

Competing Risks Quantile Regression at Work

In-depth Exploration of the Role of Public Child Support for the Duration of Maternity Leave

Dlugosz, Stephan; Lo, Simon M. S. ; Wilke, Ralf

Document Version
Accepted author manuscript

Published in:
Journal of Applied Statistics

DOI:
[10.1080/02664763.2016.1164836](https://doi.org/10.1080/02664763.2016.1164836)

Publication date:
2017

License
Unspecified

Citation for published version (APA):
Dlugosz, S., Lo, S. M. S., & Wilke, R. (2017). Competing Risks Quantile Regression at Work: In-depth Exploration of the Role of Public Child Support for the Duration of Maternity Leave. *Journal of Applied Statistics*, 44(1), 109-122. <https://doi.org/10.1080/02664763.2016.1164836>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 26. Jan. 2022



Competing Risks Quantile Regression at Work: In-depth Exploration of the Role of Public Child Support for the Duration of Maternity Leave

Stephan Dlugosz, Simon M. S. Lo, and Ralf Wilke

Journal article (Post print version)

Cite: Competing Risks Quantile Regression at Work : In-depth Exploration of the Role of Public Child Support for the Duration of Maternity Leave. / Dlugosz, Stephan; Lo, Simon M. S. ; Wilke, Ralf. In: *Journal of Applied Statistics*, 11.04.2016.

This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Applied Statistics on 11 April 2016, available online:

<http://www.tandfonline.com/10.1080/02664763.2016.1164836>

Uploaded to [Research@CBS](#): June 2016



Competing risks quantile regression at work: In-depth exploration of the role of public child support for the duration of maternity leave. *

Stephan Dlugosz[†]

Simon M. S. Lo[‡]

Ralf A. Wilke[§]

March 2016

Abstract

Despite its emergence as a frequently used method for the empirical analysis of multivariate data, quantile regression is yet to become a mainstream tool for the analysis of duration data. We present a pioneering empirical study on the grounds of a competing risks quantile regression model. We use large scale maternity duration data with multiple competing risks derived from German linked social security records to analyse how public policies are related with the length of economic inactivity of young mothers after giving birth. Our results show that the model delivers detailed insights into the distribution of transitions out of maternity leave. It is found that cumulative incidences implied by the quantile regression model differ from those implied by a proportional hazards model. To foster the use of the model we make an R-package (`cmprskQR`) available.

Keywords: dependent competing risks, quantile regression, quantile crossings

*Financial support of the German Research Foundation (DFG) through the research grant "Statistical Modelling of Errors in Administrative Labor Market Data" is gratefully acknowledged. We thank Limin Peng for making basic sample code for her model available and two reviewers for their helpful comments.

[†]ZEW Mannheim, E-mail: stephan.dlugosz@googlemail.com

[‡]Lingnan University, E-mail: simonlo@ln.edu.hk

[§]Copenhagen Business School, Department of Economics, E-mail: rw.eco@cbs.dk

1 Introduction

Since the pioneering work by Koenker and Bassett (1978), quantile regression has steadily gained popularity in the statistical analysis of multivariate data. Originally developed as a linear model for cross section data, quantile regression has been extended to a wide range of models which are applicable to a wealth of data structures including time series and panel data. See Koenker (2005) for a detailed overview. In order to adapt quantile regression for duration data it was required to handle data with a censored dependent variable. Models with a randomly censored dependent variable have been developed since the 1980s starting with Powell (1986). For a comprehensive review of censored quantile regression see Koenker (2008). While quantile regression models for duration data have been proven to provide detailed insights into the conditional distribution of failure times (see for example Koenker and Geling, 2001), the censored quantile regression model limits research to single risks models or models with independent competing risks. For a review of the relevant literature see again Koenker (2008) or Fitzenberger and Wilke (2006). Peng and Fine (2009) were the first to suggest a competing risks quantile regression (CRQR) model for dependent competing risks. While they consider quantiles defined by the cumulative incidence curve (crude quantity), Hsieh, Ding, Wang and Chi (2013) propose a CRQR model for the marginal distribution (net quantity). To our knowledge these model have not yet entered the empirical literature and thus their practical properties are still largely unknown.

This paper makes the following contributions to the literature: It is the first comprehensive application of the dependent CRQR model of Peng and Fine (2009) using an important subject problem from economics and social sciences. We explore the merits and detect practical issues of the CRQR model when applied to large scale administrative individual data with multiple competing risks. In particular we study the role of public child care policies for the duration of maternity leave. We use linked administrative social security records from Germany to investigate how leave legislation and child care provision influence the decision of young mothers after the delivery of their first child not to return to employment. In particular we make an in-depth investigation how the policy variables are related with delivering a second child before returning to the labour market or making transitions into unemployment. We propose an inversion of the estimated quantiles defined by the cumulative incidences to obtain the cumulative incidences implied by the CRQR model. Cumulative incidences are functions in duration and normally considered in applications as quantile functions are harder to relate to the subject content. In order to ensure the inverse to exist we guarantee monotonicity of conditional quantiles by eliminating any quantile crossings using the method of Chernozhukov, Fernández-Val and Galichon (2010). The resulting cumulative incidences are directly comparable with

Table 1: Sample size and share of transitions into risks

Risk	Freq.	Percent
Next birth	4,653	23.82
Unemployment	2,697	13.80
Other	11,655	59.66
Right-censored	532	2.72
Total	19,537	100.00

cumulative incidences implied by conventional proportional hazards models. As an example we derive them from the popular semiparametric model by Fine and Gray (1999). Our results suggest that the quantile regression based cumulative incidences partly differ considerably from those implied by the proportional hazards model. Building upon Peng and Fine’s (2009) sample R-code we make an R-package (`cmprskQR`) available which contains a fast and ready to use implementation of the CRQR estimator and the implied cumulative incidences. Our work should therefore pave the way for quantile regression being applied to competing risks duration data and thus bringing high level statistical theory into practice.

