statstical model for analysing the data seems practical so that we assume that Y_{ti} are independent and identically Poisson-distributed with individual and time specific intensity parameter $\lambda_{ti} = \exp(\eta_{ti})$. The linear predictor η_{ti} depends on a number of covariates

x_{ti}, say, and a set of parameters θ to be specified later. The log-likelihood contribution for time point t can be written as (see McCullagh and Nelder, 1989)

$$l_t(\theta) = \sum_{i=1}^{N_t} [Y_{ti} \log(\lambda_{ti}) - \lambda_{ti}] \tag{1}$$

In our example the number of observations N_t at each time point is rather large, in the order of 500,000 observations, summing up to 29,978,674 observations for all time points. In contrast, the number of events, that are observations with $Y_{ti} = 1$, is comparably small, about 4,000 observations for each time point summing up to 237,507 observations for all time points. Hence, about 1% of the individuals become unemployed. To handle the data in a numerically efficient way, we therefore restructure the likelihood by grouping observations with respect to their covariate values. First, we group the covariate age into $J = 4$ groups, second the educational level is grouped into $L = 3$ categories and for gender we have $K = 2$ groups. Moreover, in our database we have $T = 60$ time intervals, each representing a month, which run from January 2000 to December 2004. Let now N_{tjkl} denote the total number of observations in the specified group categories and let n_{tjkl} be the number of events in age group $j, j = 1, \dots, J$, gender $k, k = 1, 2,$ and educational group $l, l = 1, \dots, L$, in interval $t, t = 1, \dots, T$. Within the particular time dependent groups we assume a homogeneous Poisson process which simplifies the likelihood as follows. Let I_t be the index set of individuals becoming unemployed in t, that is $I_t = \{i : 1 \leq i \leq N_t, Y_{ti} = 1\}$. We define with $o_{ti} = \log(N_{tj_ik_il_i}) / n_{tj_ik_il_i}$ the offset for $i \in \mathrm{I}_t$, where j_i, k_i and l_i denote the category level of individual i. Then, the log-likelihood (1) can be written as

$$l_t(\theta) = \sum_{i \in I_t} [\log(\lambda_{ti}) - \lambda_{ti} \exp(o_{ti})] \tag{2}$$

Note that the likelihood now consists only of the individuals for which we observe the event of getting unemployed and hence it is numerically manageable. The implicit assumption is that covariates not included in the grouping have the same effect amongst all individuals in the groups. Beside the grouped covariates mentioned above we include further covariates in our model which are on a regional, calendrical, and economic information level. These are the location of the former working place, the date when the unemployment began (month and year) and the duration of unemployment during the last year. The effects of these covariates will be modeled by smooth functions while the grouped covariates will be included as fixed effects in our model. This leads us to a generalized additive Poisson model (see Hastie and Tibshirani, 1990 or Wood, 2006). Let now u_{ti} denote the duration of unemployment of individual i in the last year before unemployment in t and let s_{ti} be the location of the region where individual i gets unemployed, that is the coordinates of the centroid of the corresponding region. Note that in regions with more inhabitants or employees we can expect more individuals to get unemployed, so that we need to control for the region size by including the region as covariate. We therefore model the relative intensity λ_{ti} / E_{ti} with E_{ti} as number of employees subject to social insurance contribution in the corresponding year in the region where individual i gets unemployed. To be specific, we model the predictor $\eta_{ti} = \log(\lambda_{ti})$ in (2) to take the form

$$\eta_{ti} = \boldsymbol{x}_{ti}^T \boldsymbol{\beta} + \delta(t) + z_{ti} \xi(u_{ti}) + \phi(s_{ti}) + \log(E_{ti} / \bar{E}) \tag{3}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ are the parameters to be estimated, $\boldsymbol{x}_{ti} = (1, x_{ti1}, \dots, x_{tip})^T$ are the corresponding covariates, z_{ti} indicates if the individual was umemployed in the year before $(z_{ti} = 1)$ or not $(z_{ti} = 0)$ and $\bar{E}$ is the average number of employees per region. Moreover, $\delta(.)$ is the smooth calendrical effect of the beginning of unemployment, $\xi(.)$ specifies the smooth effects of unemployment during the last year -including only the information of those who really became unemployed during that time- and finally $\phi(.)$ describes a smooth spatial effect. We can fit model (3) by replacing the smooth functions by spline bases which are fitted in a penalized form. We refer to the Appendix for details.

3. Data Description

For our analysis we use the Scientific Use File 'Regional File 1975-2004'. A detailed description of the entire database is provided in Drews (2008). We use data from 5 years from January 2000 to December 2004 and analyse the risk of getting unemployed for 91625 men (146548 events in all time intervals out of 16715859 observations from 383769 men in all time intervals) and 66609 women (90959 events in all time intervals out of 13262815 observations from 317066 women in all time intervals) who became unemployed during the considered time. More information is shown in Table 1.

