Deutsches Institut für Wirtschaftsforschung

DIW BERLIN

www.diw.de

SOEP The German Socio-Economic Panel Study

SOEPpapers

on Multidisciplinary Panel Data Research

213

Torben Kuhlenkasper • Göran Kauermann

Duration of maternity leave in Germany: A case study of nonparametric hazard models and penalized splines

Berlin, July 2009

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

Duration of maternity leave in Germany: A case study of nonparametric hazard models and penalized splines

Torben Kuhlenkasper

Göran Kauermann

Abstract

The paper investigates maternity leave behavior in West Germany for females being employed between 1995 and 2006 using data from the German Socio Economic Panel. The observational study focuses on the investigation of individual and family-related covariate effects on the duration of maternity leave following first or second childbirth, respectively. Dynamic duration time models are used in which covariate effects are allowed to vary smoothly with duration of being in maternity leave. The intention of the paper is to demonstrate with state of the art models how effects of covariables change over time and to analyse substantial differences between maternity leaves following first and second childbirth. Particularly the personal income of mothers and the educational attainment influence the decision when to return into employment. The leave period following second birth is influenced by the mothers' attachment to the labour market between their two maternity leave periods. As fitting routine penalized spline smoothing effects is employed using available software in R (www.r-project.org).

JEL classification: C14, C23, C41, J13, J24, J60
Keywords: Duration Time Models, Dynamic Effects, Maternity Leave, Panel Data, Employment Transition

1 Introduction

In the second half of the 20th century the female labour force participation rates have risen constantly in all western European countries. Possible reasons are discussed by Fitzenberger et al. (2004) and Rubery et al. (1999). Besides this, the educational attainment has increased for both, males and females. This goes hand in hand with a longer duration of schooling and/or vocational training resulting in a shifted labour market entry to a higher age. Especially mothers of small children have expanded their labour supply disproportionally in Germany and other western countries. As discussed by Rubery et al. (1999) especially mothers of small children are dependent on social and intra-family norms. In this context Dingeldey (2000) considers Germany as a "conservative welfare state" with the consequence of disincentives for mothers to return to work after being in maternity leave, see also Kreyenfeld and Geisler (2006). With maternity leave here and in the following we understand the period where the mother is not working, that is paid or unpaid maternity leave in the classical sense but also parental leave and voluntarily non-employment due to child care. Besides a social and labour market dimension, the employment break and the reentry into employment after giving birth is traditionally a crucial point in mothers' biographies and has individual and family related aspects which motivates to look at data on both, individual and intra-family level.

This paper focuses on two questions: First, how do individual and intra-family effects like the educational attainment, the personal as well as the household income before entering maternity leave and the presence of a working spouse or life partner influence the timepoint of returning to work after bearing a child. As second aspect of our paper, we analyse the dynamic behavior of these effects and how they vary with the duration of voluntarily staying at home and possibly loose their importance on mothers' decisions when to reenter the labour market as the time off the job continues. The aspired analysis is carried out with data from the German Socio Economic Panel (GSOEP, www.diw.de/soep).

The analysis of female labour market participation has been pursued in numerous papers before. Fitzenberger et al. (2004) and Fitzenberger and Wunderlich (2004) analyse this issue from a more macroeconomic and aggregate focus, while other papers like Beblo et al. (2006), El Lahga and Moreau (2007), Hank and Kreyenfeld (2000) or Kreyenfeld et al. (2007) focus on single effects concerning motherhood. The withdrawal of mothers from the labour market and the transition from maternity leave to employment has been under investigation before in a number of countries. We refer exemplary to Shirley et al. (1998) for the UK and to Desai and Waite (1991) for the US, respectively. Several recent papers focus on the loss of human capital, especially for highly educated mothers while being out of the labour market, see for instance Baum (2002) or Gutierrez-Domenech (2005). After giving birth, mothers face a trade-off between the costs of institutional child care and a proposed continuing loss of their personal human capital while staying at home. The personal income a woman was able to earn at the labour market prior to childbirth can be considered as the labour market value and mirrors as the opportunity cost of staying at home for childcare. As a result, mothers with high income should be more likely to return to the labour market. This goes hand in hand with the standard model of labour supply, which predicts an increasing probability of working with the wage or realized income. In contrast to recent papers, we allow the effect of personal income and the effects of all other covariables to vary over the duration of maternity leave. That is to say we capture how the effects influence the probability to return to professional life and how such effects change in time while controlling for unobserved heterogeneity.

The use of panel data ensures a reliable analysis of individual effects on the probability of re-entering the labour market after bearing a child. In this paper the analysis is based on data form the GSOEP, see Wagner et al. (2007) and Haisken-DeNew and Frick (2005) for a detailed introduction. The GSOEP provides suitable data and allows to empirically explore the returning-to-work-decision on a microlevel. We analyse the duration of maternity leave of 689 and 517 mothers for first and second maternity leave, respectively.

The statistical model being used in this paper is built upon the classical Cox model, see Cox (1972), but we allow for non-proportional hazards in the style of varying coefficients as suggested in Hastie and Tibshirani (1993), see also Gray (1994) or Therneau and Grambsch (2000). For fitting we make use of penalized splines to estimate smooth dynamic covariate effects as proposed in Kauermann (2005). Our modeling exercise extends this work by allowing for unobserved heterogeneity. To do so we include an individual latent factor which is modelled as Gamma distributed following Klein (1992). The Gamma distribution is chosen to obtain a coherent estimation framework. Alternatives are e.g. discrete approximations in a mixture distribution style, see e.g. Bover et al. (2002) or Friedl and Kauermann (2001). See also Verbeke and Lesaffre (1997) for a discussion on the specification or misspecification of the distribution of individual random effects. The Gamma approach applied here is beneficial since estimation can be easily carried out with available software.

