A Bayesian Semiparametric Competing Risk Model with Unobserved Heterogeneity*

Martin Burda,† Matthew Harding,‡ Jerry Hausman,§

September 28, 2013

Abstract

This paper generalizes existing econometric models for censored competing risks by introducing a new flexible specification based on a piecewise linear baseline hazard, time-varying regressors, and unobserved individual heterogeneity distributed as an infinite mixture of Generalized Inverse Gaussian (GIG) densities, nesting the gamma kernel as a special case. A common correlated latent time effect induces dependence among risks. Our model is based on underlying latent exit decisions in continuous time while only a time interval containing the exit time is observed, as is common in economic data. We do not make the simplifying assumption of discretizing exit decisions – our competing risk model setup allows for latent exit times of different risk types to be realized within the same time period. In this setting, we derive a tractable likelihood based on scaled GIG Laplace transforms and their higher-order derivatives. We apply our approach to analyzing the determinants of unemployment duration with exits to jobs in the same industry or a different industry among unemployment insurance recipients on nationally representative individual-level survey data from the U.S. Department of Labor. Our approach allows us to conduct a counterfactual policy experiment by changing the replacement rate: we find that the impact of its change on the probability of exit from unemployment is inelastic.

JEL: C11, C13, C14, C41, J64
Keywords: competing risk model, Bayesian semiparametric model, unemployment insurance.

*We would like to thank Arthur Lewbel, Herman van Dijk, Štěpán Jurajda, and participants of the Stanford Institute for Theoretical Economics (SITE) workshop, the Cowles Foundation Econometrics Summer Conference, Yale University 2011, the Tinbergen Institute Econometrics Lectures 2011, and seminar participants at Cornell, Harvard-MIT, Oxford University, and the University of Virginia for their useful comments and suggestions.

†Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada; Phone: (416) 978-4479; Email: martin.burda@utoronto.ca

‡Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: mch@stanford.edu

§Department of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02142; Email: jhausman@mit.edu

1. Introduction

In an economic competing risk (CR) model with censoring, an individual is associated with a current state (e.g. of being unemployed) with the possibility to exit to one of several different states (e.g. finding a job in the same industry or a different industry). However, only one such exit is observed for some individuals, while other (censored) individuals are never observed to exit their current state. The state duration before exit to each potential new state or censoring time is modeled with a separate latent variable but only the shortest such duration is actually observed for each individual. A key ingredient of CR models is the survival function which captures the probability that the individual will remain the current state beyond any given time. CR analysis then typically seeks to determine the impact of observable characteristics of the individual and the various states on the survival function that can lead to policy recommendations.[5]

In this paper we introduce a new flexible model specification for the competing risk model, extending several strands of econometric and statistical literature on duration analysis. Our model encompasses three key features: (i) we estimate non-parametrically the density of unobserved individual heterogeneity; (ii) we model correlations between the different risk types; (iii) we allow for multiple latent exits within a time period with interval outcome data. Our model nests the single exit type (so-called duration model) as a special case. We apply our method to analyzing the determinants of unemployment duration among unemployment insurance recipients using data from the U.S. Department of Labor. We conduct a counterfactual experiment by changing the replacement rate. The counterfactual results show the impact of changing key policy variables such as the replacement rate on the survival function.

We will introduce each model feature in turn and discuss the advantages of using our model over the existing alternatives. First, our model provides a flexible approach to controlling for unobserved heterogeneity in competing risk data. A typical source of unobserved

[5]Applications of CR models in economics include analyzing unemployment duration (Flinn and Heckman, 1982; Katz and Meyer, 1990; Tysse and Vaage, 1999; Alba-Ramírez, Arranz, and Muñoz-Bullón, 2007), Ph.D. completion (Booth and Satchell, 1995), teacher turnover (Dolton and van der Klaauw, 1999), studies of age at marriage or cohabitation (Berrington and Diamond; 2000), mortgage termination (Deng, Quigley, and Van Order, 2000), school dropout decisions (Jakobsen and Rosholm, 2003), and manufacturing firms' exits from the market (Esteve-Perez, Sanchis-Llopis, and Sanchis-Llopis, 2010). A comprehensive overview is given in Van den Berg (2001).

heterogeneity is the omission of important but perhaps unobservable variables from the conditioning set. As an example, more motivated individuals may exit unemployment more quickly because they put more effort into the search for a new job.

It is well established that failure to account for unobserved heterogeneity biases the estimated hazard rate and the proportional effects of explanatory variables on the population hazard (Vaupel et al 1979; Lancaster 1979, 1990). A number of semi-parametric estimators for the single risk mixed proportional hazard (MPH) model have been proposed following Elbers and Ridder (1982) proof of MPH semi-parametric identification (Heckman and Singer, 1984; Honoré 1990). Han and Hausman (1990) and Meyer (1990) propose an estimator for piecewise-constant baseline hazard and gamma distributed unobserved heterogeneity. Horowitz (1999) proposed a nonparametric estimator for both the baseline hazard and the distribution of the unobserved heterogeneity, under the assumption of constant time-invariant regressors. Hausman and Woutersen (2012) show that a nonparametric estimator of the baseline hazard with gamma heterogeneity yields inconsistent estimates for all parameters and functions if the true mixing distribution is not a gamma, stressing the importance of avoiding parametric assumptions on the unobserved heterogeneity.

As the second key feature of our model, our approach allows for correlations between the different risks in the CR model environment, even in the presence of the flexible individual heterogeneity infinite GIG mixture model component. This is important since the determinants of exit can differ depending on the risk type while being correlated across the risk types, and thus our approach provides additional information to the analyst compared to single-risk duration models.

Third, in our application, we deal with interval outcome data as is common in economics and other social sciences. Even though the underlying exit decision model is set in continuous time, only the broader period in which exit occurred is observable. Our data contain the week of exit from unemployment. Based on scaled GIG Laplace transforms and their higher-order derivatives, we provide a complete likelihood specification allowing for multiple latent exits within a single time period which is more realistic than simplifying the analysis by assuming that only one latent exit can occur in a given time period. Thus, we do not rule out by assumption an individual contemplating different job offers from firms in the same industry or in a different industry as their pre-unemployment industry within any given week.

We show that given the analytical forms newly provided in this paper our model can be implemented in a user-friendly way via a Bayesian nonparametric approach. One of the key benefits of Bayesian Markov chain Monte Carlo (MCMC) methods that we utilize is their ability to factorize a complicated joint likelihood model into a sequence of conditional tractable models, so-called Gibbs blocks, and by sampling each in turn deliver outcomes from the joint model. We detail this approach for our proposed model.

The bulk of the literature on CR model development is concentrated in the natural sciences. A recent overview of CR modeling in biostatistics is provided by Beyersmann, Schumacher, and Allignol (2012), and in medical research by Pintilie (2006). The associated estimation methods typically rely on continuous time data for the exact point of exit. In contrast, we observe only discrete time intervals within which latent exits occur.

Competing risk models suitable for economic interval outcome data have been proposed in various forms. Han and Hausman (1990), Fallick (1991), Sueyoshi (1992), and McCall (1996) provide model specifications either without or with parametric individual heterogeneity. Butler, Anderson, and Burkhauser (1989) propose a semiparametric CR model controlling for the correlation between unobserved heterogeneity components in each state, with quadratic time dependence. Lleras-Muney and Honoré (2006) analyze identification issues in a general class of competing risk models allowing for correlation among risk types. Their environment is free of many of the functional form and distributional assumptions that we impose here and hence a number of parameters are set-identified. Bierens and Carvalho (2007) consider Weibull baseline hazards and common flexible unobserved heterogeneity. Canals-Cerdá and Gurmu (2007) approximate unobserved heterogeneity distribution with Laguerre polynomials. They find that model selection rules (BIC, HQIC, and AIC) perform worse in determining the polynomial order than a naive approach of controlling for unobserved heterogeneity using simple models with a small number of points of support or a polynomial of small degree. Van den Berg, van Lomwel, and van Ours (2008) consider a model with nonparametric unobserved heterogeneity terms that is based on discrete time counts. Although the model can be derived as a time-aggregated version of an underlying continuous-time model, the latter is different from the continuous-time mixed proportional hazard model.

The literature on Bayesian nonparametric methods in the CR environment has been scant and, to our knowledge, has only been used in biostatistics for estimation of other objects of interest than individual heterogeneity.[6] Variants of Bayesian Dirichlet Process analysis have been used by Gasbarra and Karia (2000) for estimating nonparametically the overall hazard rate and in Salinas-Torres, Pereira, and Tiwari (2002) and Polpo and Sinha (2011) for the vector of risk-specific cumulative incidence functions. De Blasi and Hjort (2007) specify a beta-process prior for the baseline hazard, with asymptotic properties analyzed in De Blasi and Hjort (2009).

Identification results under various assumptions were established by Heckman and Honoré (1989), Sueyoshi (1992), Abbring and van den Berg (2003), and Lee and Lewbel (2013). In general, there have been three different approaches to identification (Honoré and Lleras-Muney, 2006): (a) to make no additional assumptions beyond the latent competing risk structure and estimate bounds on the objects of interest; (b), assume that the risks are independent conditional on a set of observed covariates and deal with a multiple duration models environment; and (c), to specify a parametric or semi-parametric model conditional on the covariates. Here we take the last approach. In particular, we do not assume that the risks are independent conditional on the observed covariates.

The remainder of the paper is organized as follows. Section 2 establishes the assumptions and building block results for a single risk duration model. Section 3 introduces assumptions and results for the competing risk model. Section 4 details our application and the counterfactual experiment and Section 5 concludes. Proofs or all theorems and additional empirical results are provided in the Appendix. The Online Appendix (Burda, Harding and Hausman, 2013) contains further results.

[6]In the single-risk, duration model case, Bayesian analysis with economics application was undertaken by Ruggiero (1994), Florens, Mouchart, and Rolin (1999), Campolieti (2001), Paserman (2004), and Li (2007).

2. Single Risk Duration Model

Denote by τ a continuous time variable with density $f(\tau)$ and distribution $F(\tau)$. Denote a latent failure (or exit) time of individual i by τ_i. Define the hazard rate $\lambda_i(\tau)$ as the failure rate at time τ conditional upon survival to time τ, $\lambda_i(\tau) = \lim_{\delta \to 0} Pr(\tau < \tau_i < \tau + \delta | \tau_i \geq \tau)/\delta$ and denote the integrated hazard by:

$$\Lambda_i(\tau) = \int_0^\tau \lambda_i(u)du \tag{2.1}$$

with survivor function

$$S_i(\tau) = \exp(-\Lambda_i(\tau)) \tag{2.2}$$

Denote by t a generic time period $[\underline{\tau}, \overline{\tau})$ with end points $\underline{\tau}$ and $\overline{\tau}$, and by t_i the time period $[\underline{\tau}_i, \overline{\tau}_i) \ni \tau_i$ in which an individual i was observed to exit from a given state into another state.

ASSUMPTION (A1)**.** *The data $\{t_i\}_{i=1}^N$ consists of single spells censored at time T and drawn from a single risk process.*

ASSUMPTION (A2)**.** *The hazard rate is parameterized as*

$$\lambda_i(\tau) = \lambda_0(\tau) \exp(X_i(\tau)\beta + V_i) \tag{2.3}$$

where $\lambda_0(\tau)$ is the baseline hazard, $X_i(\tau)$ are observed covariates that are allowed to vary over time, β are model parameters, and V_i is an unobserved heterogeneity component.