2 Data

Our analysis uses biographical data of selected insurance agencies in Germany (BASiD). These data are linked administrative records from the German statutory pension insurance scheme (Deutsche Rentenversicherung) and the Federal Employment Agency (Bundesagentur für Arbeit). They comprise a 1% random sample of pension account holders which are around 579K individuals. The data contain daily spell information about employment periods, periods of training and education, periods of registered unemployment and unemployment compensation claim spells. Therefore they cover large parts of individual employment histories. The data also contain information about salaries, basic demographic variables such as age and gender, firm characteristics and regional identifiers among others. For more details about the BASiD see Hochfellner et al. (2012). For our analysis we use a very similar sample as in Arntz, Dlugosz and Wilke (2014) but only report the results for 2 of their 6 competing risks. It will be explained in Section 3 that restricting the analysis to two risks and treating all other observable exits as a pooled remainder state does not affect the results for the two risks of interest. Arntz et al. (2014) restrict their sample to females aged 18–45 who give birth to their first child in the period 1985–2005 and who were full-time employed at the time of conception. The sample comprises of 19,537 maternity durations.

Table 1 reports the number of observed transitions and the share of observed destination

states in the sample. It is apparent that around 24% deliver their second child and add a subsequent maternity duration, 14% enter unemployment, while almost 60% enter the pooled remainder state (among those around 45% enter employment). These risks are not assumed to be independent but there is independent censoring (end of data in 2009). Our model contains the same covariates as the model used in Arntz et al. (2014). The complete list of variables and descriptive statistics for the sample are given in Table 3 in the Appendix. There had been a number of policy changes in Germany during our observation period which altered the job protection period and maternity benefit entitlements (compare Figure 3 in the Appendix). In our analysis we focus on how child care policies and leave legislation are related with transition times into next birth and unemployment. We therefore only report the results for the following covariates: child care places per 100 children (aged < 3), job protection period (3 variables) and maximum entitlement for maternity benefits (in 1,000 Euros). For further details about the sample and detailed variable definitions see Arntz et al. (2014). Despite that the model controls for a wealth of variables, the analysis likely omits important variables. This includes the household background of the mother (household size and income and spousal job status), the proximity and age of the grandparents and performance indicators of the last employer (such as profits). While it is hard to believe that these variables are correlated with most of the institutional variables, the entitlement for maternity benefits is partly depending on household wealth and thus estimation results for the latter possibly incur a sizable omitted variable bias. We restrict the presentation of results to two risks and a small number of covariates because we believe this suffices to explore the practical properties of the CRQR model. As a flexible model produces a wealth of results, this paper would considerably gain in length if we included all of them. The interested reader can request additional CRQR results from the first author. The presentation of the full set of results obtained by a proportional hazards model with all six risks and all covariates can be found in Arntz et al. (2014). We focus our discussion of the results on whether quantile regression enables us to obtain new insights into the distribution of maternity leave than a conventional duration model.

3 Competing Risks Quantile Regression

We consider the following competing risks duration model to relate maternity leave duration to a number of variables. As described in the previous section we consider three exit states (next birth, unemployment, other). We therefore consider a model with $j = 1, \dots, 3$ competing random variables $T_j \in \mathbb{R}_+$. $\mathbf{X} \in \mathbb{R}^K$ is a $K \times 1$ vector of observable covariates. For consistency of the approach it is required that all covariates of the population model are observable. Although, we do not expect sizable inconsistencies in our estimation results due to the large

number of included variables, we restrict the discussion of the results to elaborating partial statistical relationships.

Due to the competing risks structure it is only possible to observe $(U, \epsilon, \mathbf{X})$ with $\epsilon = \operatorname{argmin}_j \{T_j\}$ and $U = \min_j \{T_j\}$. Independent censoring is characterised by censoring point $C \in R_+$. Thus observable duration is $T = \min\{U, C\}$ and let $\Delta = \mathbb{I}(U \leq C)\epsilon$ with indicator function \mathbb{I} . Let $(t_{ji}, c_i, \delta_i, \mathbf{x}_i)$ be $i = 1, \dots, N$ realisations of $(T_j, C, \Delta, \mathbf{X})$ and $F_j(t; \mathbf{x}) = \Pr(T_j \leq t, \Delta = j; \mathbf{x})$ be the cumulative incidence curve for risk j . $F_j(t; \mathbf{x})$ describes the probability of observing a transition into j and can be estimated without assumptions on the dependence structure between competing risks. It is remarked that this is not the same as estimating $\Pr(T_j \leq t; \mathbf{x})$, the distribution of the competing T_j . $\Pr(T_j \leq t; \mathbf{x})$ for all j describe the data generating process but their estimation requires assumptions on the dependence structure between competing risks. Following Peng and Fine (2009) we consider the τ 's conditional quantile function for the cumulative incidences, namely

$$\begin{aligned} Q_j(\tau; \mathbf{x}) &= \inf\{t : F_j(t; \mathbf{x}) \geq \tau\} \\ &= g(\mathbf{x}'\boldsymbol{\beta}_j(\tau)) \end{aligned} \quad (1)$$

for $0 < \tau_L < \tau < \tau_U < 1$ and g is a known monotonic link function. The parameters $\boldsymbol{\beta}_j(\tau)$ are unknown and can change in τ .