Table 1. Distribution of the events and the total amount of observations for the IAB 'Regional File 1975-2004' for men and women separated by the year the individual became unemployed

Year	Men		Women	
	Events	Observations	Events	Observations
2000	25718	3406426	16707	2623568
2001	25842	3426328	16719	2679894
2002	30863	3366198	18447	2685901
2003	32096	3283558	19724	2644449
2004	32029	3233349	19362	2629003
Σ	146548	16715859	90959	13262815

The covariate age is categorized into: up to 30 years, between 30 and 39 years (reference category), between 40 and 49 years, and 50 years of age and over. Following Wichert and Wilke (2012), we divide the educational background during the last period of employment into three groups to avoid missclassification within the categories of the second and third educational group. However, we do not include the category 'missing' as the aforementioned authors did. The first group consists of individuals without vocational training. The second group (reference category) contains individuals who attended a secondary general school or intermediate secondary school and successfully completed a vocational training and individuals with A-levels and with or without vocational training. The third group includes graduates from a university or comparable. Individuals with missing information on their educational background are excluded unless there is educational information before the required time point of getting unemployed. In this case, the individual's highest educational level until that time point is used as educational background. The data set also contains local information with the region of the workplace. All in all, there are 343 defined regions in Germany in the data set. We use the centroid of the corresponding region as spatial information. As calendar time we use the date (month and year) when the individual became unemployed. The duration of unemployment (in days) during the last year before unemployment is included as well. Similar covariates were also used in analyses of unemployment risks and durations to which we want to compare our results, see for example Wichert and Wilke (2012), Fahrmeir, Lang, Wolff, and Bender (2003) and Kauermann and Westerheide (2012).

4. Data Analysis

We estimate the model seperately for men and women and include interaction for the parametric effects between age and educational groups. Positive values of the estimated effects indicate a positive effect and hence go along with a higher risk of getting unemployed. The estimated intercept $\hat{\beta}_0$ differs slightly between the model for men ($\hat{\beta}_0 = -4.966$) compared to the model for women ($\hat{\beta}_0 = -5.420$), i.e. the risk of getting unemployed in the reference category (30-39 years old individuals with vocational training) is slightly lower for women compared to men. The estimated parametric effects $\hat{\beta}_r$ including the interactions are visualised in Figure 1 and listed in number in Table 2. The effects show a similar tendency for both genders. However, the effects for women compared to those for men vary less strongly, i.e. the different parametric effects for women do not influence the risk of getting unemployed as much as do the parametric effects for men. In Figure 2 and 3 we show the resulting fit of the smooth effects in equation (3) which will be discussed later. The model has been evaluated using an approximative Akaike Information Criterion (AIC) (see Wood, 2006, p. 230) and dropping any effect from the final model increased the AIC value. All effects are now discussed and interpreted in detail.

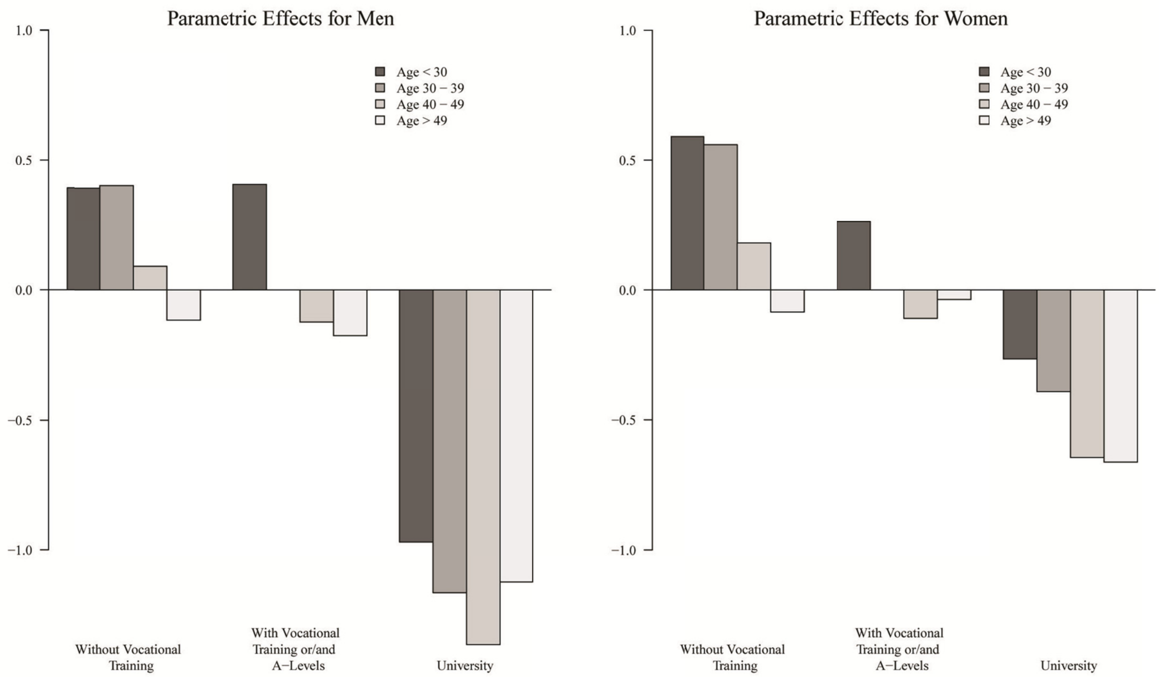

Figure 1. Estimated parametric effects $\hat{\beta}_r$ (including interaction) for getting unemployed for the IAB 'Regional File 1975-2004' for men and women, $r = 1, \dots, p$

Table 2. Estimated non-dynamic parametric effects $\hat{\beta}_r$ and interactions as well as standard errors of the IAB 'Regional File 1975-2004' for men and women

IAB Employment Samples Regional File 1975 – 2004	Men $\hat{\beta}_r$ (Std. Error)	Women $\hat{\beta}_r$ (Std. Error)
Intercept	-4.966 (0.006)	-5.420 (0.008)
Educational Group 1	0.402 (0.013)	0.559 (0.018)
Educational Group 2	reference	reference
Educational Group 3	-1.166 (0.022)	-0.392 (0.024)
Age Group 1	0.407 (0.008)	0.263 (0.011)
Age Group 2	reference	reference
Age Group 3	-0.123 (0.009)	-0.109 (0.011)
Age Group 4	-0.176 (0.009)	-0.037 (0.012)
Interactions		
Educational Group 1: Age Group 1	-0.417 (0.016)	-0.231 (0.022)
Educational Group 1: Age Group 3	-0.188 (0.020)	-0.269 (0.025)
Educational Group 1: Age Group 4	-0.342 (0.021)	-0.607 (0.026)
Educational Group 3: Age Group 1	-0.210 (0.050)	-0.136 (0.045)
Educational Group 3: Age Group 3	-0.076 (0.034)	-0.144 (0.038)
Educational Group 3: Age Group 4	0.217 (0.033)	-0.234 (0.044)