Applying the estimation routine to the data at hand we can graphically investigate the dynamics of the overall probability of returning into paid employment after maternity leave. Looking at the covariate effects we show that the effects of realized personal income as well as educational attainment of the mothers significantly change over the duration of maternity leave. This allows for an advanced interpretation compared to the classical but in our case misspecified coefficients of a proportional hazard-model. Overall the decision for returning into paid employment underlies different effects also depending on the mothers' attachment to the labour market expressed here by a possible reentering in the labour market between the first and second leave period.

The paper is organized as follows. In Section 2 we give details about the data and show some exploratory analysis based on Kaplan Meier curves. Section 3 introduces penalized spline smoothing and suggests some ideas of model selection. Section 4 gives the data analysis before we conclude in Section 5.

2 Panel Data and the Duration of Maternity Leave

Our analysis of two different periods of maternity leave is based on data from January 1995 to December 2006. As maternity leave we define the period off the job due to pregnancy in the last weeks prior to the birth and subsequently the time after childbirth staying at home. In this definition maternity leave is not restricted to a job-protection-period granted by law and also includes unpaid maternity leave due to child care, where the mother stays (voluntarily) at home and is not available for the labour-market. As event we concider the return into any kind of paid employment (including full and part time employment as well as self-employment). The maximum duration times of maternity leave observed in the data are 62 and 72 months after first and second childbirth respectively and correspond to the latest events observed. In addition to the maximum duration times an observation can be censored due to panel mortality, a further childbirth within maternity leave or a transition into a involuntarily non-employed status, i.e. being registered as unemployed and seeking for a job. According to the GSOEP-questionaires we distinguish therefore between being a housewife and officially being registered as unemployed, which corresponds to voluntarily and involuntarily being out of the official labour-market, respectively. The analysis is restricted to mothers who were employed (full or part time) before having their child, i.e. before entering maternity leave. The data consists of individual spells from West Germany starting 1995. As covariates we focus on information about income, education and other personal variables. The realized personal income of mothers prior to their first and second maternity-leave period is defined as the maximum amount of labour income the mother earned in the five years anteceding birth, measured in euros[1]. The personal income

[1]See Projectgroup SOEP - DIW (2007b) for detailed outline and information about imputations.

in our discussion can be considered as the personal labour market value, which the mother has been able to realize at the labour market in advance of bearing a child. Apparently, this also mirrors as the opportunity cost for the female for staying at home. For our analysis we categorize the personal income into three levels: less than 1250 euros monthly income, between 1250 and 2250 euros monthly income (taken as reference category) and more than 2250 euros monthly income before entering maternity leave. Mothers, who withdrew voluntarily from the labour market between their childbirths for more than five years are excluded from the analysis of the second maternity leave. The thresholds correspond roughly to the 25% and 75% quartiles based on the data. Additionally to the personal income we look at the household income of the household the mother lives in while being off the job. We define household income as the maximum value of the provided generated net household income in the five years anteceding birth, see Projectgroup SOEP - DIW (2007a) for details. The household income is categorized into three groups: less than 2100 euros monthly household-income, between 2100 and 3500 euros monthly household income (taken as reference category) and more than 3500 euros monthly household income. The thresholds again correspond to the first and third quartiles of both data sets.

The educational attainment of a mother is measured with the "International Standard Classification of Education (ISCED)" which is available in the version of 1997 for the GSOEP data and used for our analysis[2]. Our analysis is carried out on three different groups of ISCED-levels: a lower group consisting of levels 0 to 2, a medium group consisting of levels 3 and 4 (taken as

[2]See UNESCO Institute for Statistics (2006) for details.

reference category) and finally a higher group with levels 5 and 6. The age of the mother at time of her first or second childbirth is also categorized into three groups: younger than 26 years, between 26 and 32 years (taken as reference category) and older than 32 years with the thresholds corresponding to first and third quartiles. Besides a binary factor indicating whether the mother has a migration background, we constructed a variable focussing on the spouse or life partner of the mother at time of the birth. We differentiate between mothers having a partner that is working at time of childbirth and mothers who do not have a working partner (including mothers living without a spouse or life partner). Two additional covariate effects are constructed for mothers being in their second maternity leave: First, we add an effect indicating whether the first child is older or younger than 3 years of age at the timepoint of the second delivery. Secondly, we observe whether the mother has been available for the labour market since the first maternity leave, i.e. being employed or seeking for a job and registered unemployed.

Simple Kaplan-Meier estimators are shown in Figure 1. The structure of this plot and subsequent one is as follows: The first and third column display the effect for the first maternity leave, the second and fourth column give results for the second maternity leave period. The overall survivor curves (top left row) show a strong decrease and a jump like decrease at 4 and 36 months, which corresponds to the length of the mother-protection-period and the job-protection-period granted by law in Germany respectively. Overall, the decrease for the second child is weaker, i.e. more females remain in maternity leave, especially after 36 months of duration. The estimated probability of extending the maternity leave after 36 months is approximately 50% after first childbirth and about 60% after second bearing. Looking at the effect of education, higher educated women tend to return to work earlier than lower educated ones. Concerning income, the Kaplan-Meier curves of mothers in the high wage group are nearly always below the curves of individuals with lower personal income, concluding that a higher labour market income prior to maternity leave lets mothers return to work earlier. This conclusion can also to drawn from looking at the household income concerning at least for the second child. Interestingly enough the effect is not seen for the first maternity leave. While no clear difference can be found by looking at the effect of the migration background of the mother, the age at the time of giving birth seems to have a (small) effect after second childbirth indicating that older mothers return to the labour market earlier than younger mothers. Looking at mothers who do not have a working partner, the corresponding effect indicates a higher chance of returning to the labour market soon after childbirth. Mothers who worked between the end of their first and the beginning of their second leave period, or at least have been seeking for a job, reveal a strong attachment to the labour market with a higher probability for returning to a job after second delivery soon. Finally the presence of a first child younger than 3 years has an effect on the duration in maternity leave, since mothers tend to stay at home as the second leave-period continues and both children are at home. In the remaining of the paper we will model these data using non-proportional as well as proportional hazard effects.