Hence, using (2.1) and (2.3) the integrated hazard is given by

$$\Lambda_i(\tau) = \int_0^\tau \lambda_0(u) \exp(X_i(u)\beta + V_i) du \tag{2.4}$$

ASSUMPTION (A3)**.** *The baseline hazard $\lambda_0(u)$ and the values of the covariates are constant for each time period t.*

Assumptions 1 and 2 are common in the literature. Assumption A3 is based on Han and Hausman (1990). Given Assumption A3, instead of $\lambda_i(\tau)$ we can consider the integrated baseline hazard in the form

$$\mu_{0t} = \int_{\underline{\tau}}^{\overline{\tau}} \lambda_0(u)du, \tag{2.5}$$

where we denote the vector $(\mu_{01,}, \ldots, \mu_{0T})$ by μ_0.

For notational convenience, we will use subscripts for the time index and denote by Λ_{it} the quantity $\Lambda_i(\overline{\tau})$ at the end of the time period t, and similarly for other variables. Denote the probability of the exit event in time period t by $P(t_i = t)$. Conditional on V_i, for outcomes that are not censored ($t_i \leq T$),

$$P(t_i = t) = F_{it} - F_{i(t-1)} = (1 - S_{it}) - \left(1 - S_{i(t-1)}\right) = S_{i(t-1)} - S_{it} \tag{2.6}$$

When the duration observations are censored at the end of time period T,

$$P(t_i > T) = 1 - F_{iT} = S_{iT} \tag{2.7}$$

The following result is familiar in the literature and we include it here for the sake of completeness as a benchmark of comparison for the new competing risk model developed in the next Section.

RESULT 1. *Under Assumptions A1–A3, conditional on V_i, for uncensored observations*

$$P(t_i = t|V_i) = \exp\left(-\sum_{j=1}^{t-1} \mu_{0j} \exp\left(X_{ij}\beta + V_i\right)\right) - \exp\left(-\sum_{j=1}^{t} \mu_{0j} \exp\left(X_{ij}\beta + V_i\right)\right) \tag{2.8}$$

and for the censored case

$$P(t_i > T|V_i) = \exp\left(-\sum_{j=1}^{T} \mu_{0j} \exp\left(X_{ij}\beta + V_i\right)\right) \tag{2.9}$$

2.1. **Parametric Heterogeneity**

ASSUMPTION (A4)**.** *Let $v_i \equiv \exp(V_i) \sim \mathcal{G}(v)$ where $\mathcal{G}(v)$ is a generic probability measure with density $g(v)$.*

From (2.4), (2.5), and Assumption A3, we have

$$\widetilde{\Lambda}_{it} = \sum_{j=1}^{t} \mu_{0j} \exp\left(X_{ij}\beta\right) \tag{2.10}$$

$$\Lambda_{it} = v_i \widetilde{\Lambda}_{it} \tag{2.11}$$

If v is a random variable with probability density function $g(v)$ then the Laplace transform of $g(v)$ evaluated at $s \in \mathbb{R}$ is defined as

$$\mathcal{L}(s) \equiv E_v[\exp(-vs)] \tag{2.12}$$

Using (2.2), (2.11), and (2.12), the expectation of the survival function can be linked to the Laplace transform of the integrated hazard function (Hougaard, 2000) as follows:

$$E_v\left[S_{it}\right] = \mathcal{L}(\widetilde{\Lambda}_{it}) \tag{2.13}$$

Using (2.10), (2.11), and (2.13) yields the unconditional exit probability of Result 1 as follows:

RESULT 2. *The expectation of (2.6) for the uncensored observations is*

$$E_{v_i}\left[P(t_i = t)\right] = \mathcal{L}(\widetilde{\Lambda}_{i(t-1)}) - \mathcal{L}(\widetilde{\Lambda}_{it}) \tag{2.14}$$

and the expectation of (2.7) for the censored observations takes the form

$$E_{v_i}\left[P(t_i > T)\right] = \mathcal{L}(\widetilde{\Lambda}_{iT}) \tag{2.15}$$

Since the individual heterogeneity term v_i defined in Assumption A4 is non-negative, a suitable family of distributions $\mathcal{G}(v)$ with support over $[0,\infty)$ and tractable closed-form Laplace transforms is Generalized Inverse Gaussian (GIG) class of distributions, whose special case is the gamma distribution popular in duration analysis.

ASSUMPTION (A5a)**.** *The unobserved heterogeneity term v_i is distributed according to the Generalized Inverse Gaussian distribution, $\mathcal{G}(v) = \mathcal{G}^{GIG}(v;\kappa,\varphi,\theta)$.*

The GIG has the density

$$g^{GIG}(v;\kappa,\varphi,\theta) = \frac{2^{\kappa-1}}{K_\kappa(\varphi)}\frac{\theta}{\varphi^\kappa}(\theta v)^{\kappa-1}\exp\left\{-\theta v - \frac{\varphi^2}{4\theta v}\right\} \tag{2.16}$$

for $\varphi,\theta > 0$, $\kappa \in \mathbb{R}$, where $K_\kappa\left(\varphi\right)$ is the modified Bessel function of the second kind of order κ evaluated at φ (Hougaard, 2000). The GIG Laplace transform is given by

$$\mathcal{L}^{GIG}(s;\kappa,\varphi,\theta) = (1+s/\theta)^{-\kappa/2}\,\frac{K_\kappa\left(\varphi\left(1+s/\theta\right)^{1/2}\right)}{K_\kappa\left(\varphi\right)} \tag{2.17}$$

The GIG family includes as special cases the gamma distribution for $\varphi = 0$, the Inverse gamma distribution for $\theta = 0$, and the Inverse Gaussian distribution for $\kappa = -\frac{1}{2}$, among others.

Application of the Laplace transform of the GIG distribution (2.17) in Result 2 yields the following result that appears to not have been previously stated in the literature:

RESULT 3. *Under the Assumptions A1–A4, and A5a*

$$E_{v_i}^{GIG}\left[P(t_i = t)\right] = \left(1+\frac{1}{\theta}\widetilde{\Lambda}_{i(t-1)}\right)^{-\kappa/2} \frac{K_\kappa\left(\varphi\left(1+\frac{1}{\theta}\widetilde{\Lambda}_{i(t-1)}\right)^{1/2}\right)}{K_\kappa\left(\varphi\right)} - \left(1+\frac{1}{\theta}\widetilde{\Lambda}_{it}\right)^{-\kappa/2} \frac{K_\kappa\left(\varphi\left(1+\frac{1}{\theta}\widetilde{\Lambda}_{it}\right)^{1/2}\right)}{K_\kappa\left(\varphi\right)} \tag{2.18}$$

and for the censored observations

$$E_{v_i}^{GIG}\left[P(t_i > T)\right] = \left(1+\frac{1}{\theta}\widetilde{\Lambda}_{iT}\right)^{-\kappa/2} \frac{K_\kappa\left(\varphi\left(1+\frac{1}{\theta}\widetilde{\Lambda}_{iT}\right)^{1/2}\right)}{K_\kappa\left(\varphi\right)} \tag{2.19}$$

A special case of the GIG distribution is the gamma distribution, obtained from the GIG density function (2.16) when $\varphi = 0$. We use the gamma distribution for v_i as a benchmark model under the following alternative to Assumption 5a.

ASSUMPTION (A5b)**.** *The unobserved heterogeneity term v_i is distributed according to the gamma distribution, $\mathcal{G}(v) = \mathcal{G}^G(v;\gamma,\theta)$.*

The gamma density is parameterized as

$$g^G(v;\gamma,\theta) = \frac{\theta}{\Gamma(\gamma)}\left(\theta v\right)^{\gamma-1}\exp(-\theta v) \tag{2.20}$$

and its Laplace transform is given by

$$\mathcal{L}^G(s;\gamma,\theta) = (1+s/\theta)^{-\gamma} \tag{2.21}$$

In the gamma density (2.20) the parameter $\gamma > 0$ corresponds to the GIG parameter $\kappa \in \mathbb{R}$ in (2.16) restricted to the positive part of the real line. Using the gamma distribution in place of the GIG constitutes a special case of Result 3:

RESULT 4. *Under the Assumptions A1–A4, and A5b,*

$$E_{v_i}^{G}\left[P(t_i = t)\right] = \left(1+\frac{1}{\theta}\widetilde{\Lambda}_{i(t-1)}\right)^{-\gamma} - \left(1+\frac{1}{\theta}\widetilde{\Lambda}_{it}\right)^{-\gamma} \tag{2.22}$$

and

$$E_{v_i}^{G}\left[P(t_i > t)\right] = \left(1+\frac{1}{\theta}\widetilde{\Lambda}_{it}\right)^{-\gamma} \tag{2.23}$$

Result 4 was obtained in Han and Hausman (1990) and Meyer (1990).

In both gamma and GIG distributions, the scale parameter θ performs the same role. Specifically, for any $c \in \mathbb{R}_+$, if $v \sim \mathcal{G}^G(v;\gamma,\theta)$ then $cv \sim \mathcal{G}^G(v;\gamma,\theta/c)$, and if $v \sim \mathcal{G}^{GIG}(v;\kappa,\varphi,\theta)$ then $cv \sim \mathcal{G}^{GIG}(v;\kappa,\varphi,\theta/c)$. Due to this property, c and hence its inverse $s \equiv c^{-1}$ are not separately identified from θ in the Laplace transform expressions (2.17) and (2.21). Since all likelihood expressions are evaluated at $s = \widetilde{\Lambda}_{i(\cdot)}$ which is proportional to μ_{0j} for all j, as specified in (2.10), any change in θ only rescales the baseline hazard parameters μ_{0j}, leaving the likelihood unchanged. Hence, θ needs to be normalized to identify μ_{0j}. In the gamma case, typically this normalization takes the form $\theta = \gamma$ so that $E[v] = 1$. We use the equivalent normalization for the GIG case in order to nest the normalized gamma as a special case and to maintain the moment restriction $E[v] = 1$.

2.2. **Flexible Heterogeneity**

We now depart from the parametric form of the unobserved heterogeneity and consider a nonparametric infinite mixture for the distribution of v_i, as formulated in the following assumption.

ASSUMPTION (A6)**.** *The prior for* $v_i|G \sim G$ *takes the form of the hierarchical model* $G \sim DP(G_0,\alpha)$, $\alpha \sim g^G(a_0,b_0)$, $E[v_i] = 1$.

In Assumption A6, G is a random probability measure distributed according to a Dirichlet Process (DP) prior (Hirano, 2002; Chib and Hamilton, 2002). The DP prior is indexed by two hyperparameters: a so-called baseline distribution G_0 that defines the "location" of the DP prior, and a positive scalar precision parameter α. The distribution G_0 may be viewed as the prior that would be used in a typical parametric analysis. The flexibility of the DP mixture model environment stems from allowing G to stochastically deviate from G_0. The precision parameter α determines the concentration of the prior for G around the DP prior location G_0 and thus measures the strength of belief in G_0. For large values of α, a sampled G is very likely to be close to G_0, and vice versa. Assumption A6 is then completed by specifying the baseline measure G_0. We consider two cases:

ASSUMPTION (A7a)**.** *In Assumption A6,* $G_0 = \mathcal{G}^{GIG}(\kappa,\varphi,\theta)$.

Implementation of the GIG mixture model under Assumptions A1–A3, A6, and A7a uses the probabilities (2.8), (2.9), (2.18) and (2.19). Further implementation details are given in the Online Appendix (Burda, Harding and Hausman, 2013).