4.1 Parametric Effects

Effects for Men

Looking at the educational effects, it becomes clear that men with a higher education such as a university degree have a lower risk of getting unemployed compared to individuals with a lower education. This applies to all age groups. Thus, men with university degrees have the lowest risk to lose their jobs compared to all the other educational and age groups. Looking only at men with vocational training and/or A-Levels, individuals from 40 years on have -compared to those between 30 and 39 years- a lower risk of getting unemployed. In contrast, individuals up to 30 years with vocational training and/or A-Levels lose their job faster than the older ones and overall have the highest risk to lose their job (0.407). Men without vocational training up to 49 years have a higher risk of getting unemployed than men above 49 years of the same educational level and those of the reference category (30-39 year old men with vocational training and/or A-Levels).

To summarise, with increasing education the risk of getting unemployed decreases in each age group. This does not hold for under 30-year-old individuals with vocational training and/or A-Levels. Men of this age group with vocational training and/or A-Levels have a slightly higher risk to lose their job compared to men without vocational training of the same age group. Moreover, the effect of age differs within the educational groups. In general, it can be said that a higher educational level reduces the risk of losing the job, while the age effects depend on the educational level.

Effects for Women

In general, women with a higher education, i.e. holding a university degree, have a lower risk of getting unemployed than women with a lower education in the same age group and overall. Comparing the different age groups with regard to the educational background, it arises a similar picture as for men. Better educated women have a lower risk of getting unemployed in each age group, apart from over 49-year-old women with vocational training and/or A-levels. Altogether, women with a university degree and aged over 49 years have the lowest risk of getting unemployed (-0.663). Looking only at women without vocational training, women over 49 years have the lowest risk of getting unemployed compared to the other age groups. Women younger than 30 years have the highest risk of getting unemployed in all different educational groups while women without vocational training in this age group have the highest risk of all (0.590). Within each educational group the age effects show the tendency that older individuals have a lower risk of getting unemployed apart from women with vocational training and/or A-Levels: women between 40 and 49 years have the lowest risk of getting unemployed in the considered group.

Discussion

In Wichert and Wilke (2012) similar results concerning education were found, but not concerning age. In both logit analyses -which use data from the original administrative data set and a corrected version- higher education decreases the risk of getting unemployed, but the results are more distinguished for the original data than the corrected version. Looking at the age effects, Wichert and Wilke (2012) found that individuals above 55 years have a higher risk of getting unemployed compared to those between 25 and 50 years. Our results rather confirm those of Reinberg and Hummel (2002, 2003, 2005) who analysed the unemployment rates in different qualification groups for Germany: higher educated employees have a distinct lower risk of getting unemployed than lower educated men and women in East and West Germany. This is also true for older employees with a higher education which have lower unemployment rates than younger less educated individuals, see Reinberg and Hummel (2003, 2005). In 2004, higher educated employees between 55 and 64 years had the lowest unemployment rates compared to the younger age groups, see Reinberg and Hummel (2005). At first glance this result seems to stand in contrast to the unemployment behaviour in different age groups as analysed, for instance, in Hunt (1995), Hujer and Schneider (1995), Westerheide and Kauermann (2012) or Kauermann and Westerheide (2012). These papers, however, analyse the duration of unemployment, i.e. the chances for finding reemployment. Here, however, the focus is on becoming unemployed which apparently shows a different pattern.

Looking again at educational effects, Steiner and Schmitz (2010) concluded that an investment in education reduces the risk of unemployment. Wilke (2004) found out that on the one hand education has a high impact on a lower risk of unemployment especially for men, on the other hand he found only very small variation for women. Regarding personal characteristics, Lurweg (2010) observed in her analysis that an increase in education lowers the chance of getting unemployed. The results of Lauer (2003) concerning the risk of getting unemployed differ somewhat. She found out that individuals without vocational training have the highest risk of getting unemployed while individuals with vocational qualifications of an intermediate level have the lowest risk. University graduates have a higher risk of getting unemployed than individuals with an intermediate qualification level. In addition, she found out that women have a higher risk of getting unemployed in all educational groups compared to men. Generally, it can be inferred that a better education reduces the risk of getting unemployed. This result matches with our analysis. Looking at educational effects in papers analysing the unemployment duration, similar results can be found, that is education generally improves the re-employment probability, see for instance Lauer (2003), Westerheide and Kauermann (2012) or Kauermann and Westerheide (2012). Overall, a higher education seems to have a positive impact on the individual's labour market conditions.

4.2 Effects of a Former Duration of Unemployment

Referring to the effect for men illustrated in the first row of the left panel in Figure 2, we find that the accumulated duration of unemployment during the last year before unemployment has a positive effect and with it a higher risk of getting unemployed for individuals which have been unemployed between 60 and 120 days. The effect shows a peak at around 90 days, i.e. three months. Men who were only unemployed for some weeks in the last year have a lower risk of getting unemployed as well as individuals with a duration of over 200 days. We now look at the first row of the right hand panel of Figure 2 which represents the effect for women. The effect of the duration of unemployment during the last year for women differs slightly from the effect for men and is equally weak. The effect of former unemployment is slightly less pronounced and shows no positive effect.