3 Dynamic Hazard Model and Penalized Spline Smoothing

We denote with $h(t, x)$ the hazard function which mirrors the probability of returning to professional life after t months in maternity leave. The hazard depends on the covariates discussed in the previous section, notated here with x. The typical Cox type model takes the form

$$h(t, x) = \exp\{\beta_0(t)\} \exp\{x\beta_x\}, \tag{1}$$

where $h_0(t) = \exp\{\beta_0(t)\}$ is the baseline hazard and β_x give the covariate effects, see Cox (1972). The effects expressed in β_x are constant over time, so that model (1) implies proportional hazards. Looking at Figure 1 the proportionality assumption seems questionable since the Kaplan-Meier curves do not mirror proportionality. We therefore allow covariate effects to change with the duration of maternity leave. This interaction of effects is incorporated in the model in a functional form by setting

$$h(t, x) = \exp\{\beta_0(t)\} \exp\{x\beta_x(t)\}, \tag{2}$$

where $\beta_x(t)$ is a functional effect, which is assumed to change smoothly, that is not rapidly, with duration time t. Estimation is carried out using penalized splines. We follow thereby closely Kauermann and Khomski (2006), so that we refrain from providing details here, but sketch some ideas in the Appendix. The basic idea is to replace the smooth functions $\beta_0(t)$ and $\beta_x(t)$ by some high dimensional spline bases and to achieve smoothness a penalty component is added to the likelihood. It can be shown that the likelihood can be approximated by a Poisson type Mixed Model, which in fact allows to obtain the fit relatively easy. Some more details are sketched in the Appendix. A user-friendly implementation to fit the model is provided in R, see R Development Core Team (2008), with the R-package `TwoWaySurvival`, which can be downloaded from the CRAN server at `www.r-project.org`. The package is an enhancement of the routines provided with Ruppert et al. (2003) and allows to fit the model easily and relatively quickly. Moreover, using standard asymptotic arguments, one can derive variance formulae from the estimates, making use of asymptotic normality statements. This allows not only to fit functional shapes but also to provide confidence bands for the functional effects.

The model (2) is on a population basis and does not incorporate individual latent effects, that is unobserved heterogeneity among the females. We therefore extend (2) so that the i-th female has the individual hazard

$$h_i(t, x_i) = h(t, x_i)v_i \tag{3}$$

with v_i as unobserved latent effect with $E(v_i) = 1$ to maintain identifiability. We assume a Gamma distribution for v_i, hence we use the conjugate distribution to Poisson which allows for numerically simple estimation of our model. The variance of v_i is estimated from the data and takes value 1.788 and 1.866 in our data example for first and second childbirth, respectively. Details on the algorithm are provided in the Appendix, see also Klein (1992).

4 Data analysis

Maternity leave in Germany is mostly regulated in two federal laws. An important role plays the law on the protection of expectant and nursing mothers (*Mutterschutzgesetz MuSchG*), which originally was introduced in 1952. This law regulates the rights for pregnant women and mothers after delivery. It has been modified several time in the last decades, mostly concerning the type of work a pregnant female is allowed to do on the job. In combination with the federal law on child support (*Bundeserziehungsgeldgesetz BErzGG*), which is replaced by the federal law of parental leave and financial support (*Bundeselterngeld- und Elternzeitgesetz BEEG*) coming into effect January 2007, mothers and fathers have the right to leave their paid employment, partly with ongoing payment, see John and Stutzer (2002) and Gottschall and Bird (2003) for a discussion on the legal framework. In 1993 the job protection period was expanded to 36 months. Other minor changes concern the amount of parental leave benefits, see Buchner and Becker (2008) for details. It is reasonable to assume that changes in the federal regulation of maternity and parental leave change the individual behavior of mothers, but the period we consider (1995 - 2006) did not see drastically amendments to the law so that we can assume a time constant legal framework. The aim of our analysis is now to analyse the effects of individual and family-related covariates on the decision to go back to the job within the legal framework. In Figure 2 we present the resulting fit of model (3) including 95% (pointwise) confidence bands for first and second maternity leave in comparison. The distributions of the covariates are listed in Appendix B, while the estimated proportional effects from a Cox-model are added as dotted lines in figure