ASSUMPTION (A7b)**.** *In Assumption A6,* $G_0 = \mathcal{G}^G(\gamma, \theta)$.

Under Assumptions A6 and A7b, as a special limit case, putting all the prior probability on the baseline distribution G_0 by setting $\alpha \to \infty$ would result in forcing $G = G_0 = \mathcal{G}^G(\gamma, \theta)$ which yields the parametric model of Han and Hausman (1990). Here we allow α and hence G to vary stochastically. Furthermore, the gamma baseline in Assumption A7b results as a special case of the GIG baseline in Assumption A7a for the hyperparameter value $\varphi = 0$. Hence, both the gamma flexible case with $G \sim DP(\mathcal{G}^G, \alpha)$ and the parametric benchmark Han and Hausman (1990) case with $G = \mathcal{G}^G$ are nested within our full GIG mixture model specification.

3. Competing Risk Model

We will now generalize the results from the single-risk case to the competing risk (CR) environment with several different potential types of exit. Let the risk type be indexed by $k = 1, \ldots, K$. Define the latent failure (or exit) times as $\tau_{1i}, \ldots, \tau_{Ki}$ corresponding to each risk type k, for each individual i. Define their minimum by $\tau_i \equiv \min(\tau_{1i}, \ldots, \tau_{Ki})$. In our CR model for interval outcome data, τ_i is not directly observed. Instead, the observed quantity is the time interval $[\underline{\tau}_i, \overline{\tau}_i)$ labeled as t_i which contains τ_i. This is in contrast to a large class of other types of CR models where the exact failure time τ_i is observed, as is typical in biostatistics. Intrinsically, the lifetimes of other risk types, τ_{ji} for $j \neq k$, and their corresponding time intervals, remain unobserved.

Denote by $f(u_1, u_2)$ the joint density of failure at time $u = (u_1, u_2)$. The functional form of $f(u_1, u_2)$ is provided in the Online Appendix, (OA.2.4). For two risk types with $K = 2$, this yields the probability of exit in time period t of the form

$$(3.1) \qquad P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}) = \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{u_1}^{\overline{\tau}_i} f(u_1, u_2) du_2 du_1 + \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{\overline{\tau}_i}^{\infty} f(u_1, u_2) du_2 du_1$$

The first right-hand side term in (3.1) gives the probability that the second latent exit time occurred within the same time interval t as the first latent exit time. The second right-hand side term in (3.1) is then the probability that the second latent exit time occurred in a later time interval than t. A key difficulty with evaluating (3.1), precluding direct factorization, is the presence of the outer integrand u_1 in the lower bound in the inner integral of the first term. We deal with this issue and derive a closed-form solution for (3.1), under various assumptions on the latent model components. The joint density $f(u_1, u_2)$ is obtained as a function of covariates and unobserved heterogeneity from the parameterization of risk-specific hazard functions, in a direct analogy to the single-risk case. Previous work using CR interval outcome data has either bypassed this link (e.g. by assuming a multivariate Gaussian density for $f(u_1, u_2)$) or employed a discrete time approximation whereby only one exit type can occur per any one time period. Our model explicitly accounts for the continuous-time nature of the exit decisions. The statistical background for the stochastic environment of our CR model is given in the Online Appendix (Burda, Harding and Hausman, 2013).

For clarity of exposition, the numbering of the Assumptions and Theorems in this Section provides a direct counterpart to the Assumptions and Results of the single-risk case in the previous Section. We first treat the parametric case under the GIG and gamma distributions of unobserved heterogeneity, adding a common latent component for all risk types, and then proceed to infinite mixture modeling.

ASSUMPTION (B1)**.** *The data consists of single spell data, drawn from a process with two risks* $k = 1, 2$, *and is censored at* T_k.

Assumption B1 readily generalizes to an arbitrary number of risks. Without loss of generality, suppose that the failure type is of type 1 so that $\tau_{1i} = \min(\tau_{1i}, \tau_{2i})$.

ASSUMPTION (B2)**.** *The risk-specific hazard rate is parameterized as*

$$\lambda_{ki}(\tau) = \lambda_{0k}(\tau) \exp(X_i(\tau)\beta_k + V_{ki} + \zeta_k(\tau))$$

where $\lambda_{0k}(\tau) > 0$ *is the baseline hazard whose logarithm is independent across* k, $X_i(\tau)$ *are covariates that are allowed to vary over time with full support on all of the real line for any given* τ *and one covariate common to all* k, β_k *are model parameters,* V_{ki} *is an unobserved heterogeneity component, and* $\zeta_k(\tau)$ *is a common time component correlated across* k *normalized to have mean zero in each* k.

ASSUMPTION (B3)**.** *For each k, the baseline hazard $\lambda_{0k}(\tau)$, the values of the covariates, and $\zeta_k(\tau)$ are constant within each time period t.*

Thus, the log-baseline hazard could also be stated as $\delta_{0kt} = \log(\lambda_{0kt}) + \zeta_{kt}$ where δ_{0kt} is correlated across k due to the presence of ζ_{kt}. Let $\zeta_k = (\zeta_{k1}, \ldots, \zeta_{kT})$, $\zeta = (\zeta_1, \zeta_2)$, $V_i = \{V_{ki}\}_{k=1}^K$, and $V = \{V_i\}_{i=1}^N$. The probability (3.1), conditional on (V_i, ζ) and a set of covariates, is

(3.2)

$$P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}|V_i, \zeta) = \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{u_1}^{\overline{\tau}_i} f(u_1, u_2|V_i, \zeta)du_2du_1 + \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{\overline{\tau}_i}^{\infty} f(u_1, u_2|V_i, \zeta)du_2du_1$$

Let $S_{kit} = \exp(-\Lambda_{kit})$ denote the risk-specific survivor function. We derive a closed-form solution to (3.2) in the following Theorem which extends Result 1 to our CR model environment.

THEOREM 1. *Under Assumptions B1–B3, conditional on the latent vector (V_i, ζ) and a set of covariates,*

(3.3)

$$P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}|V_i, \zeta) = S_{2i(t-1)}S_{1i(t-1)}\lambda_{1it}\left(\lambda_{2it} + \lambda_{1it}\right)^{-1}\left[1 - \exp\left(-\left(\lambda_{2it} + \lambda_{1it}\right)\right)\right]$$

for uncensored observations, and

$$P(t_{1i} > T,\ t_{2i} > T|V_i, \zeta) = \left(1 - F_{1iT}\right)\left(1 - F_{2iT}\right) = S_{1iT}S_{2iT} \tag{3.4}$$

for censored observations.

The proof is provided in the Online Appendix.

3.1. **Parametric Heterogeneity in the CR Model**

ASSUMPTION (B4)**.** *Let $v_{ki} \equiv \exp(V_{ki}) \sim \mathcal{G}_k(v_k)$ where $\mathcal{G}_k(v_k)$ is a generic probability measure with density $g_k(v_k)$. Let the correlation structure of ζ_{kt} be given by*

$$\begin{pmatrix} \zeta_{1t} \\ \zeta_{2t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}\right)$$

with parameters ρ, σ_1, σ_2.

As in the single-risk case, we consider two alternative forms of the distribution of unobserved heterogeneity $\mathcal{G}(v)$ in Assumption B4. The first form is parametric given either by the GIG or gamma density. We provide new results regarding the model likelihood for the model B1–B4. These will be used in the nonparametric mixture model. This approach is different from Han and Hausman (1990) who considered the truncated multivariate Normal likelihood.

For the expected likelihood, we have two new expression for the expected probability of (3.3): one based on a quadrature, and another one with a series expansion without the need for a quadrature. The following Theorem extends Result 2 into the CR model environment.

THEOREM 2. *Under Assumptions B1–B4,*

$$E_v P(t_{1i} = t, \tau_{2i} > \tau_{1i}) = -\widetilde{\lambda}_{1it} \int_0^1 \mathcal{L}_2\left(\widetilde{\Lambda}_{2i(t-1)} + \widetilde{\lambda}_{2it} s_1\right) \mathcal{L}_1^{(1)}\left(\widetilde{\Lambda}_{1i(t-1)} + \widetilde{\lambda}_{1it} s_1\right) ds_1 \tag{3.5}$$

or

$$E_v P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}) = \sum_{r_2=0}^{\infty} \sum_{r_1=0}^{\infty} \frac{(-1)^{2r_1+2r_2+1}}{r_2! r_1! \left(r_2 + r_1 + 1\right)} \widetilde{\lambda}_{1it}^{r_1+1} \widetilde{\lambda}_{2it}^{r_2} \times \mathcal{L}_1^{(r_1+1)}\left(\widetilde{\Lambda}_{1i(t-1)}\right) \mathcal{L}_2^{(r_2)}\left(\widetilde{\Lambda}_{2i(t-1)}\right) \tag{3.6}$$

for uncensored observations, and

$$E_v P(t_{1i} > T,\ t_{2i} > T) = \mathcal{L}_1\left(\widetilde{\Lambda}_{1iT}\right) \mathcal{L}_2\left(\widetilde{\Lambda}_{2iT}\right) \tag{3.7}$$

for censored observations, where $\mathcal{L}_k^{(r)}(s)$ is the $r-th$ derivative of the Laplace transform.

The proof is given in the Online Appendix. Theorem 2 is derived for a generic distribution of the unobserved individual heterogeneity term v_i and provides a direct extension of (2.14) and (2.15) to the competing risk model environment. Specific alternative distributional assumptions with corresponding likelihood expressions are provided next.

ASSUMPTION (B5a)**.** *The unobserved heterogeneity term v_i is distributed according to the Generalized Inverse Gaussian distribution, $\mathcal{G}(v) = \mathcal{G}^{GIG}(v; \kappa, \varphi, \theta)$.*

ASSUMPTION (B5b)**.** *The unobserved heterogeneity term v_i is distributed according to the gamma distribution, $\mathcal{G}(v) = \mathcal{G}^{G}(v; \gamma, \theta)$.*

The derivatives of the Laplace transform in (3.5) and (3.6) Theorem 2 depend on the functional form of the density kernel of v_i, given in Assumptions B5a and B5b. The formulas for the derivatives of arbitrary order of the Laplace transform for the GIG or gamma densities do not appear to be available in the literature; we derive them in the Online Appendix. Using those expressions in Theorem 2 yields the following two Corollaries, extending Result 3 and 4, respectively, from the single-risk case to the competing risk model environment.

Corollary 1 (to Theorem 2)**.** *Under Assumptions B1–B4 and B5a, the functional forms of Theorem 2 are given in (OA.2.43), (OA.2.44), and (OA.2.45) in the Online Appendix.*

Corollary 2 (to Theorem 2)**.** *Under Assumptions B1–B4 and B5b, the functional forms of Theorem 2 are given in (OA.2.47), (OA.2.48), and (OA.2.49) in the Online Appendix*

3.2. **Flexible Heterogeneity in the CR Model**

We will now proceed to an infinite mixture model for the distribution of v_{ki}.

ASSUMPTION (B6)**.** *The prior for* $v_{ki}|G_k \sim G_k$ *is specified as the hierarchical model* $G_k \sim DP(G_{0k}, \alpha_k)$, $\alpha_k \sim g^G(a_{0k}, b_{0k})$, $E[v_{ki}] = 1$.

The roles of the individual model components are described in Assumption A6 and generalize to the CR framework. Similarly to the single risk model environment, we consider two cases for the functional form of the baseline measure G_0:

ASSUMPTION (B7a)**.** *In Assumption B6,* $G_{0k} = \mathcal{G}^{GIG}(\kappa_k, \varphi_k, \theta_k)$.