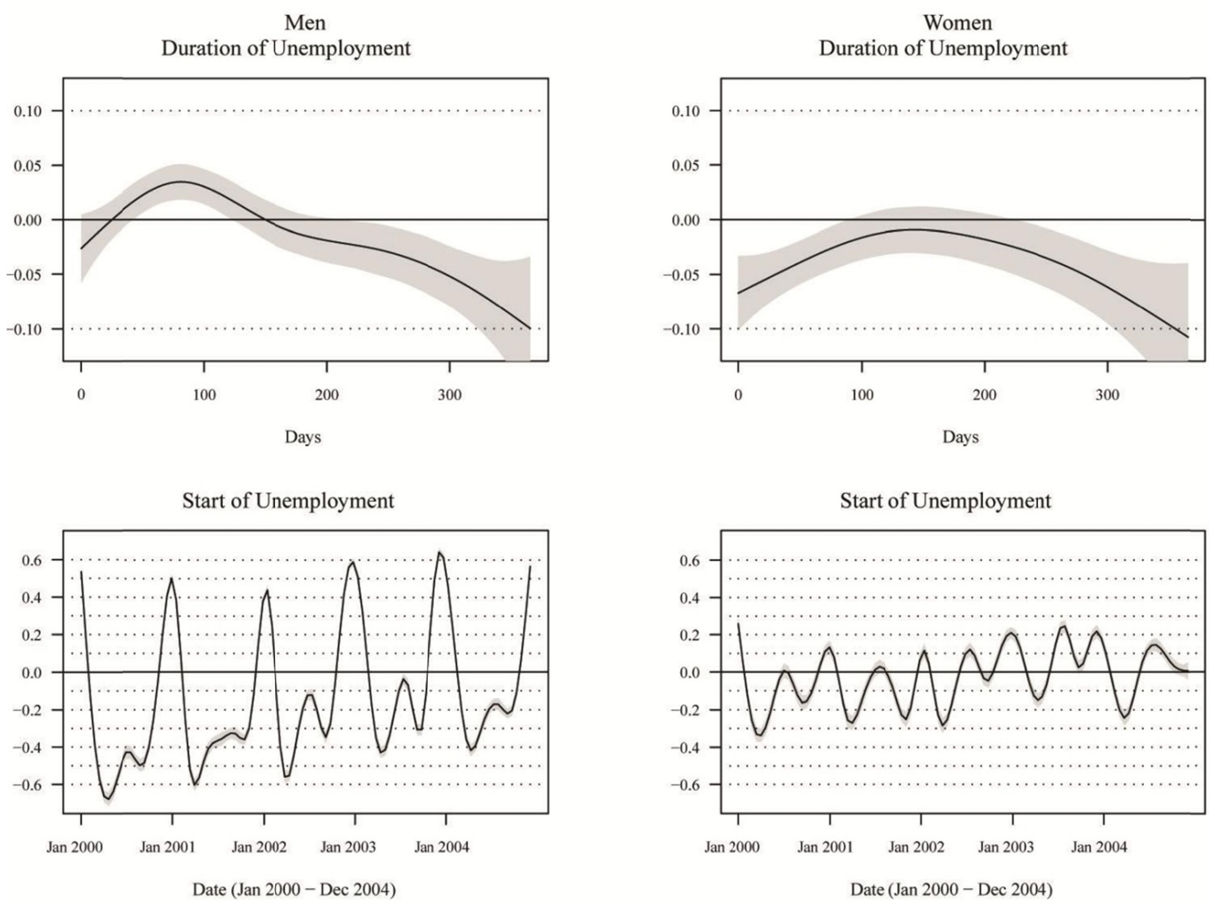

Figure 2. Fitted smooth effects with 95%-confidence intervals for the IAB 'Regional File 1975-2004' for men (left column) and women (right column)

Discussion

Wichert and Wilke (2012) figured out that individuals with a job tenure between 7 and 12 months have a slightly higher risk of getting unemployed compared to those with a tenure of less than 7 months. Their results may be linked to ours and may confirm the peak at around 3 months which could be detected for men. Individuals who have been unemployed between 1 and 4 months may belong to those who have a job tenure between 7 and 12 months. All in all, this covariate effect is small in size and does not strongly influence the risk of getting unemployed.

4.3 Calendrical Effects

Referring to the effect for men shown in the second row of the left panel in Figure 2, the calendar time has a distinct effect and a regular pattern can be observed. The effect always shows a high peak in December/January and a lower peak in June/July, i.e. the risk of getting unemployed is the highest in December/January and is still more pronounced in June/July than in the surrounding months. During spring men have the lowest risk of losing their jobs. The calendrical effect for women -displayed in the second row of the right hand panel of Figure 2- is distinct, but it does not show such a seasonal pattern as it could be seen for the calendrical effect for men. One can identify a peak during winter and summer over the observed period, but the peaks in wintertime are not so pronounced. Similarly, the trend of getting unemployed over the observed period is not so distinct for women.

Discussion

Especially, the effect of men behaves similar to observed unemployment data of other years in Germany, see Rudolph (1998) or Institut für Arbeitsmarkt- und Berufsforschung (2009) and go along with the results of Wichert and Wilke (2012). Beside these seasonal effects one can find a slightly decreasing trend between January 2000 and January 2002 and a slightly increasing trend of getting unemployed between January 2002 and December 2004. This trend goes along with the German business cycle, see Schirwitz (2009).