2. These estimates may be seen as benchmark and our smooth estimation clearly indicates that the proportional hazard assumption is void. The first and third column show the fitted effects for first maternity leave, the second and fourth column give the corresponding fitted effects for the second child related maternity leave. The first plots in the first and second column show the baseline effects $\beta_0(t)$. These effects represent German-born mothers with a personal monthly income between 1250 and 2250 euros with an average achieved ISCED-level, who gave birth in the age between 26 and 32 and live with a working partner in a household with a monthly (net) income between 2000 and 3500 euros. Additionally, the baseline effect for mothers analyzed for second childbirth represents mothers who have been available for the labour market between their two leave periods and have a first child older than 3 years. We see a steep increase in the probability of returning into paid employment until about 4 months. Note that a fixed mother-protection period starts 6 weeks in advance of the scheduled birth date and ends 8 weeks after childbirth. During that time mothers are not allowed to work officially and they can return to their previous job at the earliest after being 4 months in maternity leave. Both baselines (first and second column on child respectively) also show a strong peak at 36 months of duration time. This timepoint indicates the end of the law-regulated job-protection period. The effect of the educational attainment are different for low- and high educated mothers. The effect of females being in the low ISCED group shows a linear dynamic behavior postulating a negative effect on the chance to return into a job. On the other hand, highly educated mothers seem to be affected by their educational attainment in an interesting dynamic way: The effects

reveal peaks around 4 and 36 months while being in their first maternity leave. These peaks cannot be observed for highly educated mothers being in their second leave period. The effect of high ISCED level is less significant after bearing a second child.
Significant effects can be observed by looking at mothers with a high personal income realized at the labour market anteceding birth: while observing an almost constant positive effect after first childbirth, the effect is only strong positive in the first two years of second maternity leave. Afterwards the effect weakens and even becomes insignificant. In contrast, mothers with low personal income are not affected by their realized income significantly.
An almost constant positive effect can be observed by looking at the presence of a partner who is working at time of both, first and second birth. In contrast, the income of the entire household where the mothers lives in does not effect the decision when to reenter the labour market.
A weak positive and negative effect can be seen by looking at mothers with a migration background in the first year in maternity leave for first and second leave period respectively. Looking at the age of the mother a small constant positive effect can be seen for mothers being older than 32 years at time of their second childbirth. Mothers being in their second maternity leave while having a child younger than 3 years are affected by the age of their first child until about 20 months after second delivery with a negative and decreasing effect. Finally, the most dominant factor effecting the length of the second leave period is the labour market attachment of the mother since her first childbirth: Mothers who have not been available for the labour market in this time window are significantly negatively affected with a minor increase

of the effect as the duration time continues.

5 Conclusion

The fitted smooth baseline and covariate effects reveal a two-dimensional framework for reentering the labour market after maternity leave: First, the legal framework of maternity leave in Germany drives some mothers back to their job after the maternity-protection-period which ends at 4 months and after the job-protection-period-ends which lasts 36 months. This is mirrored by the baseline effects and even strengthened by the additive covariate effects of high educational attainment and high personal income realized at the labour market before giving birth. Secondly, the personal labour market income before entering maternity leave and the educational attainment of the mother are the most dominant factors on the decision when to return into a paid employment after first birth. Mothers with high personal income earned prior to the childbirth are more likely to return to a job after first childbirth. Assuming the income as opportunity costs of not working due to child care, high opportunity costs force mothers back to their jobs after giving birth to their first child. For the second child the effect of personal income is only strong in the first two years being in maternity leave and fading away thereafter, indicating decreasing importance of high income earned prior to childbirth and decreasing opportunity costs as duration time continues. The effect of high realized personal income show a dynamic behavior, attenuating the general conclusion from Kreyenfeld and Geisler (2006) for Germany, who propose a static strong positive effect of high income on the likelihood of returning into paid work. Low-paid mothers show weaker incentive to get reemployed after first childbirth. For these mothers, the opportunity costs of not-working are lower and the effects of low income are weaker than for mothers expecting high salary after maternity leave. It is not surprising that highly educated mothers are more likely to return to a job shortly after their childbirth (personal income and educational level are highly correlated in the data). This underlies the general conclusion of Kreyenfeld and Geisler (2006) who state that highly qualified women establish a dual-earner-model in their families in contrast to the male-breadwinner-model. Looking at our fitted effects however, our analysis reveals a dynamic and vanishing effect of high education as maternity leave continues. A classical dual-earner-model, as discussed by Kreyenfeld et al. (2007) can only be obtained for families with mothers returning to work shortly after delivery. However, our analysis underlines Dingeldey (2000), considering Germany as a "conservative welfare state" with disincentives for mothers to work when the spouse or life partner contributes to the family income. Mothers without a working partner, including single mothers, strive back to the labour market sooner than mothers living with a partner being employed. Therefore, not only individual factors effect the length of the maternity leave. This underlines the proposed dependence on intra-family norms. In contrast however, the intra-family effect of household income only has a minor impact on the decision to reenter paid employment. The individual human capital of the mother, built up before entering motherhood and realized at the labour market in addition with the status of her spouse or life-partner is crucial for reentering the labour force in Germany after first childbirth. As Gustafsson et al. (1996) conclude, this holds for the UK as well, but not necessarily for all western countries. However, this does not hold as the family planing continues and the mothers enter their second maternity leave period. The attachment to the labour market, expressed by a readoptment or seeking of work between the two leave periods is more crucial to the decision when to reenter the labour market after second childbirth. The effects of high income and high ISCED level fade away and are only crucial for reentering the labour market after second childbirth in the first months after giving birth. A working partner however allows for an elongation of maternity leave during the entire leave-period after both, first and second childbirth, independently of income and educational attainment. A complete withdrawal from the labour market between the leave periods keeps mothers off the job as the child care of the second child continues.