ASSUMPTION (B7b)**.** *In Assumption B6,* $G_{0k} = \mathcal{G}^{G}(\gamma_k, \theta_k)$.

Implementation of the mixture models under Assumptions B1–B3, B6, and B7a or B7b uses the probabilities derived in Theorem 1, Corollary 1, and Corollary 2. Further implementation details are given in the next Section and in the Online Appendix.

Heckman and Honoré (1989) show how the introduction of covariates allows identification of a large class of dependent competing risks models without invoking distributional assumptions. Nonetheless, normalization assumptions are necessary for parameter identification. The normalization constraints generalize directly from the single-risk case and we impose them for each risk type. Assumptions B2 and B4 impose explicit restrictions on the model behavior in continuous time within each time period. These restrictions allow us to invoke identification conditions of Heckman and Honoré (1989). A detailed discussion of identification is provided in the Online Appendix.

4. MCMC Posterior Sampling

4.1. Single Risk Model

All technical implementation details are provided in the Online Appendix. In this Section we summarize the main points. For the implementation of the Dirichlet Process Mixture model (Assumptions A6 and A7a,b) we used the Bayesian generalized Pólya urn scheme (Neal 2000 Algorithm 2; West, Müller, and Escobar, 1994; Bush and MacEachern, 1996). Implementation of the GIG mixture model (Assumptions A1–A3, A6, and A7a) uses the functions (2.8), (2.9), (2.18), and (2.19). The gamma mixture model (Assumptions A1–A3, A6, and A7b) uses (2.8), (2.9), (2.22) and (2.23), respectively. The remaining model parameters were sampled in standard Gibbs blocks using Hybrid Monte Carlo (Neal 2011) with diffuse priors unless stated otherwise above. The results are discussed in our application below.

4.2. Competing Risk model

Similarly to the single-risk case, for the implementation of the Dirichlet Process Mixture model in the competing risk environment (Assumptions B6 and B7a,b) we also used the Bayesian generalized Pólya urn scheme. Implementation of the GIG mixture model (Assumptions B1–B3, B6, and B7a) uses the functions (3.3) and (3.4) from Theorem 1, and Corollary 1 to Theorem 2. The gamma mixture model (Assumptions B1–B3, B6, and B7b) uses (OA.2.47) and (OA.2.49) from Corollary 2 to Theorem 2. The remaining model parameters were sampled in standard Gibbs blocks using Hybrid Monte Carlo (Neal 2011) with diffuse priors unless stated otherwise above, except the covariance matrix of Assumption B4 with parameters σ_1, σ_2, ρ for which we specify a proper Inverse Wishart prior with maximum possible dispersion, $IW(K+1, I_K)$ where I_K is the identity matrix of dimension $K = 2$. The results are discussed in our application below.

5. Application

Since its introduction in 1935 as part of Roosevelt's *Social Security Act*, unemployment insurance (UI) benefits provide partial insurance to workers who become unemployed. Most states offer unemployment insurance for up to 26 weeks. Neoclassical economic thought suggests that higher benefits also lead to reduced incentives to search for a job, thus prolonging the period of time an individual spends out of employment (Mulligan, 2012). As a result policy makers have placed increased emphasis on reforming the UI system by rewarding personal responsibility rather than bad luck. This has lead to a shift away from the unqualified provision of UI benefits towards a system that is search intensive, making benefits conditional on providing evidence that the potential recipient engaged in a certain minimum amount of job search. Additionally, schemes whereby individuals are provided one-off grants that attempt to alleviate temporary hardship rather than longer term UI benefits are advocated. At the same time the recent economic crisis has forced policy makers to extend the duration of UI benefits for up to 99 weeks.[7]

Applied economists require econometric tools to accurately estimate the impact of unemployment insurance on the duration of unemployment, while accounting for state unemployment rates, generosity of unemployment insurance benefits and workers' observed and unobserved heterogeneity. In this section we apply our approach to analyzing the determinants of unemployment duration among unemployment insurance recipients. We will stress the importance of relaxing the parametric assumptions of the econometric models and accounting for correlations in the competing exit choices faced by workers. One of the major advantages of our approach is that it is possible to simulate counterfactual policy changes which can inform policy makers on the relative merits of various changes that may be contemplated. We will illustrate this feature by evaluating the impact of a change in the replacement rate on the duration of unemployment.

5.1. Data

We use data from the Needels et. al. (2001) report submitted to the U.S. Department of Labor that is based on a nationally representative sample built from individual-level surveys of unemployment insurance (UI) recipients in 25 states between 1998 and 2001.

[7]The unemployment extension legislation was set to expire on January 1, 2013.

Candidates for the survey are selected on the basis of administrative records and are sampled from the pool of unemployed individuals that started collecting UI benefits at some point during the year 1998.

We are interested in analyzing the effect of unemployment insurance on the duration of unemployment. The duration of unemployment is measured in weeks. At the time of the survey and from the states that were included in the survey only two states provided UI benefits for a maximum of 30 weeks, the rest providing UI benefits for a maximum of 26 weeks. Theoretical models of the impact of UI benefits on unemployment duration, such as Mortensen (1977) and Moffitt and Nicholson (1982), predict an increasing hazard up to the point of benefit exhaustion and a flat one afterwards. We limit our study to the first 24 weeks of unemployment due to the recognized change in behavior in week 26 when UI benefits cease for a significant part of the sample (see, e.g., Han and Hausman, 1990), which would affect the econometric model.

The data contain individual-level information about labor market and other activities from the time the person entered the UI system through the time of the interview. The data include information about the individual's pre-UI job, other income or assistance received, and demographic information. We use two indicator variables, race (defined as an indicator for black) and age (defined as an indicator for over 50). We further use the replacement rate, which is the weekly benefit amount divided by the UI recipient's base period earnings. Lastly, we utilize the state unemployment rate of the state from which the individual received UI benefits during the period in which the individual filed for benefits. This variable changes over time. The Needels et. al. (2001) data shares certain similarities to the PSID dataset used in Han and Hausman (1990). The UI recipients are mostly white, young, poorly educated workers who find themselves below or very near the poverty line.[8]

Below we will estimate a single risk duration model for the duration to re-employment and also a competing risk duration model for the duration to re-employment using our proposed approach, which differentiates between workers who find a job in the same

[8] Note that the labor market conditions captured in this dataset are substantially different than the ones experienced today. According to the Bureau of Labor Statistics (BLS), the latest figures broken down by state for September 2012 show the mean state unemployment rate is 7.5% and varies between 3% and 11.8%. In contrast, in our dataset the state unemployment rate is approximately 4.5%.

industry and workers who find a job in a different industry. For both analyses we employ subsamples from the Needels et. al. (2001) survey data, after removing outliers and observations with missing values. For the single risk model we use a subsample consisting of 15,358 individuals. Summary statistics for this sample are given in Table 2. For a subsample of 1,243 of the Needels et. al. (2001) data we also know the SIC codes of the employer before and after the unemployment spell. We denote individuals who find a job in the same industry as individuals with risk type 1, while those who find a job in a different industry as individuals with risk type 2. Summary statistics for this subsample are given in Table 2. We note a marked difference in the unemployment durations of these two groups of individuals. Figure 1 provides plots of the number of individuals who exited in each time period, shown separately for workers who find a job in the same industry and those who do not (risk type 2). Individuals who find a job in the same industry exit unemployment much faster in the first few weeks after they lose their job but conditional on not having found employment by week 8 their exit pattern resembles that of the other individuals. This is empirically interesting as it points towards the importance of industry specific human capital. We would expect variation among the industry specific human capital to vary by industry but also by individual reflecting their level of experience and motivation. It is thus to be expected that some individuals have accumulated a significant degree of human capital which makes them attractive to other employers in the same industry. Switching industries usually entails learning new skills and the incentive to do so may be affected by the duration and generosity of the unemployment insurance benefits. We would thus expect substantial behavioral differences between workers facing these two competing risks.

5.2. **Single Risk Duration Model with Flexible Heterogeneity**

Estimation results of the semiparametric duration model with a flexible form of unobserved heterogeneity under GIG mixing (Assumptions A1–A3, A6, and A7a) are presented in Table 3. In the Online Appendix we also present estimation results for two benchmark parametric models. Estimation results of a model with parametric gamma heterogeneity (Han and Hausman, 1990; Meyer 1990), as specified in Assumptions A1–A4 and A5b, are given in Table OA.1. In Table OA.2 we present estimation results for another benchmark model with parametric GIG heterogeneity (Assumptions A1–A4 and A5a).

In addition to the above mentioned censoring at $T = 24$ weeks, we also include the benchmark cases where we censor at $T = 6$ and $T = 13$. In the GIG mixture model, all of our variables (state unemployment rate, race, age, and replacement rate) are estimated to have a negative significant impact on the hazard rate of exiting unemployment. Note however, that when comparing the estimates of the β coefficients, the scaling changes depending on the variance of the estimated heterogeneity distribution. Thus, the ratios of the coefficients should be compared, as opposed to their absolute values. This makes the interpretation of the coefficients less transparent. We note however that the coefficient estimates obtained from the flexible model are substantively different than those obtained from the parametric model. As discussed above, the parametric restriction on the heterogeneity distribution can lead to inconsistent estimates if the true mixing distribution does not exactly correspond to the parametric specification.

We would expect to obtain similar results irrespective of the truncation point and thus the coefficients obtained for the models truncated at 6, 13, and 24 weeks to be very similar. While the coefficient ratios are not constant they tend to be relatively similar. The one exception comes from the ratios involving the replacement rate, in particular for the model with censoring at 24 periods. This is similar to the results in Hausman and Woutersen (2012) and might be explained by behavioral changes as individuals approach the date of UI exhaustion.

The estimated distribution of the unobserved individual heterogeneity is presented in Figure 2. The estimated distributions can only be very roughly approximated by the gamma distribution. While in all three cases the mode is negative, as we increase the number of periods used in the estimation the distribution acquires a more pronounced left tail. This indicates that as we observe individuals over a longer period of time the model captures to a larger extent the part of unobserved heterogeneity which prevents workers from finding employment and thus becomes indicative of the propensity for long term unemployment.

The survival function estimates along with 95% confidence bands are presented in Figure 3, featuring the anticipated downward sloping shape. The smoothing parameter α of the Dirichlet Process (DP) Mixture model introduced in Assumption A6 controls the extent to which the DP draws mixture distributions that are more or less "similar" to the baseline parametric distribution G_0. In the limiting case of $\alpha \rightarrow \infty$ the mixture distribution becomes equivalent to G_0, while in the other extreme $\alpha \rightarrow 0$ the mixture distribution limits to a convolution of density kernels centered at each data point without any influence of the DP prior. The posterior distribution estimates of α are plotted in Figure OA.1 in the Online Appendix. The distributions are concentrated around a well-defined mode with a value of less than 1 indicating a strong influence of data relative to the baseline prior distribution thereby providing a high degree of support in favor of our nonparametric approach.

5.3. **Competing Risk Model with Flexible Heterogeneity**

We now present the results of our newly proposed competing risk model with a flexible form of unobserved heterogeneity using GIG mixing and correlated risks (Assumptions B1-B3, B6, and B7a). Note that in our example risk 1 corresponds to the event that a worker find a job in the same industry, while risk 2 corresponds to the event that she finds a new job in an industry that is different than the industry of her previous employer. We present the estimated coefficients in Table 4. For all three censoring times ($T = 6, 13, 24$), the partial effects of race and age are not statistically significant, with a few isolated exceptions. This could be due to smaller sample size available for the competing risk case as opposed to the single-risk case, with the former consisting of less than 10% of observations of the latter. For all three censoring times, the relative influence of the replacement rate is declining from $T = 13$ to $T = 24$ indicating the impact of benefit exhaustion.