4.4 Regional Effects

Finally, the spatial effects for men and women shown in Figure 3 reveal outstanding lower risks of getting unemployed in large cities like Berlin, Hamburg or Munich and the surrounding regions as well as in other metropolitan areas like the Ruhr region or the Frankfurt area. In weakly populated regions as for instance Mecklenburg-West Pomerania in the northeast of Germany or the northwest of Lower Saxony in the northwest of Germany there is a higher risk of getting unemployed.

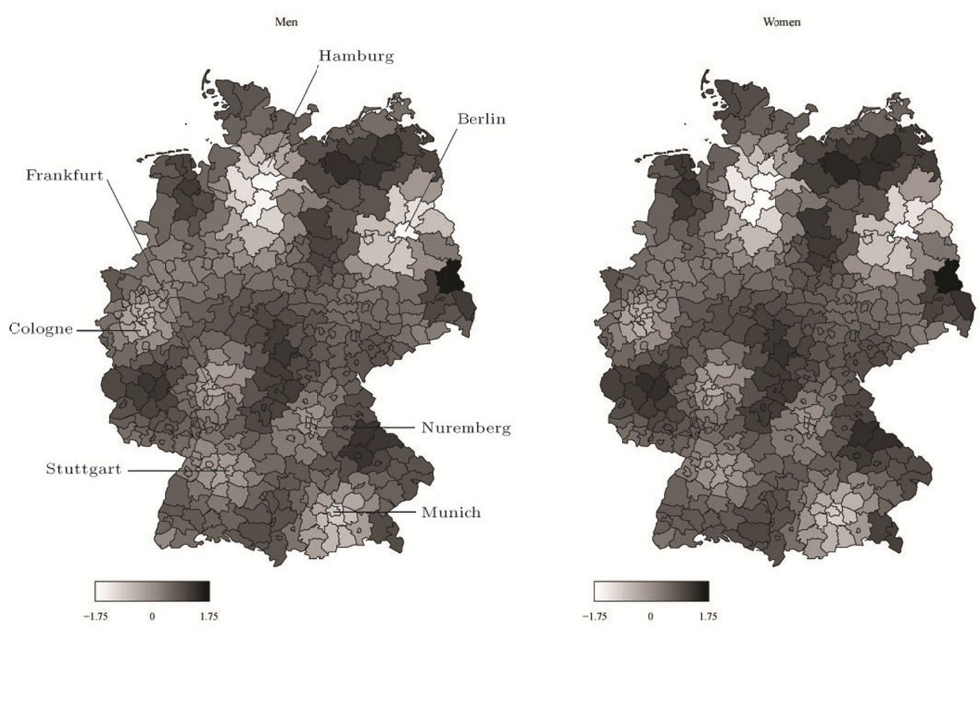

Figure 3. Fitted spatial effects for the IAB 'Regional File 1975-2004' for men (left) and women (right). Dark-coloured districts correspond to a positive effect and light-coloured districts correspond to a negative effect on the risk of getting unemployed

Discussion

Lurweg (2010) discovered a higher risk of unemployment for East German households compared to West German households. We do not find a specific east-west effect. Looking at analyses of the unemployment duration, spatial effects are clearly found. Kauermann and Westerheide (2012), who explored the chance of getting reemployed in Germany using also the IAB 'Regional File 1975-2004', found out that this covariate has a significant influence on the individual's re-employment chances, but the effect shows a different pattern. Arntz and Wilke (2009) detected only small differences between the unemployment durations in West and East Germany. Analysing the unemployment duration in West Germany, Fahrmeir, Lang, Wolff, and Bender (2003) also found spatial heterogeneity, but the spatial pattern is not comparable to ours.

5. Conclusion

In this paper we analysed the risk of getting unemployed in Germany using an additive Poisson model. We studied fixed individual covariate effects of men and women of different age and educational groups as well as smooth flexible covariate effects of calendrical, economic or spatial information. Looking at the educational effects, we can conclude that differences in the level of education strongly influence the risk of getting unemployed within each age group. The higher the educational level, the lower the risk of unemployment. This rule is true for all but one educational effect of the different age groups respecting men and women. A higher education is one of the essentials to be successful on the labour market. This conclusion goes along with analyses of qualification-specific unemployment rates and unemployment durations as well as with some other studies concerning the risk of unemployment. The risk of getting unemployed is lower for older better educated individuals (men and women) than for lesser educated younger individuals and can be found in analyses of age- and qualification-specific unemployment rates, too. Looking at the smooth flexible effects, the calendrical effect has a high influence on the unemployment risk of both genders. For men a regular cyclical pattern can be seen with the highest risk in wintertime and the lowest risk during spring. This effect is associated with the seasonal unemployment rate. For women this pattern is not so regular, but similar weaker risks are visible. Regarding the smooth effect of the region of the former working place, there is a strong influence on the individual's risk of getting unemployed. The spatial pattern is different from the spatial pattern of analyses concerning unemployment duration. Hence, the region specific effect of finding a job is different from the region specific effect of losing a job. Following Reinberg and Hummel (2002, 2003, 2005), that lower unemployment rates indicate a lower unemployment risk, we get similar results for educational effects but not for regional effects, for instance, we can not detect a specific east-west effect that could be in connection with the different regional unemployment rates. Our analysis shows that it is not always sufficient to analyse pure unemployment rates or other macroeconomic measurements to gain information about the risk of getting unemployed. However, conclusions drawn from analyses of unemployment duration can also not be taken to make an impact on the individual's risk of unemployment. As it could be seen in our analysis, the usage of an additive Poisson model seems to be a good way to obtain more detailed information about the influence of covariate effects on the unemployment risk.