We are reluctand to explain the different performances solely by different personal income, educational achievements and a readoptment of labour between two employment-breaks even though it seems plausible that this contributes to it. It is worth noting that the discussed socio-economic effects loose their impact as the mother continues to withdraw from the labour market voluntarily. The analysis therefore ends with the explanatory message based on our data analysis but does not go deeper into political and economical explanation. An investigation of major changes due to changed role allocation within German families and changed federal laws as well as a comparison of mothers living in West and East Germany is left for further research. The analysis however demonstrates the flexibility and capacity of penalized spline smoothing as estimation routine for functional data. This ensures the detection of time changing effects that even turn from positive to negative and vice versa during the analyzed periods of maternity leave. Especially the most crucial effects that influence mothers' decision to reenter the labour market (personal income, educational achievement, working between the leave periods) show a dynamical behavior. This cannot be observed specifying a classical proportional hazard model. A Cox-model averts a detailled analysis of the behavior of females facing an employment break due to childcare. Given that the software is available and the analysis did not require additional implementation, it seems inviting to make use of the non-proportional hazard model in other settings as well.

A Technical Details

For simplicity of notation and presentation of the penalized spline idea we ignore covariates x in model (2) for the moment. The underlying idea for estimation is to replace the unknown smooth function by some high dimensional parametric function. This means, for instance, we model $\beta_0(t)$ as $B_0(t)u_0$ with $B_0(\cdot)$ as high dimensional basis. For fitting we impose a penalty on coefficient vector u_0 which guarantees that the resulting fitted curve $\hat{\beta}_0(t) = B_0(t)\hat{u}_0$ is smooth. This is achieved by adding the penalty component $\lambda_{0u}u_0^T D_{0u}u_0$ to the log likelihood, with D_{0u} as penalty matrix and λ_0 as penalty parameter steering the amount of smoothness.

Denote now with (t_i, δ_i) the observations (again omitting covariates for simplicity of presentation), where t_i is the length of maternity leave and δ_i the censoring indicator. The penalized likelihood for coefficients u_0 results now with classical theory, see Cox and Oakes (1984), to

$$\ell(u_0, \lambda_{0u}) = \sum_{i=1}^{n} \left\{ \delta_i B_0(t_i) u_0 - \int_0^{t_i} \exp(B_0(z) u_0)\, dz \right\} - \lambda_{0u} u_0^T D_{0u} u_0. \qquad (4)$$

For estimation two further aspects have to be considered. First, one has to numerically solve the integral in (4) resulting from the integrated Hazard function. A simple and numerically feasible way to do so is to use a trapezoid approximation. In formulae this boils down to discretizing the continuous time scale. In our example we use trapezoids of width corresponding to one month which is also the finest resolution of the time scale. The second aspect is to select the smoothing parameter λ_{0u} appropriately, that is data driven. This can be done by comprehending the penalty as a priori normality imposed on the coefficient. In this case λ_0 becomes a parameter which can be estimated by maximizing the corresponding likelihood. In particular, trapezoid approximation and writing the penalty as a priori normal distribution lead to a generalized linear mixed model (GLMM) and the model can be easily fitted with available software.

To be more specific, let $0 = \tau_0 < \tau_1 < \ldots < \tau_K$ denote integration point at which we anchor our trapezoid approximation. In principle these can be the observed time points, even recommendable if duration times are observed on a discrete, rounded scale, like months. The integral component in (4) becomes now with trapezoid approximation and some simple calculus

$$\int_0^{t_i} \exp\left(B_0(z) u_0\right) dz \approx \sum_{k=0}^{K_i} \exp\left(B_0(\tau_k) u_0 + o_{ik}\right)$$

where $o_{i0} = \log\{\tilde{\tau}_{i0}\}$ and $o_{ik} = \log\{1/2[\tau_{i(k+1)} - \tilde{\tau}_{ik}]\}$ is a known offset with $\tilde{\tau}_{ik} = \min(\tau_k, t_i)$ and $K_i = \operatorname{argmax}\{t_i \le \tau_k\}$. Inserting this sum into (4) yields a penalized likelihood for artificial random variables d_{ik} taking values $d_{ik} = 0$ for $k < K_i$ and $d_{ik} = \delta_i$ for $k = K_i$ and having the Poisson distribution

$$d_{ik}|u_0 \sim Poisson\,(\lambda_{ik} = \exp\{B_0(\tau_k)u_0 + o_{ik}\}) \tag{5}$$

The next step is to formulate the penalty as normal prior leading to

$$u_0 \sim N(0, \lambda_{0u}^{-1} D_{0u}^{-}) \tag{6}$$

with D^{-} as (generalized) inverse. With (5) and (6) we obtain a Generalized Linear Mixed Model (GLMM) and the smoothing or penalty parameter becomes an a priori variance component which could be estimated following the likelihood principle. This idea has proved to be quite powerful, both in theory as well as in its numerical performance. For further details we refer to Wand (2003) and Kauermann (2005). The model can now be fitted using software for GLMMs in the style of Breslow and Clayton (1993). The idea is to treat spline coefficient u_{0t} as random so that the likelihood to be maximized results by integrating out the random terms. The latter is done by Laplace approximation. Clearly, the idea of penalized splines and its connection to GLMMs extends to model (2), that is for fitting the smooth covariates effect $\beta_x(t)$.

We now extend model (5) by assuming λ_{ik} to depend on some unobservable heterogeneity as well. We replace (5) by

$$d_{ik}|u_0, v_i \sim Poisson\,(\lambda_{ik} v_i) \tag{7}$$

with

$$v_i \sim Gamma\left(\frac{1}{\alpha}, \alpha\right) = \frac{1}{\left(\frac{1}{\alpha}\right)^{\alpha}\Gamma(\alpha)} v_i^{\alpha-1} \exp\left(-\alpha v_i\right). \tag{8}$$

Note that $E(v_i) = 1$ and $Var(v_i) = \frac{1}{\alpha}$. The Gamma distribution is the conjugate distribution for the Poisson distribution so that $v_i|(d_{ij}, j = 1, ..., K_i)$ is again Gamma distributed. This allows to easily apply an EM algorithm for estimation as shown in detail in Klein (1992).