The estimated density[9] of unobserved heterogeneity V_{ik} is shown in Figure 4 for the GIG mixture model for both risk types $k = 1, 2$, each centered at the time average of the risk-specific latent common time effect ζ_{kt} to reflect the overall influence of the unobserved heterogeneity component $\zeta_{kt} + V_{ik}$. The differences between the density of unobserved individual heterogeneity further highlight the importance of distinguishing between the different risk types in the competing risk model environment as compared to the single-risk duration case. In particular the two distributions of are distinct and well-separated

[9] In Table 4 we report the estimated GIG mixture model coefficients but as these enter all mixing kernel moments their interpretation is not immediate. Hence it appears more informative to examine the resulting mixture density estimate.

indicating that conditional on observed covariates there is a significant degree of sorting between workers finding jobs in the same or in a different industry. While the mode of both distributions is negative, workers who find a job in the same industry possess latent attributes that make them more desirable than workers who are not. This may indicate the presence of unobserved industry specific human capital for these workers which make them more attractive to employers in the same industry.

The survival function estimates along with 95% confidence bands are presented in Figure 5 for both types of risks. The differences in the shapes for the first few weeks are striking and also indicative of the differences in exit rates between workers finding a job in the same industry compared who workers that find a job in a different industry. The estimated survival function for workers who find a job in the same industry is convex while that for workers who do not is concave, indicating a much slower overall re-integration into the labor market. This appears to confer a long term advantage with the overall probability of being unemployed being substantially higher for workers with no apparent industry specific human capital and which have no particular advantage in terms of finding a job in the same industry as their pre-unemployment firm.

Figure 6 shows the estimated correlation structure of the latent time variables ζ_{kt} common to all individuals, defined in Assumption B4, for the GIG mixture model in terms of the estimated densities for the variances σ_1^2, σ_2^2, and the correlation coefficient ρ between ζ_{1t} and ζ_{2t} for the two different risk types. Interestingly, most probability mass for the density of ρ is negative for $T = 6$, around zero for $T = 13$ and positive for $T = 24$. This may correspond to a negative correlation of common shocks for jobs in the same and in different industries for the first several weeks of unemployment with a subsequent correlation reversal in later time periods, but may also be influenced by other factors, as estimates of other model elements also change. Nonetheless, such correlation pattern would explain the exit counts shown in Figure 1 for the two risks, with high ratios of jobs in the same industry to different industries for the first few weeks abetting to parity around week six.

The posterior distribution estimates of the smoothing parameter α of the Dirichlet Process Mixture model (Assumption B6) are plotted in Figure OA.2 in the Online Appendix.

The distributions are concentrated around a well-defined mode of value less than five, indicating a strong influence of data relative to the baseline prior distribution.

It is informative to contrast the estimates from our preferred model with those obtained under different modeling assumptions on the unobserved heterogeneity which are detailed in the Online Appendix. Table OA.3 presents the estimated coefficients from a model that ignores the presence of individual heterogeneity, Table OA.4 corresponds to a model which assumes parametric gamma distributed heterogeneity, Table OA.5 estimates a competing risk model with parametric GIG heterogeneity, and Table OA.6 presents the estimation results from a flexible model which estimates the unobserved semi-parametrically using an infinite mixture of GIG distributions but also further imposes the assumption of independence between the different risks. In Table OA.7 we present estimation results from our single risk GIG mixture model but applied to the subsample of observations which records the outcomes for the two competing risks.

Given the large number of parameters to be considered it is helpful to compare these different models in a graphical setting. In order to facilitate the comparison between the model which pools the two risks and the models which do not we can combine the two risks into a common survival function, as discussed in the Online Appendix. Thus, in Figure 7 we compare the estimated survival function of our CR GIG mixture model (Assumptions 1–3, B6 and B7a, labeled as "CR full") with its estimates in three restricted model versions: 1) the parametric GIG case (labeled as "CR param"), 2) the independent risks case where we estimate a single-risk model separately for each risk type of data and then merge their survival functions ex-post (labeled as "CR indep"); 3) the case without individual unobserved heterogeneity under the restriction $V_{ik} = 0$ (labeled as "CR no ihet"); and 4) the single-risk case where we do not distinguish between risk types in the competing risk data (labeled as "SR full").

Two features are particularly significant. First, we notice that if we enforce the assumption of independence of the two risk types, the resulting common survival function is severely downward biased. The magnitude of the bias dominates the other modeling choices which we make on the specification of unobserved heterogeneity. This could be due to the distributional effects of the risk correlations (Figure 7) that are absent in the independent risk model. Second, we plot the confidence bounds for our proposed model which allows for a flexible specification of the unobserved heterogeneity but also for correlated competing risks. We notice that all other more restrictive specifications are downward biased and the differences become statistically significant as the number of time periods increases.

5.4. **Counterfactual Policy Evaluation**

One of the advantages of our model consists in the explicit estimation of the unobserved heterogeneity components which enables us to evaluate the effectiveness of counterfactual policy experiments taking into account the distributional effects of individual heterogeneity. As discussed above, one of the main policy questions currently faced by economists is the extent to which the generosity of unemployment insurance benefits impacts the workers' incentives to find employment once their lose their job. On the one hand, more generous benefits are expensive to provide given the ongoing debt crisis and may actually prove detrimental in the long run as they may erode workers' incentives to find a job quickly. Thus they would ultimately contribute to increasing long run unemployment. On the other hand, low levels of unemployment insurance benefits can make unemployment very difficult for many low income families. Poverty can also have a negative effect on their ability to find employment since job search is costly and in the absence of unemployment insurance benefits many workers may find themselves unable to support their families while also searching for an adequate job. As a result workers may end up underemployed or leave the labor market altogether. The relative magnitude of the impact of incentives over poverty is an empirical question and a counterfactual analysis using model estimates can provide some evidence in this debate.

In the context of our model we can consider changing the replacement rate in order to investigate its impact on the probability of exit from unemployment as captured by the survival function. Note that we assume that this policy change does not impact the distribution of heterogeneity. We can perform this policy counterfactual using both the single risk and the competing risk model. For clarity, we combine the two risk types in the CR model into a common survival function as described in the Online Appendix. The counterfactual experiment consists in increasing and decreasing the replacement rate by 10%. We present counterfactual results from our preferred specification which flexibly models the unobserved heterogeneity as an infinite GIG mixture. The estimated and counterfactual survival curves under the two scenarios are presented in Figure 8 (SR) and Figure 9 (CR).

Both Figures show that the survival function moves in the anticipated direction: for a replacement rate decrease the probability of staying unemployed is lower, and for replacement rate increase the probability of continued unemployment is higher. However, the changes are relatively small. For example, for $T = 24$ in the final period the survival function changes by -1.7% and 1.6% for the CR data, respectively. This suggests that while the estimated impact of a change in unemployment benefit generosity has the sign predicted by economic theory, the magnitude of the impact on the probability of unemployment exit is inelastic. Policy makers may thus wish to consider the extent to which cutting unemployment benefits may ultimately influence an unemployed worker's welfare.

In the Online Appendix we further report on the results of two extensions of the counterfactual experiment that we briefly summarize here. First, we split individuals into two different subsamples based on their unobserved heterogeneity component – below and above the median. The results indicate that individuals with higher unobserved heterogeneity react more to replacement rate changes and have better chances exiting unemployment faster than their counterparts with lower unobserved heterogeneity. Second, we specify a time-varying replacement rate counterfactual change under two scenarios: one with sharp initial change and one with sharp late change. On average the survival function changes more under the former scenario, albeit inelastically in absolute terms.

6. Conclusion

We introduced a new flexible model specification for the competing risk model with piecewise linear baseline hazard, time-varying regressors, risk-specific unobserved individual heterogeneity distributed as an infinite mixture of density kernels, and a common correlated latent effect. Unobserved individual heterogeneity is assumed to be distributed according a Bayesian Dirichlet Process mixture model with a data-driven stochastic number of mixture components estimated along with other model parameters. We derive a tractable likelihood for Generalized Inverse Gaussian (GIG) mixing based on scaled GIG Laplace transforms and their higher-order derivatives. We find that mixing under a special case of the GIG, the gamma kernel, leads to degenerate outcomes in nonparametric mixtures motivating the use of the more flexible GIG. We apply our approach to analyzing unemployment duration with exits to jobs in the same industry and to a different industry among unemployment insurance recipients on nationally representative individual-level survey data from the U.S. Department of Labor. We also conduct a counterfactual policy experiment that changes the replacement rate and find that the extent to which cuts in unemployment benefits incentivize unemployed workers is relatively very small.

References

Abbring, J., and G. J. van den Berg (2003): "The identifiability of the mixed proportional hazards competing risks model," *Journal of the Royal Statistical Society Series B*, 65, 701–710.

Alba-Ramírez, A., J. M. Arranz, and F. Muñoz Bullón (2007): "Exits from unemployment: Recall or new job," *Labor Economics*, 14, 788–810.

Baker, M., and A. Melino (2000): "Duration dependence and nonparametric heterogeneity: A Monte Carlo study," *Journal of Econometrics*, 96(2), 357–393.

Berrington, A., and I. Diamond (2000): "Marriage or Cohabitation: A Competing Risks Analysis of First-Partnership Formation among the 1958 British Birth Cohort," *Journal of the Royal Statistical Society, Series A*, 163(2), 127–151.

Beyersmann, J., M. Schumacher, and A. Allignol (2012): *Competing Risks and Multistate Models in R*. Springer, New York.

Bierens, H. J., and J. R. Carvalho (2007): "Semi-Nonparametric Competing Risks Analysis Of Recidivism," *Journal of Applied Econometrics*, 22, 971–993.

Bijwaard, G. E. (2012): "Multistate event history analysis with frailty," forthcoming, *Demographic Research*.

Blackwell, D., and J. B. MacQueen (1973): "Ferguson Distributions via Pólya Urn Schemes," *Annals of Statistics*, 1, 353–355.

Booth, A. L., and S. E. Satchell (1995): "The Hazards of doing a PhD: An Analysis of Completion and Withdrawal Rates of British PhD Students in the 1980s," *Journal of the Royal Statistical Society, Series A*, 158(2), 297–318.

Burda, M., M. Harding, and J. A. Hausman (2013): "Online Appendix to "A Bayesian Semiparametric Competing Risk Model with Unobserved Heterogeneity"," *Journal of Applied Econometrics*.

Bush, C. A., and S. N. MacEachem (1999): "A Semiparametric Bayesian Model for Randomised Block Designs," *Biometrika*, 83, 275–285.

Butler, J. S., K. H. Anderson, and R. V. Burkhauser (1989): "Work and Health after Retirement: A Competing Risks Model with Semiparametric Unobserved Heterogeneity," *The Review of Economics and Statistics*, 71, 46–53.

Campolieti, M. (2001): "Bayesian semiparametric estimation of discrete duration models: an application of the dirichlet process prior," *Journal of Applied Econometrics*, 16(1), 1–22.

Canals-Cerdá, J., and S. Gurmu (2007): "Semiparametric competing risks analysis," *Econometrics Journal*, 10, 193–215.