Given that we consider a model for the cumulative incidence, we can focus on a small number of observable risks and pool all other risks into a remainder state. Although, the pooled remainder risk does not have a direct interpretation, the pooling does not change the cumulative incidences for the risks of main interest which can be still estimated separately. We consider a nonparametric model for the distribution of durations to avoid misspecification. By denoting $T_j^* = g(\mathbf{x}'\boldsymbol{\beta}_j(\tau))$ we have $g^{-1}(T_j^*) = \mathbf{x}'\boldsymbol{\beta}_j(\tau)$ and thus the estimator for $\boldsymbol{\beta}_j(\tau)$ can be obtained by conventional quantile regression models. In particular, in the absence of censoring $\hat{\boldsymbol{\beta}}_j(\tau)$ is the solution to

$$\hat{\boldsymbol{\beta}}_j(\tau) = \min_{\mathbf{b}} \sum_{i=1}^N \rho_\tau(g^{-1}(T_{ji}^*) - \mathbf{x}_i'\mathbf{b})$$

with check function $\rho_\tau(v) = v\{\tau - \mathbb{I}(v \leq 0)\}$ and $T_{ji}^* = \mathbb{I}(\epsilon_i = j) * t_{ji} + \mathbb{I}(\epsilon_i \neq j) * \infty$. Equivalently the estimator can be obtained by solving the estimating equation with respect to \mathbf{b}

$$N^{-1/2} \sum_{i=1}^N \mathbf{x}_i [\mathbb{I}\{g^{-1}(t_i) \leq \mathbf{x}_i'\mathbf{b}, \epsilon_i = j\} - \tau] = 0.$$

In the presence of censoring Peng and Fine (2009) suggest a weighted version of the above estimating equation by using inverse probability of censoring weighting which is

$$S_N(\mathbf{b}, \tau) = N^{-1/2} \sum_{i=1}^N \mathbf{x}_i \left(\frac{\mathbb{I}\{t_i \leq g(\mathbf{x}_i'\mathbf{b})\} \mathbb{I}\{\epsilon_i = j\} - \tau}{\hat{G}(t_i)} \right)$$

with $\hat{G}(t)$ is the Kaplan-Meier estimator for $Pr(C \geq t)$. Due to not being continuous Peng and Fine (2009) suggest a related L_1 type function which can be more easily numerically minimized. Under some technical conditions they show that the resulting estimator is \sqrt{N} -consistent and asymptotically normal and they derive the variance matrix of the estimator.

Empirical researchers are often primarily interested in reporting results as a function of duration or failure time rather than conditional quantiles. Since the CRQR model estimates the latter, we suggest that $F_j(t; \mathbf{x})$ can be obtained as follows:

$$\begin{aligned}
Q_j^{-1}(t; \mathbf{x}) &= \sup\{\tau : \tau \leq F_j(t; \mathbf{x})\} \\
&= \sup\{\tau : Q_j(\tau; \mathbf{x}) \leq t\} \\
&= \sup\{\tau : g_j\{\xi_{\mathbf{x},j}(\tau)\} \leq t\} \\
&= \sup\{\tau : \xi_{\mathbf{x},j}(\tau) \leq g^{-1}(t)\} \\
&= \sup\{\tau : \tau \leq \xi_{\mathbf{x},j}^{-1}[g^{-1}(t)]\}
\end{aligned} \tag{2}$$

with $\xi_{\mathbf{x},j}(\tau) = \mathbf{x}'\boldsymbol{\beta}_j(\tau)$. The last step requires that $\xi_{\mathbf{x},j}(\tau)$ is invertible which is the case if $\xi_{\mathbf{x},j}(\tau)$ is monotonic in τ for all \mathbf{x} . This condition is violated if there are any crossings of conditional quantiles. Quantile crossings can occur due to misspecification of the quantile regression model but they are also present if the model is correctly specified. It is therefore an empirical exercise whether $\xi_{\mathbf{x},j}(\tau)$ is invertible when estimated $\boldsymbol{\beta}_j(\tau)$ are used. In order to ensure monotonicity of conditional quantiles- and therefore their invertibility- we employ the method of Chernozhukov et al. (2010). In particular let $\hat{Q}_j(\tau; \mathbf{x})$ be the conditional quantile implied by the model. Then the monotone rearranged quantile curve $\hat{Q}_j^*(\tau; \mathbf{x})$ is computed as the τ -th quantile of $\{\hat{Q}_j(\tau_1; \mathbf{x}), \dots, \hat{Q}_j(\tau_k; \mathbf{x})\}$, where $\{\tau_1, \dots, \tau_k\}$ is a sufficiently fine grid of equidistant values in $(0, 1)$.

In an application τ_L and τ_U are determined by the technical condition C4 of Peng and Fine and they differ across risks. While τ_L is likely a small number greater than zero, τ_U is typically in the range of the share of observed transitions into the relevant risk. The cumulative incidence may not attain higher values than this share and thus higher quantile functions are not observable and thus cannot be estimated. This means the quantile regression model cannot be estimated for τ too large because $F_j(t; \mathbf{x})$ attains a plateau level for t large enough and the estimated $\boldsymbol{\beta}_j(\tau)$ would tend to infinity. Condition C4 of Peng and Fine is met if $N^{-1} \sum_i \mathbf{x}_i \mathbf{x}_i'$ is non singular and if $\inf_{\mathbf{b} \in \mathcal{B}_j} \partial F_j(g(\mathbf{x}'\mathbf{b}); \mathbf{x}) / \partial t$ is bounded below by a positive constant, where \mathcal{B}_j is a set of \mathbf{b} which contains elements sufficiently close to the true $\boldsymbol{\beta}_j(\tau)$. These properties can be verified with the data. τ_U is then such that F_j is below its plateau level and therefore there are no implied $F_j(t; \mathbf{x})$ for t large. In our application we set $g(v) = \exp(v)$. The model is estimated on an equidistant 0.001 grid on τ_L and τ_U . τ_L and τ_U are automatically determined for each risk.