B Distribution of covariates

			job status		job returns in year			
		all	no return	return	1st	2nd	3rd	4th or later
ISCED-level	low	121	75	46	12	8	12	14
	middle	382	220	162	52	42	43	25
	high	186	82	104	56	20	14	14
personal	< 1250	122	81	41	18	8	13	2
income	1250 - 2250	347	190	157	49	34	41	33
[EUR]	> 2250	220	106	114	53	28	15	18
working	yes	557	306	251	94	60	53	44
life-partner	no	132	71	61	26	10	16	9
household	< 2100	149	86	63	25	13	15	10
income	2100 - 3500	351	183	168	58	39	42	29
[EUR]	> 3500	189	108	81	37	18	12	14
migration	yes	104	56	48	23	8	7	10
background	no	585	321	264	97	62	62	43
age of mother	< 26	210	127	83	29	15	24	15
	26 - 32	370	192	178	67	44	39	28
	> 32	109	58	51	24	11	6	10
	Total	689	377	312	120	70	69	53

Table 1: Distribution of covariates for first maternity leave

		all	job status: no return	job status: return	job returns in year: 1st	2nd	3rd	4th or later
ISCED-level	low	90	59	31	10	5	6	10
	middle	298	144	154	50	26	31	47
	high	129	59	70	32	15	12	11
personal income [EUR]	< 1250	176	95	81	21	19	16	25
	1250 - 2250	234	117	117	37	17	24	39
	> 2250	107	50	57	34	10	9	4
working life-partner	yes	467	241	226	77	43	43	63
	no	50	21	29	15	3	6	5
household income [EUR]	< 2100	109	64	45	9	10	8	18
	2100 - 3500	284	142	142	50	22	30	40
	> 3500	124	56	68	33	14	11	10
migration background	yes	87	41	46	14	8	11	13
	no	430	221	209	78	38	38	55
age of mother	< 26	72	45	27	3	4	8	12
	26 - 32	304	157	147	48	25	25	49
	> 32	141	60	81	41	17	16	7
time since last birth [months]	< 36	226	123	103	31	9	26	37
	> 36	291	139	152	61	37	23	31
worked between the leave periods	yes	381	157	224	91	44	37	45
	no	136	105	31	1	2	5	23
	Total	517	262	255	92	46	49	68

Table 2: Distribution of covariates for second maternity leave

References

Baum, C. L. (2002). A dynamic analysis of the effect of child care costs on the work decision of low-income mothers with infants. *Demography 39*(1), 139–164.

Beblo, M., S. Bender, and E. Wolf (2006). *The Wage Effects of Entering Motherhood. A Within-firm Matching Approach.* Mannheim [u.a.]: ZEW Discussion Paper No. 06-053.

Bover, O., M. Arellano, and S. Bentolila (2002). Unemployment Duration, Benefit Duration and the Business Cycle. *The Economic Journal 112*(479), 223–265.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in Generalized Mixed Model. *Journal of the American Statistical Association. 88*, 9–25.

Buchner, H. and U. Becker (2008). *Mutterschutzgesetz und Bundeselterngeld- und Elternzeitgesetz : Kommentar* (8. ed.). München: Beck.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B 34*, 187–220.

Cox, D. R. and D. Oakes (1984). *Analysis of Survival Data.* London: Chapman & Hall.

Desai, S. and L. Waite (1991). Women's employment during pregnancy and after the first birth: occupational characteristics and work commitment. *American Sociological Review 56*(4), 551–566.

Dingeldey, I. (2000). Einkommenssteuersysteme und familiale erwerbsmuster im europäischen vergleich. In I. Dingeldey (Ed.), *Erwerbstätigkeit und Familie in Steuer- und Sozialversicherungssystemen. Begünstigungen und Belastungen verschiedener familialer Erwerbsmuster im Ländervergleich*, Chapter 1, pp. 11–47. Opladen: Leske + Budrich.

El Lahga, A. and N. Moreau (2007). *Would you Marry me? The Effects of Marriage on German Couples' Allocation of Time*. Mannheim [u.a.]: ZEW Discussion Paper No. 07-024.

Fitzenberger, B., R. Schnabel, and G. Wunderlich (2004). The gender gap in labor market participation and employment: A cohort analysis for west germany. *Journal of Population Economics 17*, 83–116.

Fitzenberger, B. and G. Wunderlich (2004). The changing life cycle pattern in female employment: A comparison of Germany and the UK. *Scottish Journal of Political Economy 51*(3), 302–328.

Friedl, H. and G. E. Kauermann (2001). Statistical modelling. In *Special Issue on Mixed Models, Euroworkshop on Statistical Modelling*, Volume 3. 233-234.

Gottschall, K. and K. Bird (2003). Family leave policies and labor market segregation in germany: Reinvention or reform of the male breadwinner model? *Review of Policy Research:* Special issue on Gender and Work Place Policies *20*(1), 115–134.

Gray, R. (1994). Spline-based tests in survival analysis. *Biometrics 50*, 640–652.