Chakrabarty, B., H. Tyurin, Z. Han, and X. Zheng (2006): "A Competing Risk Analysis of Executions and Cancellations in a Limit Order Market," Discussion paper, CAEPR Working Paper No. 2006-015.

Chib, S., and B. Hamilton (2002): "Semiparametric bayes analysis of longitudinal data treatment models," *Journal of Econometrics*, 110, 67–89.

Cox, D. R. (1962): *Renewal Theory*. Methuen, London.

De Blasi, P., and N. Hjort (2007): "Bayesian survival analysis in proportional hazard models with logistic relative risks," *Scandinavian Journal of Statistics*, 24(2), 229–257.

——— (2009): "The Bernstein-von Mises theorem in semiparametric competing risks models," *Journal of Statistical Planning and Inference*, 139(2), 2316–2328.

Deng, Y., J. M. Quigley, and R. Van Order (2000): "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options," *Econometrica*, 68(2), 275307.

Dolton, P., and W. van der Klaauw (1999): "The Turnover Of Teachers: A Competing Risks Explanation," *The Review of Economics and Statistics*, 81, 543–552.

Elbers, C., and G. Ridder (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *The Review of Economic Studies*, 49(3), 403–409.

Escobar, M. D., and M. West (1995): "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90(430), 577–588.

Esteve-Pérez, S., A. Sanchis-Llopis, and J. A. Sanchis-Llopis (2010): "A competing risks analysis of firms exit," *Empirical Economics*, 38, 281–304.

Fallick, B. C. (1991): "Unemployment Insurance and the Rate of Re-Employment of Displaced Workers," *The Review of Economics and Statistics*, 73, 228–235.

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

Florens, J.-P., M. Mouchart, and J.-M. Rolin (1999): "Semi- and Non-parametric Bayesian Analysis of Duration Models with Dirichlet Priors: A Survey," *International Statistical Review*, 67, 187–210.

Gasbarra, D., and S. R. Karia (2000): "Analysis of Competing Risks by Using Bayesian Smoothing," *Scandinavian Journal of Statistics*, 27(4), 605–617.

Han, A., and J. A. Hausman (1990): "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics*, 5(1), 1–28.

Hausman, J. A., and T. Woutersen (2012): "Estimating a Semi-Parametric Duration Model without Specifying Heterogeneity," forthcoming, *Journal of Econometrics*.

Heckman, J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2), 271–320.

Heckman, J. J., and B. E. Honoré (1989): "The Identifiability of the Competing Risks Model," *Biometrika*, 76(2), 325–330.

Hirano, K. (2002): "Semiparametric bayesian inference in autoregressive panel data models," *Econometrica*, 70, 781–799.

Hjort, N. L., C. Holmes, P. Müller, and S. G. Walker (2010): *Bayesian Nonparametrics*. Cambridge University Press, New York.

Honoré, B. E. (1990): "Simple Estimation of a Duration Model with Unobserved Heterogeneity," *Econometrica*, 58(2), 453–473.

Honoré, B. E., and A. Lleras-Muney (2006): "Bounds in Competing Risks Models and the War on Cancer," *Econometrica*, 74(6), 1675–1698.

Horowitz, J. L. (1999): "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity," *Econometrica*, 67(5), 1001–1028.

Hougaard, P. (2000): *Analysis of multivariate survival data*. Springer, New York.

Jakobsen, V., and M. Rosholm (2003): "Dropping out of school? A competing risk Analysis of Young Immigrants' Progress in the Educational System," Discussion paper no. 918., IZA.

Katz, L. F., and B. D. Meyer (1990): "Unemployment Insurance, Recall Expectations, and Unemployment Outcomes," *Quarterly Journal of Economics*, 105, 973–1002.

Lancaster, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47(4), 939–56.

Lancaster, T. (1990): *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge.

Lee, S., and A. Lewbel (2013): "Nonparametric Identification of Accelerated Failure Time Competing Risks Models," forthcoming, *Econometric Theory*.

Li, M. (2007): "Bayesian Proportional Hazard Analysis of the Timing of High School Dropout Decisions," *Econometric Reviews*, 26(5), 529–556.

Lillard, L. A. (1993): "Simultaneous equations for hazards : Marriage duration and fertility timing," *Journal of Econometrics*, 56(1-2), 189–217.

McCall, B. P. (1996): "Unemployment Insurance Rules, Joblessness, and Part-Time Work," *Econometrica*, 64(3), 647–682.

Meyer, B. D. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58(4), 757–82.

Moffitt, R., and W. Nicholson (1982): "The Effect of Unemployment Insurance on Unemployment: The Case of Federal Supplemental Benefits," *The Review of Economics and Statistics*, 64, 1–11.

Mortensen, D. (1977): "Unemployment Insurance and Job Search Decisions," *Industrial and Labor Relations Review*, 30, 505–517.

Mulligan, C. (2012): *The Redistribution Recession: How Labor Market Distortions Contracted the Economy*. Oxford University Press.

Neal, R. (2000): "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9(2), 249–265.

Neal, R. M. (2011): "MCMC using Hamiltonian dynamics," in *Handbook of Markov Chain Monte Carlo*, ed. by S. Brooks, A. Gelman, G. Jones, , and X.-L. Meng, pp. 113–162. Chapman & Hall / CRC Press.

Needels, K., W. Corson, and W. Nicholson (2001): "Left Out of the Boom Economy: UI Recipients in the Late 1990s," U.S. Department of Labor Report.

Paserman, M. D. (2004): "Bayesian Inference for Duration Data with Unobserved and Unknown Heterogeneity: Monte Carlo Evidence and an Application," IZA discussion papers, Institute for the Study of Labor (IZA).

Pintilie, M. (2006): *Competing risks: a practical perspective*. John Wiley & Sons, Hoboken, NJ.

Polpo, A., and D. Sinha (2011): "Correction in Bayesian nonparametric estimation in a series system or a competing-risk model," *Statistics & Probability Letters*, 81(12), 1756–1759.

Porta, N., G. Gómez, and M. Calle (2008): "The role of survival functions in competing risks," Technical report DR 2008/06, Department of Statistics and Operations Research, Universitat Politècnica de Catalunya.

Prentice, R., and L. Gloeckler (1978): "Regression Analysis of Grouped Survival Data with an Application to Breast Cancer Data," *Biometrics*, 34, 57–67.

Ruggiero, M. (1994): "Bayesian semiparametric estimation of proportional hazard models," *Journal of Econometrics*, 62, 277–300.

Salinas-Torres, V., C. Pereira, and R. Tiwari (2002): "Bayesian nonparametric estimation in a series system or a competing-risk model," *Journal of Nonparametric Statistics*, 14(4), 449–458.

Sueyoshi, G. T. (1992): "Semiparametric proportional hazards estimation of competing risks models with time-varying covariates," *Journal of Econometrics*, 51(1-2), 25–58.

Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK.

Tsiatis, A. (1975): "A nonidentifiability aspect of the problem of competing risks," *Proceedings of the National Academy of Sciences*, 72, 20–2.

Tysse, T. I., and K. Vaage (1999): "Unemployment of Older Norwegian Workers: A Competing Risk Analysis," Research report 99/14, Statistics Norway, Research Department.

Van den Berg, G. J. (2001): "Duration models: specification, identification and multiple durations," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5, chap. 55, pp. 3381–3460. Elsevier.

van den Berg, G. J., A. G. C. van Lomwel, and J. C. van Ours (2008): "Nonparametric estimation of a dependent competing risks model for unemployment durations," *Empirical Economics*, 34, 477–491.

Vaupel, J., K. Manton, and E. Stallard (1979): "The impact of heterogeneity in individual frailty on the dynamics of mortality," *Demography*, 16, 439–54.

West, M., P. Müller, and M. D. Escobar (1994): "Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation," in *Aspects of Uncertainty*, ed. by P. R. Freeman, and A. F. M. Smith, pp. 363–386. Wiley, New York.

7. Appendix: Tables and Figures

TABLE 1. Overview of Assumptions

Heterogeneity type	*Single Risk*	*Competing Risks*
No heterogeneity	A1–A3	B1–B3
Parametric GIG	A1–A4, A5a	B1–B4, B5a
Parametric gamma	A1–A4, A5b	B1–B4, B5b
Flexible GIG mixture	A1–A3, A6, A7a	B1–B3, B6, B7a
Flexible gamma mixture	A1–A3, A6, A7b	B1–B3, B6, B7b

TABLE 2. Summary Statistics

	Duration Data				Competing Risk Data			
Variable	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Duration	23.245	16.334	1	63	27.958	28.410	1	140
Race	0.117	0.321	0	1	0.107	0.310	0	1
Age	0.177	0.382	0	1	0.181	0.385	0	1
Replacement Rate	0.6558	0.3082	0.016	1.8434	0.660	0.324	0.015	2.173
State unemp rate:								
period 1	4.686	1.087	2	6.9	4.579	1.125	2	7.5
period 2	4.672	1.083	2	6.9	4.568	1.120	2	7.5
period 3	4.660	1.079	2	6.9	4.562	1.107	2	7.5
period 4	4.645	1.074	2	6.9	4.553	1.103	2	7.8
period 5	4.630	1.069	2	6.9	4.536	1.092	2	8.1
period 6	4.621	1.064	2	6.9	4.533	1.083	2	8.1
period 7	4.616	1.066	2	6.9	4.538	1.080	2	8.1
period 8	4.598	1.064	2	6.9	4.527	1.071	2	7.4
period 9	4.570	1.061	2	6.9	4.518	1.070	2	7.2
period 10	4.538	1.061	2	6.9	4.486	1.070	2	7.2
period 11	4.531	1.063	2	6.9	4.481	1.072	2	7.2
period 12	4.509	1.067	2	6.9	4.458	1.082	2	6.9
period 13	4.483	1.075	2	6.9	4.438	1.091	2	6.9
period 14	4.462	1.080	2	6.9	4.412	1.098	2	6.9
period 15	4.460	1.075	2	6.9	4.404	1.102	2	6.9
period 16	4.449	1.073	2	6.9	4.390	1.097	2	6.9
period 17	4.439	1.067	2	6.9	4.376	1.088	2	7.8
period 18	4.440	1.055	2	6.9	4.368	1.078	2	7.8
period 19	4.431	1.054	2	6.9	4.357	1.073	2	7.8
period 20	4.420	1.045	2	6.9	4.342	1.066	2	7.8
period 21	4.423	1.033	2	6.9	4.338	1.052	2	7.5
period 22	4.431	1.029	2	6.8	4.335	1.038	2	7.4
period 23	4.436	1.024	2	6.7	4.345	1.037	2	7.4
period 24	4.441	1.015	2	6.7	4.353	1.037	2	7.4
Observations	15,358				1,243			