In our application we compare estimates of the CRQR model based F_j with implied F_j obtained by a proportional hazards model. For the latter we use the model by Fine and Grey (1999) who consider the cause-specific subdistribution hazard function

$$\begin{aligned}\lambda_j^s(t; \mathbf{x}) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T \leq t + \Delta t, \delta = j; T \geq t \cup (T \leq t \cap \delta \neq j), \mathbf{x}) \\ &= -\frac{\partial \ln[1 - F_j(t; \mathbf{x})]}{\partial t}.\end{aligned}$$

In their model $\lambda_j^s(t; \mathbf{x})$ is assumed to be a proportional hazards model of the form

$$\lambda_j^s(t; \mathbf{x}) = \lambda_{j0}^s(t) \exp(\mathbf{x}'\boldsymbol{\gamma}_j),$$

where $\lambda_{j0}^s(t)$ is the nonparametric baseline subdistribution hazard and $\boldsymbol{\gamma}_j$ are unknown parameters. The cumulative incidence in this model is

$$F_j(t; \mathbf{x}) = 1 - \exp[-\Lambda_{j0}^s(t) \exp(\mathbf{x}'\boldsymbol{\gamma}_j)], \quad (3)$$

where $\Lambda_{j0}^s(t; \mathbf{x}) = \int_0^t \lambda_{j0}^s(u) du$ is the cumulative baseline subdistribution hazard. The marginal effect of a continuous x_k on $F_j(t; \mathbf{x})$ is not simply γ_{jk} but

$$\frac{\partial F_j(t; \mathbf{x})}{\partial x_k} = (F_j(t; \mathbf{x}) - 1) \ln(1 - F_j(t; \mathbf{x})) \gamma_{jk}. \quad (4)$$

Thus the magnitude of the partial effect varies with t but its sign is determined by the sign of γ_{jk} . Therefore the direction of the effect is the same for all t . This implies that the conditional quantile functions for this model are also shifted only in one direction for all τ when a covariate changes. The Cox proportional hazards model has a similar property, that the sign of the partial covariate effect on the conditional quantile does not change across quantiles (Koenker and Geling, 2001). In contrast, the sign of $\boldsymbol{\beta}_j(\tau)$ in model (1) is not restricted across τ . A covariate can therefore shift downward the conditional quantile function for lower quantiles (shorter duration) and shift it upward for larger quantiles (longer duration). Therefore, the CRQR model offers more flexibility than the proportional hazards model in terms of the covariate effect. But, as pointed out by a referee, the proportional hazards model is characterised by a greater flexibility in the time effect due to the baseline hazard being nonparametric. Given that these two models are non-nested, we are unable to jump to the conclusion with certainty that the proportional hazards model is misspecified, when a change in the sign of the CRQR coefficients (when varying τ) is found. This, however, hints at evidence of conflicting results of these two models, and such evidence might point to a violation of the property of the proportional model in (4). In our application we compute and compare the estimated cumulative incidences implied by the different models as given in (2) and (3) to assess how the different sets of restrictions possibly lead to different results. We use the bootstrap in order to determine standard errors and

inference statistics for the estimated cumulative incidences. We make an R-package `cmprskQR` available which comprises of the estimator for $\beta_j(\tau)$ in (1) and the implied cumulative incidence in (2). For the estimation of the Fine and Gray (1999) model we use the `cmprsk` R-package by Bob Gray.