References

Arai, M., & R. Vilhelmsson. (2004, July). Unemployment-risk differentials between immigrant and native workers in Sweden. *Industrial Relations, 43*(3), 690-698. http://dx.doi.org/10.1111/j.0019-8676.2004.00355.x

Arntz, M., & R. Wilke. (2009). Unemployment duration in Germany: Individual and regional determinants of local job finding. *Regional Studies, 43*(1), 43-61. http://dx.doi.org/10.1080/00343400701654145

Arrow, J. (1996). Estimating the influence of health as a risk factor on unemployment: A survival analysis of employment durations for workers surveyed in the German Socio-Economic Panel (1984 – 1990). *Social Science & Medicine, 42*(12), 1651-1659. http://dx.doi.org/10.1016/0277-9536(95)00329-0

Blanchard, O. (2006). European unemployment: The evolution of facts and ideas. *Economic Policy, 21*(45), 5-59. http://dx.doi.org/10.1111/j.1468-0327.2006.00153.x

Böheim, R., & M. P. Taylor. (2000, July). Unemployment duration and exit states in Britain. CEPR Discussion Paper No. 2500, Centre for Economic Policy Research, London.

Bover, O., M. Arellano, & S. Bentolila. (2002). Unemployment duration, benefit duration and the business cycle. *Economic Journal, 112*(479), 223-265. http://dx.doi.org/10.1111/1468-0297.00034

Covizzi, I. (2008). Does union dissolution lead to unemployment? A longitudinal study of health and risk of unemployment for women and men undergoing separation. *European Sociological Review, 24*(3), 347-361.

http://dx.doi.org/10.1093/esr/jcn006

Drews, N. (2008). Das Regionalfile der IAB-Beschäftigtenstichprobe 1975-2004: Handbuch-Version 1.0.2. FDZ Datenreport. Documentation on Labour Market Data 2008/02(DE), Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

Eilers, P. H. C., & B. D. Marx. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science, 11*(2), 89-121. http://dx.doi.org/10.1214/ss/1038425655

European Commission. (2009). *Employment in Europe 2009*. Luxembourg: Office for Official Publications of the European Communities.

Eurostat. (2009). *Europe in Figures - Eurostat Yearbook 2009*. Luxembourg: Office for Official Publications of the European Communities.

Fahrmeir, L., S. Lang, J. Wolff, & S. Bender. (2003). Semiparametric bayesian time-space analysis of unemployment duration. *Allgemeines Statistisches Archiv, 87*, 281-307.

Fieldhouse, E. A. (1996). Putting unemployment in its place: Using the samples of anonymized records to explore the risk of unemployment in Great Britain in 1991. *Regional Studies, 30*(2), 119-133. http://dx.doi.org/10.1080/00343409612331349528

Galiani, S., & H. A. Hopenhayn. (2003, June). Duration and risk of unemployment in Argentina. *Journal of Development Economics, 71*(1), 199-212. http://dx.doi.org/10.1016/S0304-3878(02)00138-4

Hammer, T. (1997). History dependence in youth unemployment. *European Sociological Review, 13*(1), 17-33. http://dx.doi.org/10.1093/oxfordjournals.esr.a018204

Hastie, T. (1996). Pseudosplines. *J. Roy. Statist. Soc. Ser. B 58*, 379-396.

Hastie, T., & R. Tibshirani. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Hujer, R., & H. Schneider. (1995). Institutionelle und strukturelle Determinanten der Arbeitslosigkeit in Westdeutschland: Eine mikroökonometrische Analyse mit Paneldaten. In B. Gahlen, H. Hesse, and H. J. Ramser (Eds.), *Arbeitslosigkeit und Möglichkeiten ihrer Überwindung*, Volume 25 of *Wirtschaftswissenschaftliches Seminar Ottenbeuren*, pp. 53-76. Tübingen: J. C. B. Mohr.

Hunt, J. (1995). The effect of unemployment compensation on unemployment duration in Germany. *Journal of Labor Economics, 13*(1), 88-120. http://dx.doi.org/10.1086/298369

Institut für Arbeitsmarkt- und Berufsforschung (Ed.). (2009). *IAB-Jahresbericht 2008*. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit.

Kauermann, G., & J. Opsomer. (2011). Data-driven selection of the spline dimension in penalized spline regression. *Biometrika, 98*(1), 225-230. http://dx.doi.org/10.1093/biomet/asq081

Kauermann, G., & N. Westerheide. (2012). To move or not to move to find a new job - spatial duration time model with dynamic covariate effects. *Journal of Applied Statistics, 39*(5), 995-1009. http://dx.doi.org/10.1080/02664763.2011.634394

Kauermann, G., T. Krivobokova, & L. Fahrmeir. (2009). Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society, Series B, 71*(2), 487-503. http://dx.doi.org/10.1111/j.1467-9868.2008.00691.x

Lauer, C. (2003). Education and unemployment: A French-German comparison. ZEW Discussion Papers 03-34, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Layard, P. R. G., S. J. Nickell, & R. Jackman. (2009). *Unemployment*. Oxford: Oxford Univ. Press.

Ljungqvist, L., & T. J. Sargent. (1998). The European unemployment dilemma. *Journal of Political Economy, 106*(3), 514-550. http://dx.doi.org/10.1086/250020

Lurweg, M. (2010, May). Perceived job insecurity, unemployment risk and international trade - a micro-level analysis of employees in german service industries. SOEPpapers 300, DIW Berlin, The German Socio-Economic Panel (SOEP).

McCullagh, P., & J. A. Nelder. (1989). *Generalized Linear Models* (second ed.). New York: Chapman and Hall. http://dx.doi.org/10.1007/978-1-4899-3242-6

Narendranathan, W., & M. B. Stewart. (1993). Modelling the probability of leaving unemployment: Competing risks models with flexible baseline hazards. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 42*(1), 63-83.