Gustafsson, S. S., C. M. Wetzels, J. D. Vlasblom, and S. Dex (1996). Women's labor force transitions in connection with childbirth: A panel data comparision between Germany, Sweden and Great Britain. *Journal of Population Economics 9*(3), 223–246.

Gutierrez-Domenech, M. (2005). Employment after motherhood: a European comparison. *Labour Economics 12*(1), 99–123.

Haisken-DeNew, J. P. and J. R. Frick (2005, December). *DTC – Desktop Companion the the German Socio-Economic Panel (SOEP)* (8. ed.). Essen, Berlin: RWI Essen & DIW Berlin.

Hank, K. and M. Kreyenfeld (2000). Does the availability of childcare influence the employment of mothers? – Findings from western Germany. MPIDR WORKING PAPER WP 2000-003, Max-Planck-Institut für demografische Forschung, Rostock.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B 55*, 757–796.

John, B. and E. Stutzer (2002). Professional activity during maternal leave. *Zeitschrift für Familienforschung 14*(3), 215–233.

Kauermann, G. (2005). Penalized spline smoothing in multivariable survival models with varying coefficients. *Computational Statistics & Data Analysis 49*, 169–186.

Kauermann, G. and P. Khomski (2006). Additive two way hazards model with varying coefficients. *Computational Statistics & Data Analysis 51(3)*, 1944–1956.

Klein, J. P. (1992). Semiparametric Estimation of Random Effects Using the Cox Model based on the EM Algorithm. *Biometrics 48*(3), 795–806.

Kreyenfeld, M. and E. Geisler (2006). Müttererwerbstätigkeit in ost- und westdeutschland: eine analyse mit den mikrozensen 1991-2002. *Zeitschrift für Familienforschung 18*(3), 333–360.

Kreyenfeld, M., D. Konietzka, and S. Böhm (2007). Educational differences in maternal employment in germany: an investigation based on the german microcensus data 1976-2004. *Zeitschrift für Soziologie 36*(6), 434–452.

Projectgroup SOEP - DIW (2007a). *Documentation HGEN: Household-related status variables and generated variables*. Berlin. `http://diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf`.

Projectgroup SOEP - DIW (2007b). *Documentation PGEN: Person-related status and generated variables*. Berlin. `http://diw.de/documents/dokumentenarchiv/17/60055/pgen.pdf`.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rubery, J., M. Smith, and C. Fagan (1999). *Women's Employment in Europe. Trends and Prospects*. London [u.a.]: Routledge.

Ruppert, D., M. Wand, and R. Carroll (2003). *Semiparametric regression.* Cambridge [a.o.]: Cambridge University Press.

Shirley, D., H. Joshi, S. Macran, and A. McCulloch (1998). Women's employment transitions around child bearing. *Oxford Bulletin of Economics and Statistics 60*(1), 79–98.

Therneau, T. M. and P. M. Grambsch (2000). *Modeling Survival Data – Extending the Cox Model*. New York: Springer.

UNESCO Institute for Statistics (2006). *International Standard Classification of Education ISCED 1997*. Montreal, Canada. ISBN 92-9189-035-9.

Verbeke, G. and E. Lesaffre (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis 23*(4), 541–556.

Wagner, G. G., J. Frick, and J. Schupp (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch (Journal of Applied Social Science Studies) 127*(1), 139–169.

Wand, M. (2003). Smoothing and mixed models. *Computational Statistics 18*, 223–249.