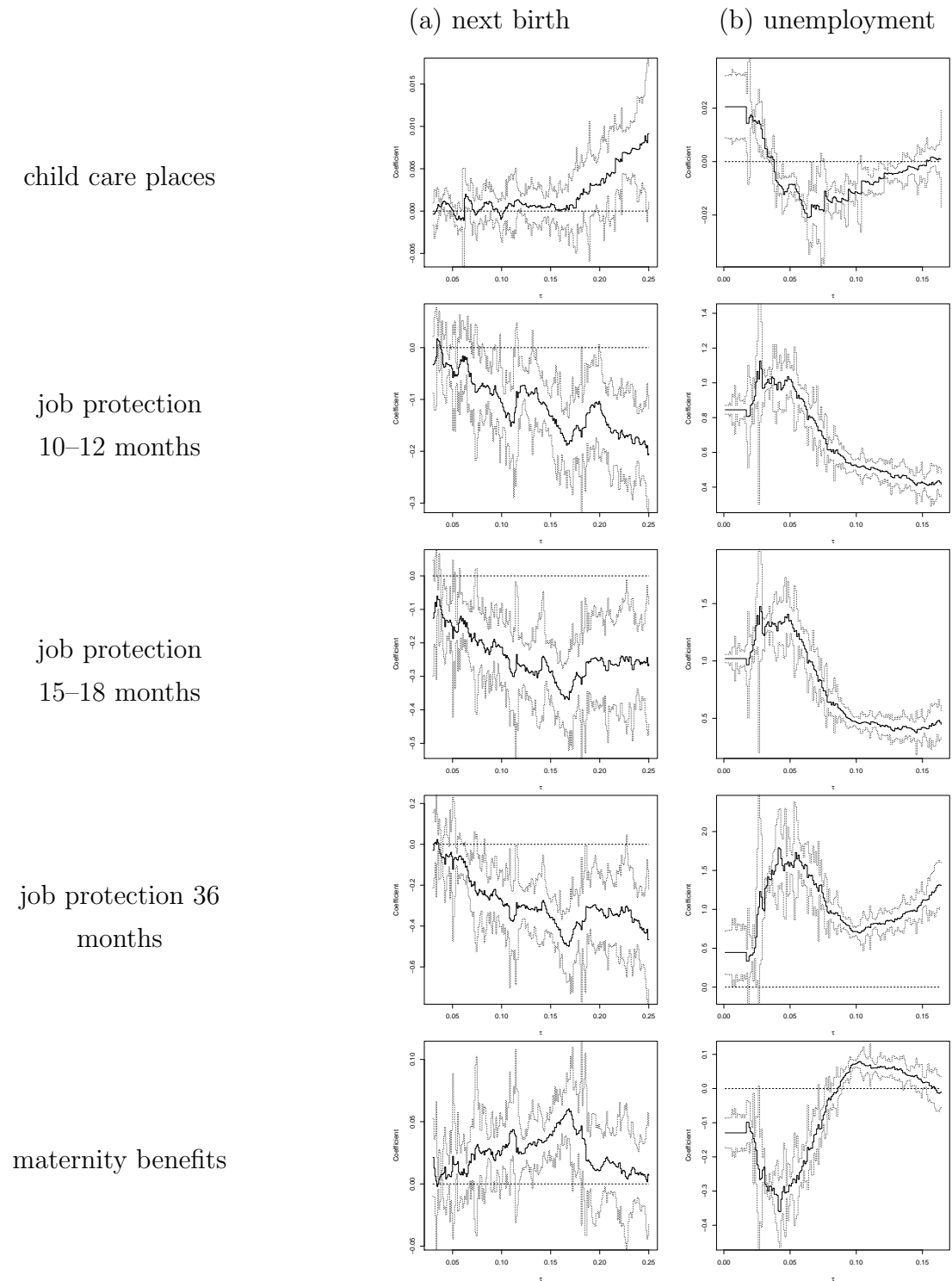
4 Empirical Results

We first present estimated quantile regression coefficients of Q_j as given in (1). Figure 1 shows selected coefficients for public policy related variables in the model. Column (a) contains the coefficients for the risk "next birth" and column (b) the coefficients for the risk "unemployment". τ_U is 0.28 for "next birth" and 0.164 for "unemployment", thus in both cases it is slightly above the share of observed transitions into these risks (compare Table 1). While τ_L is close to zero for both risks, we do not report the coefficients for risk "next birth" for $\tau < 0.03$. A natural "next birth" cannot occur within nine months which corresponds to a flat cumulative incidence during the first months of maternity leave. We do, however, observe a small number of transitions within 9 months because of adoptions of small babies. Thus, the cumulative incidence is not fully flat but we obtain large jumps in qr coefficients across the lowest quantiles. Thus, we decided against reporting them for "next birth" as this points to some instability. For comparison of the QRQR model with a conventional approach we report the coefficients obtained by the proportional hazards model for the cumulative incidence in Table 5 in the Appendix. We focus our discussion of the results on to what extent the CRQR model is capable of providing different insights into the role of covariates for the distribution of observed transitions than a conventional proportional hazards model.

It is apparent from Figure 1 that coefficients vary over risks, which highlights the importance of employing a competing risks model. Moreover some of the coefficients vary considerably across quantiles, which points to the usefulness of the CRQR model to detect these patterns. For example "child care places" is almost constant in τ for "next birth" but has a U-shape for "unemployment". For risk "unemployment" the results provide detailed insights into the partial relationships as all coefficients vary significantly across quantiles. In particular, the job protection variables are found to strongly shift the quantiles of the cumulative incidences. Moreover, the coefficients on "child care places" and "maternity benefits" change sign across quantiles. This means for instance that more generous maternity benefits are estimated to increase conditional quantiles for unemployment at shorter durations but decrease cumulative incidences at longer durations.

To contrast the results obtained by the CRQR and the proportional hazards model, Table 2 contains a summary of the direction and strength of the estimated effects. In particular

Figure 1: Selected estimated coefficients of the CRQR model with 95% asymptotic confidence intervals.



for the risk "unemployment" the proportional hazards model provides evidence for a negative relationship between "child care places" and the probability of observing a transition, while the CRQR model does not suggest this for most quantiles (except for the lowest, < 0.03). From a policy perspective the proportional hazards model therefore suggests that more child care provision is related with longer unemployment periods. The CRQR model for most quantiles suggests the contrary or does not reveal any systematic relationship, which appears to be more plausible given increased child care should improve the employability of young mothers and thus should result in a decrease in the probability of making a transition to unemployment. The CRQR model provides strong evidence for a negative relationship between the length of job protection periods and the probability of making a transition to unemployment for all quantiles, but the proportional hazards model only finds this for the longest job protection period (36 months). These contrasting results are surprising, indeed, in particular given that the CRQR model suggests that these are the most important variables in the model. From a policy perspective, the CRQR suggests that longer job protection periods of young mothers after giving birth are related with a considerably lower probability of entering unemployment, in particular at shorter durations. Given that the coefficients of the CRQR model strongly fall for higher quantiles, this points to that some mothers delay the entry into unemployment rather than it is entirely avoided. Thus, differences are smallest when cumulative incidences reach their plateau levels. Finally, as indicated in Figure 1, the CRQR model estimates that the partial relationship between maternity benefits and the probability of observing a transition to unemployment change its sign across different quantiles. The proportional hazards model in contrast estimates this relationship to be small and insignificant. Thus, the various effects of different signs estimated by the CRQR possibly level out in the estimation of the proportional hazards model. The CRQR therefore provides detailed additional insights about the role of the maternity benefits for the probability of terminating maternity which exceed and partly contradict with what can be concluded from the coefficients of the proportional hazards model. Particularly, maternity benefits tend to have contrasting effects on the employability of young mothers at different quantiles, while the proportional hazards model suggests that maternity benefits have no effect at all.