OECD. (2009). *OECD Employment Outlook: Tackling the Jobs Crisis* (2009 Edition ed.). Paris: OECD Publications.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (c/r: P519-527). *Statistical Science, 1*, 502-518. http://dx.doi.org/10.1214/ss/1177013525

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Reinberg, A., & M. Hummel. (2002). Arbeitslosigkeit: Qualifikation bestimmt Position auf dem Arbeitsmarkt. IAB Kurzbericht 15, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

Reinberg, A., & M. Hummel. (2003). Geringqualifizierte: In der Krise verdrängt, sogar im Boom vergessen. IAB Kurzbericht 19, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

Reinberg, A., & M. Hummel. (2005). Vertrauter Befund: Höhere Bildung schützt auch in der Krise vor Arbeitslosigkeit. IAB Kurzbericht 9, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

Røed, K., & T. Zhang. (2003). Does unemployment compensation affect unemployment duration? *Economic Journal, 113*, 190-206. http://dx.doi.org/10.1111/1468-0297.00086

Rudolph, H. (1998). Alle Jahre wieder: Saisoneffekte in der Arbeitslosigkeit. IAB Kurzbericht 12, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics, 11*, 735-757. http://dx.doi.org/10.1198/106186002853

Ruppert, D., M. Wand, & R. Carroll. (2003). *Semiparametric Regression*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511755453

Ruppert, D., M. Wand, & R. Carroll. (2009). Semiparametric regression during 2003-2007. *Electronic Journal of Statistics, 3*, 1193-1256. http://dx.doi.org/10.1214/09-EJS525

Schirwitz, B. (2009). A comprehensive German business cycle chronology. *Empirical Economics, 37*(2), 287-301. http://dx.doi.org/10.1007/s00181-008-0233-y

Steiner, V. (1997). Extended benefit entitlement periods and the duration of unemployment in West Germany. ZEW Discussion Papers 97-14, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Steiner, V. (2001). Unemployment persistence in the West German labour market: Negative duration dependence or sorting? *Oxford Bulletin of Economics and Statistics, 63*(1), 91-113. http://dx.doi.org/10.1111/1468-0084.00210

Steiner, V., & S. Schmitz. (2010). Hohe Bildungsrendieten durch Vermeidung von Arbeitslosigkeit. *Wochenbericht des DIW Berlin, 77*(5), 2-8.

Tatsiramos, K. (2009). Unemployment insurance in Europe: Unemployment duration and subsequent employment stability. *Journal of the European Economic Association, 7*(6), 1225-1260. http://dx.doi.org/10.1162/JEEA.2009.7.6.1225

Thapa, P. J. (2004, June). On the risk of unemployment: A comparative assessment of the labour market success of migrants in Australia. *Australian Journal of Labour Economics, 7*(2), 199-229.

Wahba, G. (1990). *Spline models for observational data*. CBMS NSF regional conference series in applied mathematics; 59. Philadelphia: Society for Industrial and Applied Mathematics. http://dx.doi.org/10.1137/1.9781611970128

Wand, M., & J. Ormerod. (2008). On semiparametric regression with O'Sullivan penalised splines. *Australian and New Zealand Journal of Statistics, 50*, 179-198. http://dx.doi.org/10.1111/j.1467-842X.2008.00507.x

Westerheide, N., & G. Kauermann. (2012). Flexible modelling of duration of unemployment using functional hazard models and penelized splines: A case study comparing Germany and the UK. *Studies in Nonlinear Dynamics and Econometrics, 16*(1). Article 5. http://dx.doi.org/10.1515/1558-3708.1914

Wichert, L., & R. A. Wilke. (2012). Which factors safeguard employment?: An analysis with misclassified German register data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 175*(1), 135-151. http://dx.doi.org/10.1111/j.1467-985X.2011.00698.x

Wilke, R. A. (2004). New estimates of the duration and risk of unemployment for West-Germany. ZEW Discussion Papers 04-26, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research, Mannheim.

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B), 65*(1), 95-114. http://dx.doi.org/10.1111/1467-9868.00374

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.

Wood, S. N. (2010, April). Package 'mgcv'. Online. Version 1.6-2. Retrieved June 8, 2010, from http://cran.at.r-project.org/web/packages/mgcv/mgcv.pdf

Appendix A. Estimation

First, we describe the fitting of the smooth, functional components in (3). The unknown functions are represented by a linear combination of thin plate spline basis terms, see Wahba (1990, pp. 30-34) with the popular cubic regression spline basis resulting as special case, see Wood (2006). This approach is used for all smooth functions except for the spatial effect $\phi(.)$. For the smooth function $\phi(.)$ we use thin plate regression splines, see Wood (2006). We now replace the functional components in (3) by

$$\delta(t) = B_\delta(t)b_\delta, \qquad \phi(s) = B_\phi(s)b_\phi$$
$$\xi(u) = B_\xi(u)b_\xi \tag{4}$$