Table 3. New Semiparametric Duration Model, GIG Mixture

		6 periods		13 periods		24 periods	
		Mean	s.e.	Mean	s.e.	Mean	s.e.
κ		-0.963	0.039	-1.226	0.034	-1.659	0.101
φ		2.284	0.110	3.022	0.097	4.240	0.285
Urate		-0.184	0.021	-0.214	0.014	-0.327	0.034
Race		-0.055	0.069	-0.126	0.042	-0.145	0.050
Age		-0.194	0.057	-0.178	0.035	-0.187	0.039
Rrate		-0.924	0.076	-0.473	0.051	-0.147	0.066
t	1	-2.142	0.116	-2.264	0.090	-1.971	0.181
	2	-1.716	0.109	-1.843	0.088	-1.573	0.175
	3	-2.026	0.112	-2.157	0.096	-1.878	0.182
	4	-1.733	0.112	-1.865	0.085	-1.578	0.182
	5	-2.070	0.112	-2.198	0.099	-1.928	0.185
	6	-1.833	0.102	-1.829	0.084	-1.547	0.184
	7			-2.336	0.100	-2.061	0.205
	8			-2.026	0.087	-1.743	0.189
	9			-2.347	0.097	-2.099	0.181
	10			-2.130	0.087	-1.871	0.189
	11			-2.347	0.087	-2.086	0.194
	12			-2.120	0.095	-1.856	0.196
	13			-2.277	0.074	-1.904	0.195
	14					-1.840	0.191
	15					-1.775	0.191
	16					-1.740	0.195
	17					-1.981	0.206
	18					-1.641	0.191
	19					-1.848	0.194
	20					-1.674	0.191
	21					-1.832	0.198
	22					-1.738	0.199
	23					-1.913	0.211
	24					-1.006	0.179

$N = 15,491$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

Table 4. New Semiparametric Competing Risk Model, GIG Mixture

		6 periods				13 periods				24 periods			
		Risk 1		Risk 2		Risk 1		Risk 2		Risk 1		Risk 2	
		Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
κ		-1.535	0.046	-1.516	0.046	-1.482	0.057	-1.554	0.062	-1.407	0.044	-1.492	0.047
φ		3.890	0.129	3.838	0.304	3.741	0.162	3.944	0.176	3.532	0.126	3.771	0.134
Urate		-0.111	0.055	-0.103	0.083	-0.107	0.048	-0.117	0.063	-0.216	0.040	-0.095	0.056
Race		0.054	0.225	0.033	0.424	-0.044	0.183	-0.377	0.343	-0.046	0.157	0.312	0.256
Age		-0.008	0.176	-0.562	0.357	-0.016	0.154	-0.700	0.290	-0.036	0.130	-0.461	0.212
Rrate		-1.062	0.227	-0.465	0.390	-0.903	0.181	-0.411	0.288	-0.385	0.156	-0.331	0.226
t	1	-2.553	0.314	-4.864	0.516	-2.680	0.276	-4.741	0.503	-2.496	0.256	-4.947	0.511
	2	-1.993	0.296	-4.625	0.491	-2.113	0.255	-4.514	0.504	-1.933	0.234	-4.715	0.475
	3	-2.324	0.311	-4.937	0.565	-2.445	0.272	-4.830	0.558	-2.263	0.253	-5.027	0.554
	4	-2.022	0.302	-3.633	0.400	-2.138	0.263	-3.506	0.388	-1.962	0.243	-3.728	0.372
	5	-2.470	0.325	-3.638	0.413	-2.580	0..28	-3.516	0.391	-2.406	0.273	-3.729	0.376
	6	-3.854	0.290	-3.711	0.212	-3.044	0.322	-3.327	0.386	-2.872	0.308	-3.543	0.367
	7					-2.947	0.321	-4.361	0.518	-2.764	0.304	-4.599	0.508
	8					-2.493	0.295	-3.068	0.372	-2.318	0.277	-3.285	0.355
	9					-3.750	0.422	-3.984	0.479	-3.571	0.417	-4.203	0.460
	10					-2.620	0.305	-3.950	0.471	-2.449	0.293	-4.149	0.451
	11					-3.232	0.362	-3.925	0.480	-3.078	0.355	-4.129	0.449
	12					-2.791	0.326	-3.866	0.445	-2.625	0.305	-4.093	0.454
	13					-3.799	0.341	-3.829	0.409	-2.703	0.317	-4.057	0.453
	14									-2.738	0.329	-3.897	0.437
	15									-2.511	0.309	-3.634	0.406
	16									-3.074	0.376	-3.927	0.452
	17									-2.127	0.284	-3.872	0.453
	18									-2.564	0.335	-4.589	0.569
	19									-2.513	0.322	-3.285	0.388
	20									-2.702	0.347	-3.884	0.457
	21									-1.850	0.278	-3.569	0.424
	22									-3.795	0.540	-4.719	0.652
	23									-2.133	0.307	-3.780	0.464
	24									-3.640	0.322	-3.770	0.106

$N = 1,243$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

FIGURE 1. Empirical Exit Count for Competing Risk Data

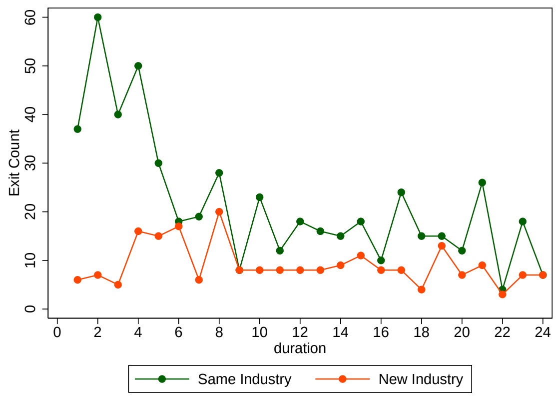


FIGURE 2. Density of individual heterogeneity component v_i, GIG mixture

$T = 6$ $T = 13$ $T = 24$

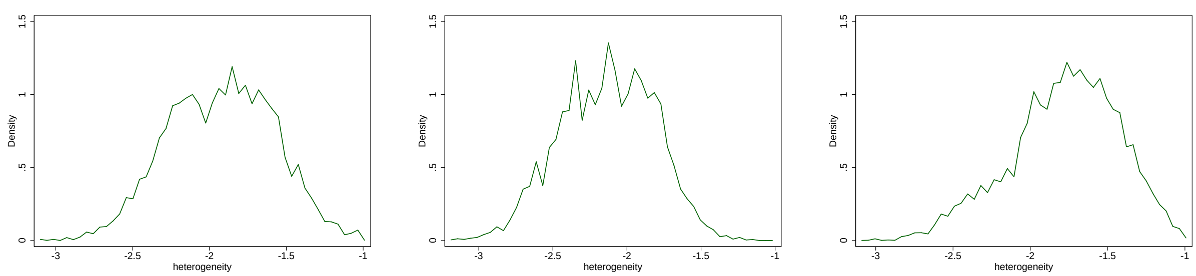


FIGURE 3. Survival function, GIG mixture

$T = 6$ $T = 13$ $T = 24$

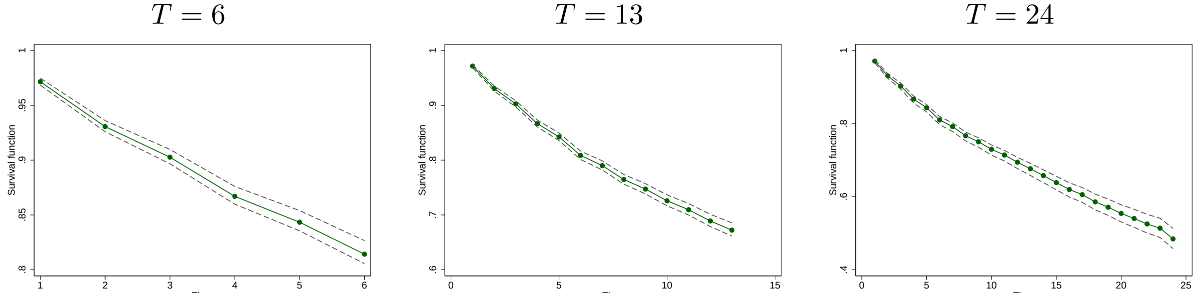


FIGURE 4. Heterogeneity density, GIG mixture

$T = 6$ $T = 13$ $T = 24$

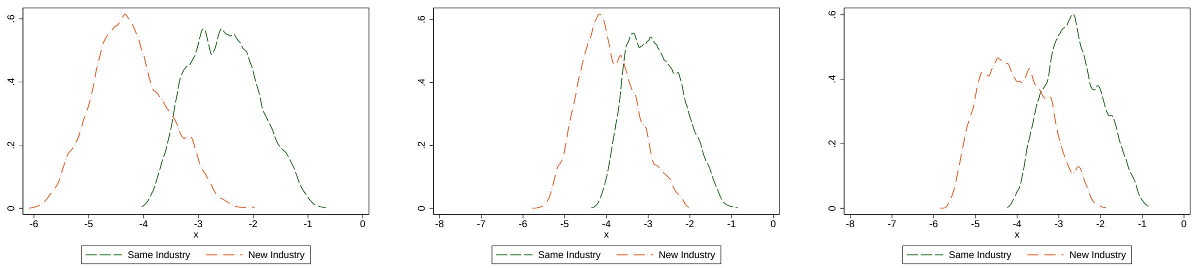


FIGURE 5. Survival function, GIG mixture

$T = 6$ $T = 13$ $T = 24$

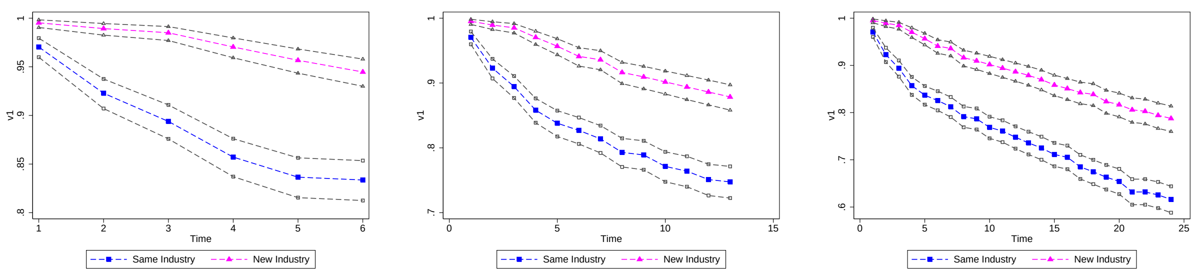


FIGURE 6. Correlation structure of ζ_t: density of σ_1^2, σ_1^2, and ρ, GIG mixture