As a next step we present a direct comparison of estimated conditional cumulative incidences obtained by the CRQR model and the proportional hazards model. We compute these functionals for a reference mother which is defined by setting continuous variables to their sample averages and most dummy variables to 0 (compare Table 4 in the Appendix). The resulting cumulative incidences are shown in Figure 2. The plots suggest that estimated cumulative incidences partly differ significantly between the two models. In particular, the CRQR model tends to produce higher cumulative incidences than the proportional hazards model for longer

Table 2: Summary of the estimated partial relationship of policy variables and the probability of observing a transition.

Risk	next birth				unemployment		
	0.05	0.15	0.25	PHM*	0.05	0.15	PHM*
child care places	0	0	-	-	+	0	-
job protection 10-12m	0	++	++	++	--	--	0
job protection 15-18m	++	++	++	++	--	--	0
job protection 36m	++	++	++	++	--	--	--
maternity benefits	0	-	0	-	++	0	0

Legend: ++ strongly positive, + positive, 0: minor effect

- negative, -- strongly negative

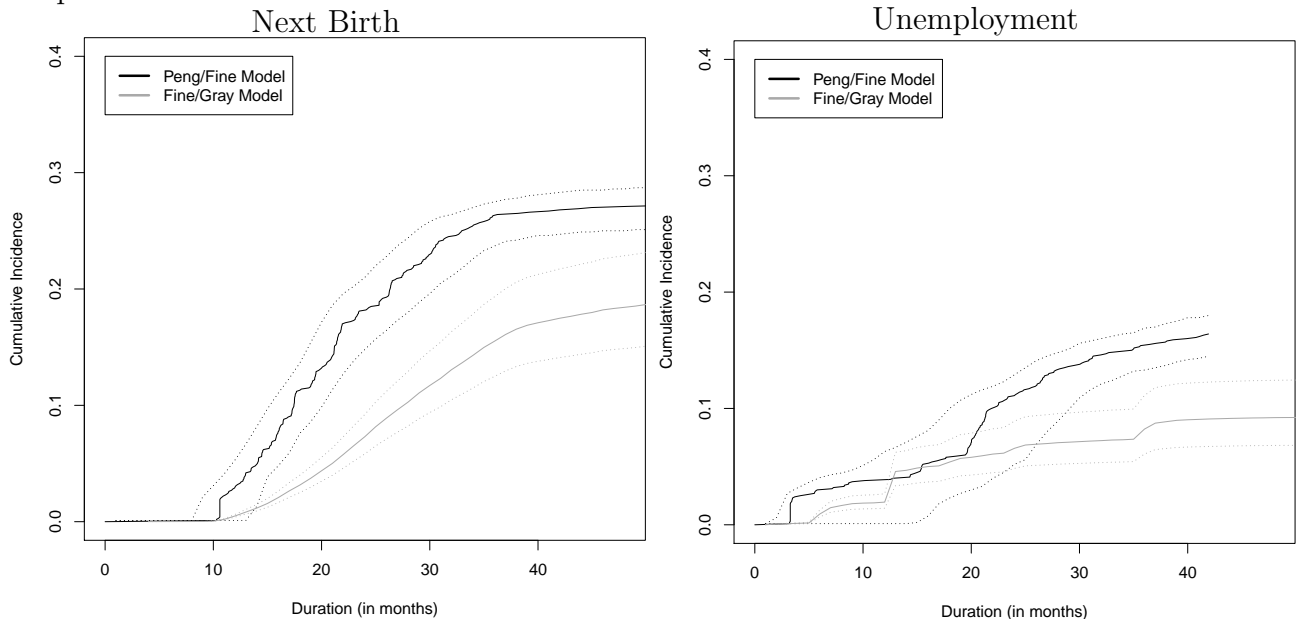
*: Proportional hazards model

durations. It is notable that the CRQR model estimates the plateau levels to be around 10% higher than those estimated by the proportional hazards model. It is also apparent from the figure that the CRQR model based cumulative incidence for "unemployment" cannot be obtained for durations greater than 42 months. This is because only few transitions into the relevant risk are observed beyond 42 months. However, we cannot reject either of these non-nested models on the grounds that implied cumulative incidences differ statistically. To shed more light on this issue, we have additionally estimated a fully nonparametric cumulative incidence functions using stratified samples based on one binary covariate and compared this with the CRQR model and proportional hazards model. This comparison provides strong evidence for the CRQR model to produce a better fit than the proportional hazards model as its implied cumulative incidences almost match the nonparametric estimates. These results are available on request. Unfortunately, we cannot compute nonparametric estimates in a higher dimensional setting due to the curse of dimensionality. It remains therefore unclear, which of the estimates in Figure 2 are preferable.

The estimation of the CRQR model is quick. In our application with around 20K observations it takes less than 16 minutes per risk on a AMD Opteron with 2.4GHz for a τ -grid of size 0.01 (i.e. 28 quantiles for risk "next birth" and 16 for "unemployment"). For comparison the R-implementation of the Fine and Gray (1999) model takes around 12 minutes per risk with the same sample. Computations of the QR model can easily be parallelised, thus computation time can be conveniently reduced.