with $B(.)$ as spline bases. We follow Hastie (1996) and Wood (2003) and use so-called 'low rank smoothing', i.e. each function works with a reduced set of knots. This set of knots is still large enough to capture the functional shape but small enough to guarantee feasible computation. This concept has been characterized by Eilers and Marx (1996) as P(enalized)-spline smoothing, see also Ruppert, Wand, and Carroll (2003, 2009). The number of knots is denoted with q. Following Wood (2006, p. 161), we set $q = 30$ for the calendar effect and $q = 60$ for the spatial effect functions, respectively. For the remaining smooth function we set $q = 10$. The model was also fitted for larger values of q but the choice of q has only small influence on the fit, see also Ruppert (2002) or Kauermann and Opsomer (2011). Suppose that $(\boldsymbol{x}_{ti}, t, s_{ti}, u_{ti})$ denote the observations for the *i*-th individual in interval t, where $i \in I_t$, that is individual i becomes unemployed in period t. Assuming that the individuals are independent the log-likelihood in (2) for parameter vector $\boldsymbol{\theta} = (\beta_0^T, \boldsymbol{\beta}_x^T, b_\delta^T, b_\xi^T, b_\phi^T)^T$ with $\boldsymbol{\beta}_x^T = (\beta_r^T, r = 1, \dots, p)$ can be expressed for all t as $l(\boldsymbol{\theta}) = \Sigma_{i \in I_t}\Sigma_{t=1}^T l_{ti}(\boldsymbol{\theta})$ where

$$\begin{aligned} l_{ti}(\boldsymbol{\theta}) = Y_{ti}[&\boldsymbol{x}_{ti}^T\boldsymbol{\beta} + B_\delta(t)b_\delta + z_{ti}B_\xi(u_{ti})b_\xi + B_\phi(s_{ti})b_\phi \\ &+\log(E_{ti}/\bar{E}) - \exp\{\boldsymbol{x}_{ti}^T\boldsymbol{\beta} + B_\delta(t)b_\delta \\ &+z_{ti}B_\xi(u_{ti})b_\xi + B_\phi(s_{ti})b_\phi + \log(E_{ti}/\bar{E}) + o_{ti}\}] \end{aligned} \tag{5}$$

Next we establish a penalty on the spline coefficients to obtain a smooth functional fit. The model is high dimensional which implies that the Maximum-Likelihood estimate will produce wiggled fitted curves. Hence, we use a penalty on the coefficients as described in Eilers and Marx (1996) and Ruppert, Wand, and Carroll (2003). Following Wand and Ormerod (2008), we rewrite the spline representation in (4) by extracting the intercept and the linear slope, i.e.

$$\gamma(t) = B(t)b_\delta = \beta_{\delta 0} + r\beta_{\delta 1} + \tilde{B}_\delta(t)\tilde{b}_\delta \tag{6}$$

where $\tilde{B}_\delta(t)$ is the reduced rank basis with intercept and linear slope extracted. For $\phi(s)$ and $\xi(u)$ we receive the reduced basis matrices $\tilde{B}_\phi(s)$ and $\tilde{B}_\xi(u)$. In the following a quadratic penalty on the spline coefficient is imposed, e.g. $\lambda_\delta\tilde{b}_\delta^T\tilde{D}_\delta\tilde{b}_\delta$. It can be demonstrated that it is equivalent to penalize with squared second order derivatives of the function (see O'Sullivan, 1986 or Wahba, 1990), or second (or higher) order differences of the spline coefficient b_δ (see Eilers and Marx, 1996). Here we make use of derivatives to penalize because this approach is implemented in the software we use for fitting the data (see end of this section). The parameter λ_δ is thereby a smoothing parameter which leads to a linear fit with $\lambda_\delta \to \infty$. This yields to the penalized log-likelihood

$$l(\boldsymbol{\beta}, \tilde{\boldsymbol{b}}, \boldsymbol{\lambda}) = \sum_{i \in I_t} \sum_{t=1}^{T} \tilde{l}_{ti} (\boldsymbol{\beta}, \tilde{\boldsymbol{b}}) - \frac{1}{2} \lambda_\delta \tilde{b}_\delta^T \tilde{D}_\delta \tilde{b}_\delta$$

$$- \tfrac{1}{2} \lambda_\xi \tilde{b}_\xi^T \tilde{D}_\xi \tilde{b}_\xi - \tfrac{1}{2} \lambda_\phi \tilde{b}_\phi^T \tilde{D}_\phi \tilde{b}_\phi \qquad (7)$$

with $\tilde{l}_{ti}$ as log-likelihood for the Poisson distributed variables and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p; (\beta_{0\delta}, \beta_{1\delta}), (\beta_{0\xi}, \beta_{1\xi}), (\beta_{0\phi}, \beta_{1\phi}))^T$, analogous definition for $\tilde{\boldsymbol{b}}$ and obvious definition for $\boldsymbol{\lambda} = (\lambda_\delta, \lambda_\xi, \lambda_\phi)^T$. The penalized log-likelihood can be fitted with standard software for generalized additive models, see Hastie and Tibshirani (1990). The only additional step which has to be done before modelling the data is to group the data to calculate the offsets. This can be easily done with simple data management as described above. For fitting our data we use the bam() procedure in **R** of the package mgcv. This procedure extends the gam() procedure and is helpful when working with large data sets, see Wood (2010). The smoothing parameters $\boldsymbol{\lambda}$ can be selected using a generalized cross validation which is embedded in the bam() procedure. We made use of REML estimation which is also implemented in this procedure. In the end, the inference for the model can be drawn. We follow thereby standard asymptotic arguments as presented in Ruppert, Wand, and Carroll (2003), Wood (2006) or Kauermann, Krivobokova, and Fahrmeir (2009). Denoting with $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \tilde{\boldsymbol{b}}^T)^T$ the complete parameter vector, the Fisher matrix can be determined with $F(\boldsymbol{\theta}, \boldsymbol{\lambda})$ and it can be demonstrated that

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}) = F^{-1}(\boldsymbol{\theta}, \boldsymbol{\lambda}) F(\boldsymbol{\theta}, \boldsymbol{\lambda} = \mathbf{0})\, F^{-1}(\boldsymbol{\theta}, \boldsymbol{\lambda}),$$

see for further information e.g. Ruppert, Wand, and Carroll (2003).