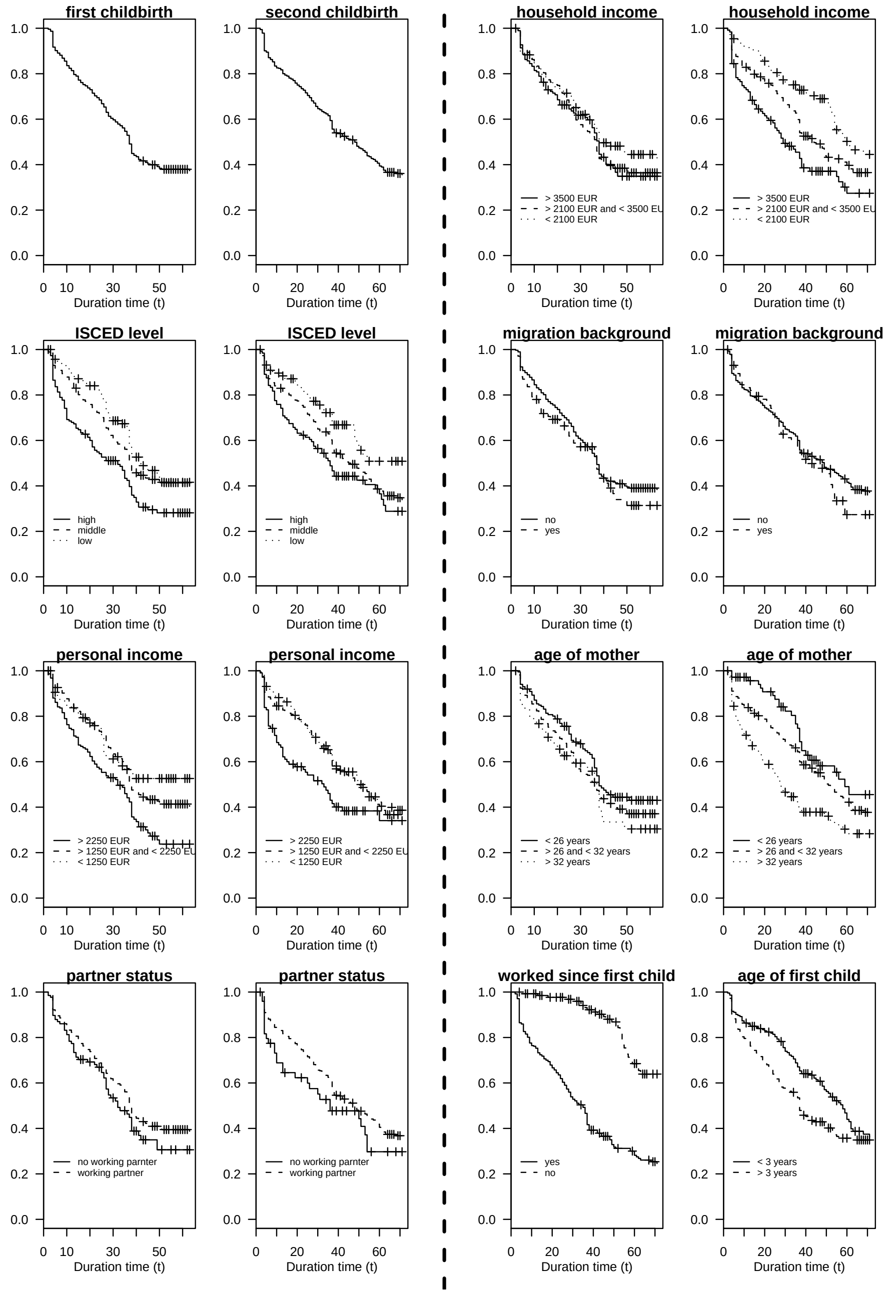

Figure 1: Kaplan-Meier curves for first and second maternity leave (left-hand and right-hand columns respectively.

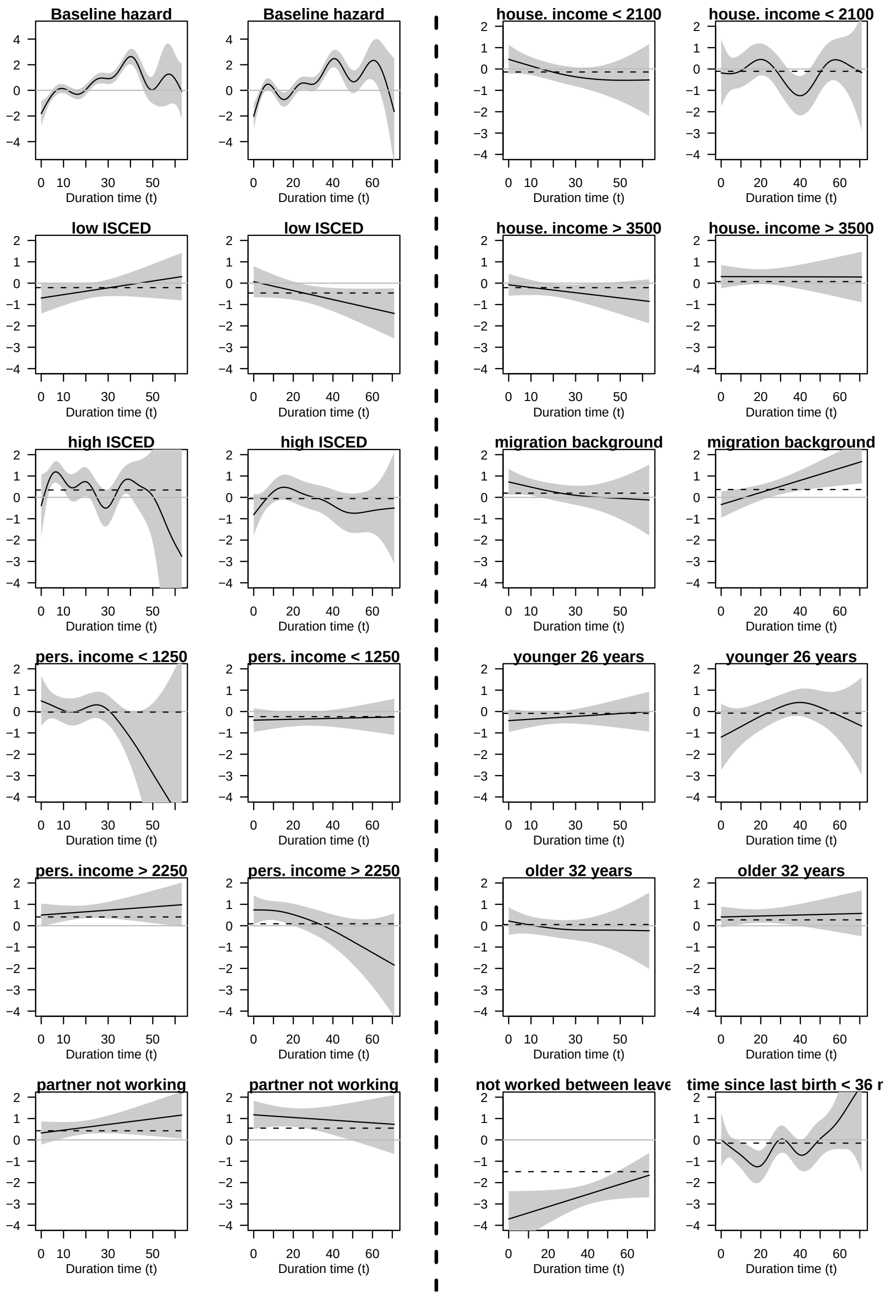

Figure 2: Fitted dynamic effects for duration time in maternity leave (in months) after first and second childbirth (first and second columns respectively.)