We have seen that there is often only little variation in coefficients across quantiles for the risk "next birth". It would be an interesting extension to impose cross quantile restrictions on coefficients as in Qian and Peng (2010) to improve efficiency of the CRQR estimator.

Figure 2: Comparison of estimated cumulative incidences for reference mother with 95% bootstrap confidence interval.



5 Conclusion

We present a first application of a competing risks quantile regression model to large administrative data. We provide a fast implementation (R-package: `cmprskQR`) for the estimation of the quantile regression model and implied cumulative incidences. The latter are directly comparable to results from conventional duration models for cumulative incidences such as proportional hazards. Our application to German administrative maternity duration data reveals several interesting result patterns for the role of public child support. Some of the estimated coefficients change considerably over duration and partly they change sign, suggesting a varying influence of some of the covariates for lower quantiles and higher quantiles and thus shorter and longer durations are affected differently. This may point to conflicting findings and insights for a policy question when compared with the results of the proportional hazards model. For example, the CRQR model provides strong evidence for a negative relationship between the length of job protection period and the probability of making a transition to unemployment for all quantiles, but the proportional hazards model only finds this for the longest job protection period (36 months). The CRQR model estimates suggest a changing direction in the partial relationship between maternity benefits and the probability of observing a transition to unemployment. The proportional hazards model in contrast estimates this relationship to be small and insignificant. Thus, if maternity leave regulations aim at providing young mothers security to avoid unemployment, the CRQR coefficients appear to produce a more detailed and partly contradicting picture of the role of the various policy variables than the coefficients of the

proportional hazards model. We therefore consider the CRQR model as a useful complement to existing methods.

Even though, quantile regression is shown to provide interesting insights into the distribution of competing risks, the large number of results causes new challenges for the practitioner. These are more difficult to report and partly more difficult to interpret for non-experts as they are measured at quantiles rather than durations. Beside the usual graphical presentation of quantile regression coefficients across quantiles, we report a summary table comparing the direction of the estimated effects of the CRQR versus the proportional hazards model in a concise manner. Our proposed construction of the conditional cumulative incidences from the CRQR results is a useful tool for practitioners to make results comparable and interpretable in the duration dimension. The implied functionals for the two models are directly comparable.

References

- [1] Arntz, M., Dlugosz, S. & Wilke, R.A. (2014). The Sorting of Female Careers after First Birth: A Competing Risks Analysis of Out of Work Duration, *ZEW Discussion Paper*, No. 14-125.
- [2] Chernozhukov V., Fernández-Val I. & Galichon, A. (2010). Quantile and probability curves without crossing, *Econometrica*, 78, 1093–1125.
- [3] Fine J.P. & Gray R.J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94, 496–509.
- [4] Fitzenberger, B. & Wilke, R. A. (2006). Using Quantile Regression for Duration Analysis. *Allgemeines Statistisches Archiv*, 90, 103–118.
- [5] Hochfellner, D., Müller, D. & Wurdack, A. (2012). Biographical Data of Social Insurance Agencies in Germany - Improving the Content of Administrative Data. *Schmollers Jahrbuch*, 132, 443–451.
- [6] Hsieh, J., Ding, A., Wang, W. & Chi, Y. (2013). Quantile Regression Based on Semi-Competing Risks Data, *Open Journal of Statistics*, 3, 12–26.
- [7] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.
- [8] Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software*, 27, 1-24.
- [9] Koenker, R. & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46, 33-50.
- [10] Koenker, R. & Geling, O. (2001). Reappraising Medfly Longevity: A Quantile Regression Survival Analysis. *Journal of the American Statistical Association*, 96, 458–468.
- [11] Peng, L., & Fine, J. P. (2009). Competing risks quantile regression, *Journal of the American Statistical Association*, 104, 1440–1453.
- [12] Powell J.L. (1986). Censored Regression Quantiles. *Journal of Econometrics* 32,143–155.
- [13] Qian, J. & Peng, L (2010). Quantile regression with partially functional effects. *Biometrika*, 97, 839–850.

Appendix

Table 3: Descriptives

Variable	next birth		unemployment		sample				
	Mean	SD	Mean	SD	Mean	SD	Min	Max	
length	30.596	14.167	20.093	13.999	24.231	25.009	1	228	
individual age	26.734	3.802	26.750	4.088	27.275	4.271	18	45	
wage, 1st quintile	0.044	0.206	0.034	0.182	0.038	0.191	0	1	
wage, 2nd quintile	0.132	0.338	0.148	0.355	0.118	0.322	0	1	
wage, 4th quintile	0.303	0.459	0.297	0.457	0.302	0.459	0	1	
wage, 5th quintile	0.294	0.456	0.264	0.441	0.325	0.468	0	1	
labour market experience (months)	6.556	3.507	6.597	3.648	6.871	3.843	0.083	25.000	
dummy for past unemployment	0.284	0.451	0.352	0.478	0.296	0.456	0	1	
illness during pregnancy	0.088	0.748	0.184	1.407	0.108	0.902	0	27	
inactivity period during pregnancy	0.025	0.155	0.034	0.181	0.026	0.160	0	1	
tenure at current firm (months)	3.455	2.965	3.309	2.876	3.566	3.146	0.083	20.750	
decrease in wage quintile	0.109	0.312	0.100	0.300	0.088	0.284	0	1	
increase in wage quintile	0.124	0.330	0.136	0.343	0.123	0.328	0	1	
agriculture	0.021	0.142	0.009	0.094	0.011	0.106	0	1	
manual	0.148	0.355	0.180	0.384	0.154	0.360	0	1	
qualified manual	0.063	0.243	0.058	0.233	0.055	0.228	0	1	
technician	0.043	0.202	0.052	0.222	0.039	0.193	0	1	
engineer	0.008	0.089			0.007	0.081	0	1	
services	0.073	0.260	0.076	0.266	0.071	0.256	0	1	
qualified services	0.135	0.341	0.130	0.337	0.131	0.338	0	1	
semi professional	0.103	0.304	0.064	0.244	0.103	0.303	0	1	
professional	0.007	0.085			0.008	0.088	0	1	
business administration	0.131	0.338	0.143	0.350	0.126	0.332	0	1	
manager	0.015	0.123	0.010	0.101	0.015	0.122	0	1	
firm size < 20	0.331	0.471	0.334	0.472	0.298	0.457	0	1	
firm size > 1000	0.134	0.341	0.113	0.316	0.144	0.351	0	1	
share of female workers	0.649	0.268	0.633	0.279	0.622	0.281	0	1	
share of young workers	0.425	0.233	0.440	0.234	0.394	0.230	0	1	
GDP growth (in %)	1.912	1.553	1.938	1.635	1.885	1.543	-1.002	5.255	
unemployment rate (in %)	3.814	1.109	3.980	1.141	3.897	1.128	1.205	10.976	
child care places per 100 children*	1.883	3.776	1.762	4.173	2.317	4.920	0.200	37.936	
job protection 10-12 months	0.120	0.325	0.141	0.348	0.121	0.326	0	1	
job protection 15-18 months	0.127	0.333	0.152	0.359	0.124	0.329	0	1	
job protection 36 months	0.561	0.496	0.366	0.482	0.529	0.499	0	1	
maternity benefits (in 1000 Euros)	5.195	2.321	4.235	2.441	5.035	2.411	1.020	7.200	
regional dummy for Hessia	0.090	0.286	0.090	0.286	0.098	0.297	0	1	
regional dummy for BW	0.233	0.423	0.125	0.330	0.193	0.395	0	1	
regional dummy for Bavaria	0.241	0.428	0.232	0.422	0.225	0.417	0	1	
dummy for years 1990 to 1999	0.505	0.500	0.377	0.485	0.423	0.494	0	1	
dummy for years 2000 to 2004	0.141	0.348	0.096	0.294	0.188	0.391	0	1	

* aged <3

Table 4: Characteristics of the reference mother in Figure 2.

Variable	Value
wage, 1st quintile	0
wage, 2nd quintile	0
wage, 4th quintile	0
wage, 5th quintile	0
labour market experience (months)	6.9
tenure at current firm (months)	3.6
decrease in wage quintile	0
increase in wage quintile	0
agriculture	0
manual	0
qualified manual	0
technician	0
engineer	0
services	0
qualified services	0
semi professional	0
professional	0
business administration	0
manager	0
firm size <20	0
firm size >1000	0
share of female workers	0.6
share of young workers	0.4
firm information missing	0
individual age	27.3
illness during pregnancy	0
inactivity period during pregnancy	0
regional dummy for Hestia	0
regional dummy for Baden-Württemberg	0
regional dummy for Bavaria	0
dummy for past unemployment	0
dummy for years 1990 to 1999	0
dummy for years 2000 to 2004	1
GDP growth (in %)	1.9
unemployment rate (in %)	3.9
child care places per 100 children*	2.3
job protection 10-12 months	0
job protection 15-18 months	0
job protection 36 months	1
maternity benefits (in 1000 Euros)	5.0

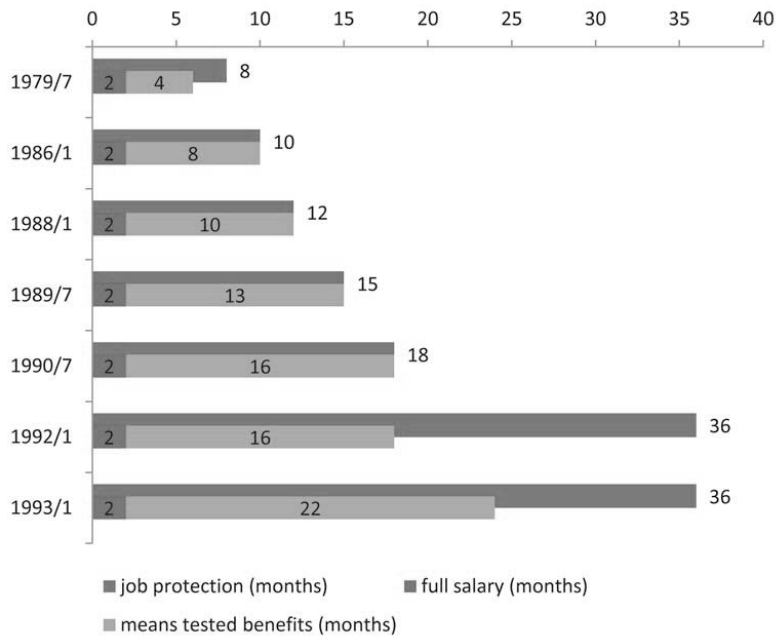
* aged <3

Table 5: Selected coefficients of the proportional hazard model

	next birth	unemployment
child care places per 100 children	-0.018***	-0.017***
job protection 10–12 months	0.380***	-0.094
job protection 15–18 months	0.532***	0.045
job protection 36 months	1.005***	-0.792***
maternity benefits (in 1000 Euros)	-0.086**	-0.051

Significance: * 10%, ** 5%, *** 1%

Figure 3: Job protection periods and duration of maternity benefits by regime (with start date).



Source: Arntz et. al (2014)