$T = 6$ $T = 13$ $T = 24$

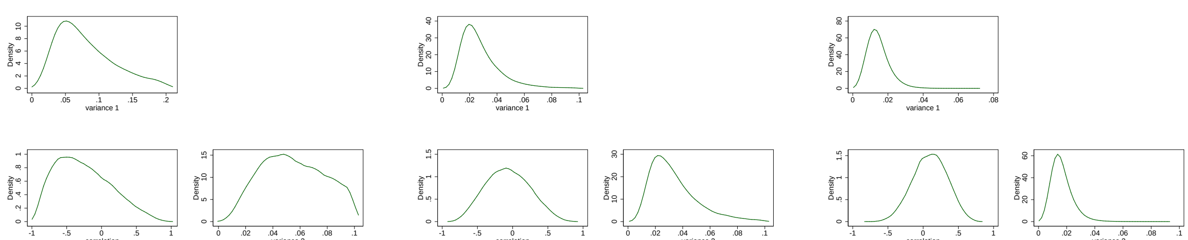


FIGURE 7. Model Comparison in Terms of Survival Functions

$T = 6$ $T = 13$ $T = 24$

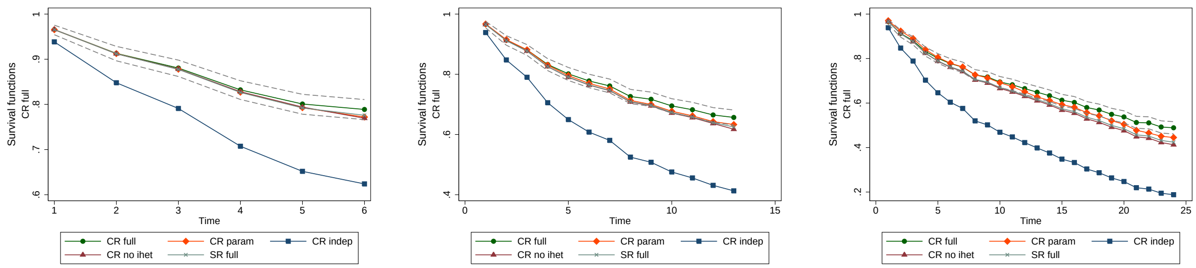


FIGURE 8. Counterfactual Experiment for the Single Risk GIG Mixture

$T = 6$ $T = 13$ $T = 24$

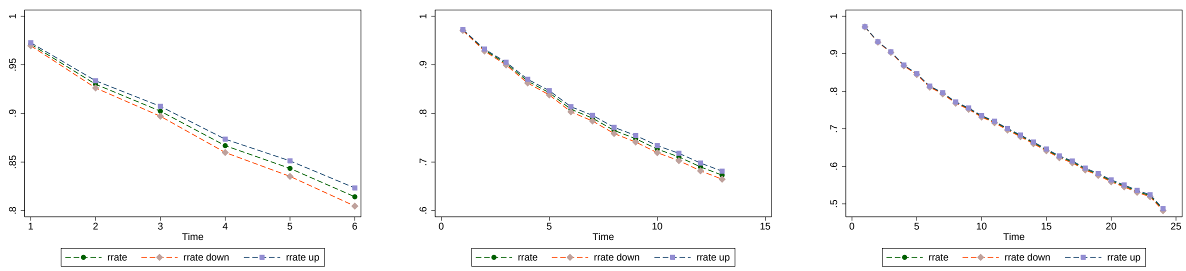
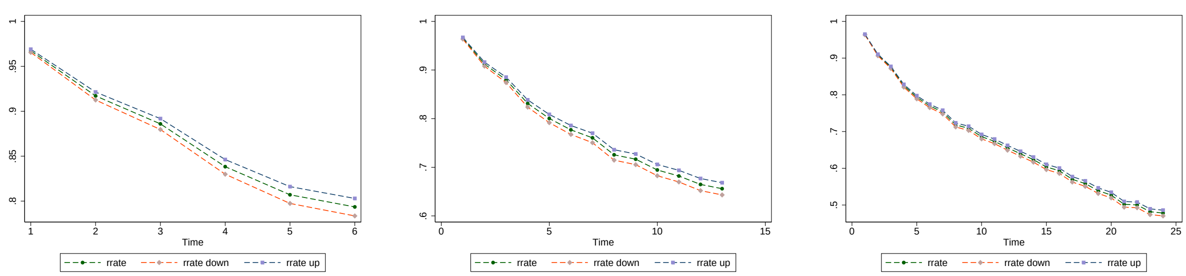


FIGURE 9. Counterfactual Experiment for the Competing Risks Model using a GIG Mixture and Combining the Risks.

$T = 6$ $T = 13$ $T = 24$



Online Appendix

"A Bayesian Semiparametric Competing Risk Model with Unobserved Heterogeneity"
by Martin Burda, Matthew Harding, and Jerry Hausman
September 14, 2013

1. Details on MCMC Posterior Sampling

For our model implementation we utilize the Gibbs sampling scheme which belongs to the class of Markov Chain Monte Carlo (MCMC) simulation methods. An attractive feature of MCMC techniques is that samples of random draws can be generated from the joint posterior densities of parameters of interest indirectly, without the need to specify the exact analytical form of the joint densities. The Gibbs sampler uses an iterative procedure to create Markov chains by simulating from conditional densities instead which are analytically tractable. The sets of draws obtained in this way can be effectively considered as samples from the joint posterior densities.

1.1. Gibbs Blocks

For the single risk case, let $\delta_{0t} = \log(\mu_{0t})$, $\delta_0 = (\delta_{01}, \ldots, \delta_{0T})$, and $V = \{V_i\}_{i=1}^N$. The model parameters consist of β, δ_0, V, hyperparameters either of the GIG mixture κ and φ or the gamma mixture γ (denote the hyperparameters generically by ψ), and the DP concentration parameter α. Under Assumptions A1–A7, the joint posterior density can be decomposed into the following Gibbs blocks:

(1) $\beta, \delta_0|V, \psi, \alpha$
(2) $V|\psi, \alpha, \beta, \delta_0$
(3) $\psi|\alpha, \beta, \delta_0, V$
(4) $\alpha|\beta, \delta_0, V, \psi$

In the competing risk case, all the above parameters are risk-specific. Moreover, due to Assumption B4, there are additional parameters ζ_{kt}, ρ, σ_1, and σ_2. Hence, let $\delta_{0kt} = \log(\lambda_{0kt}) + \zeta_{kt}$ where $\lambda_{0kt} = \mu_{0kt}$ since time intervals have unit length. Let further $\beta = \{\beta_k\}_{k=1}^K$, $\delta_{0k} = (\delta_{0k1}, \ldots, \delta_{0kT})$, $\delta_0 = (\delta_{01}, \delta_{02})$, $\sigma = \{\sigma_k\}_{k=1}^K$, $V_i = \{V_{ki}\}_{k=1}^K$, $V = \{V_i\}_{i=1}^N$, $\psi = \{\psi_k\}_{k=1}^K$, and $\alpha = \{\alpha_k\}_{k=1}^K$. In our application, $K = 2$. The Gibbs blocks are now as follows:

(1) $\beta, \delta_0|\sigma, \rho, V, \psi, \alpha$
(2) $\sigma, \rho|V, \psi, \alpha, \beta, \delta_0$
(3) $V|\psi, \alpha, \beta, \delta_0, \sigma, \rho$
(4) $\psi|\alpha, \beta, \mu_0, \sigma, \rho, V$
(5) $\alpha|\beta, \mu_0, \sigma, \rho, V, \psi$

This first Gibbs block is sampled using standard Hamiltonian Monte Carlo (HMC) with SR posterior (2.8) and (2.9), and CR posterior (3.3) and (3.4). For a detailed description of the HMC procedure, see e.g. Neal (2011), pp. 122–125. In the CR case, the covariance matrix of the second block is endowed with a proper Inverse Wishart prior with maximum dispersion, $IW(K+1, I_K)$ where I_K is the identity matrix of dimension K. Given the draws δ_0 and the Normal correlation structure in Assumption B4, the posterior can be found e.g. in Train (2003) on p. 301 and the sampling procedure on p. 302. Sampling individual heterogeneity V, hyperparameters ψ, and α is detailed in the next section.

1.2. **Individual Heterogeneity**

The distribution of the unobserved heterogeneity component v_i is modeled as a mixture with countably infinite number of mixture components. In the Bayesian framework employing a prior distribution for mixing proportions, such as the Dirichlet Process that we adopt here, leads to a relatively few of the mixture components dominating in the posterior. Using a countably infinite mixture bypasses the need to determine the "correct" number of components in a finite mixture model.

DP mixture modeling is described in detail e.g. in Hjort et al (2010). In our implementation, we use Algorithm 2 of Neal (2000). Here we provide the essence of the procedure. The prior structure of the model for v_i is specified by our Assumptions A6 (SR) and B6 (CR). It is based on two levels of hierarchy, where the first one is formed by a random measure G that stochastically deviates from the baseline measure G_0 and the second level is given by the Dirichlet Process $DP(G_0, \alpha)$. The baseline measure G_0 is specified in our Assumptions A7a,b (SR) and B7a,b (CR).

The level formed by G can be integrated out to obtain a representation of the prior in terms of successive conditional distributions of a mixture form (Blackwell and MacQueen 1973):

$$v_i | v_{-i} \sim \frac{1}{i-1+\alpha} \sum_{j=1}^{i-1} \delta(v_j) + \frac{\alpha}{i-1+\alpha} G_0 \tag{OA.1.1}$$

where v_{-i} denotes the collection of v_j, $j \neq i$, and $\delta(v_j)$ is the Dirac measure concentrated on the single point v_j. When combined with the likelihood, this yields the following conditional distribution for use in Gibbs sampling (Neal 2000):

$$\begin{aligned} v_i | v_{-i}, t_i &\sim \sum_{j=1, j\neq i}^{N} q_{ij} \delta(v_j) + q_0 H_i \\ q_{ij} &= bF(t_i, v_j) \\ q_0 &= b\alpha \int F(t_i, v) dG_0(v) \end{aligned} \tag{OA.1.2}$$

where H_i is the posterior for v_i based on the prior $G_0(v)$ and the single observation t_i with likelihood denoted by $F(t_i, v_i)$, b is a normalizing constant such that $\sum_{j\neq i} q_{ij} + q_0 = 1$, and N is the sample size. In the SR case, implementation of the GIG mixture model (Assumptions A1–A3, A6, and A7a) uses (2.8) and (2.9) for $F(t_i, v)$, while $\int F(t_i, v) dG_0(v)$ is given by (2.18) and (2.19). The gamma mixture model (Assumptions A1–A3, A6, and A7b) uses (2.8), (2.9), (2.22) and (2.23), respectively. In the CR case, the GIG mixture model (Assumptions B1–B3, B6, and B7a) uses (3.3) and (3.4) for $F(t_i, v)$ from Theorem 1, and $\int F(t_i, v) dG_0(v)$, as derived in Corollary 1 to Theorem 2. The gamma mixture model (Assumptions B1–B3, B6, and B7b) uses (OA.2.47) and (OA.2.49) for the latter integral, from Corollary 2 to Theorem 2. The hyperparameters ψ are then updated in a separate Gibbs block as given by Algorithm 2 (Neal

2000, p. 254). Gibbs updates of the concentration parameter α are detailed in Escobar and West (1995).

1.3. Compilation and Runtime

All reported posterior means were obtained from Markov Chain Monte Carlo (MCMC) chains of total length of 30,000 steps with a 10,000 burn-in section. All models were implemented using the Intel Fortran 95 compiler on a 2.8GHz Unix machine under serial compilation. For a sample of 15,358 individuals, the single risk model implementation took approximately 3 hours for $T = 6$, 4 hours for $T = 13$, and 6 hours for $T = 24$ to run. In contrast, with a sample of 1,317 individuals, the competing risk model implementation took approximately 2 hours for $T = 6$, 6 hours for $T = 13$, and 13 hours for $T = 24$.

1.4. Gamma versus GIG

In the gamma mixture model we found that the probability mass of the individual heterogeneity component was accumulating at zero, with a thin right tail diverging to infinity, leading to a degenerate outcome. We believe this to be an artefact of the gamma density kernel shape with mode at zero for mean less than or equal to one. In contrast, for the GIG density under the Assumptions A6 and A7a (SR) or Assumptions B6 and B7a (CR), we obtained a well-defined stable nonparametric heterogeneity clustering without the degenerative tendencies of the gamma. We attribute this outcome to the more flexible functional form of the GIG with a well-defined mode at a strictly positive value for v for mean values smaller than one.

2. Details on Proofs and Derivations

2.1. CR Stochastic Environment

Consider the CR model setup with interval outcome data and latent exit times, as described in the main text. In this section we will initially omit the subscripts i and t and also covariates and heterogeneity variables to focus on the general model, without loss of generality. We will then include these elements into the model as needed. Denote the latent exit time variables by $\tau = (\tau_1, \ldots, \tau_K)$ while the time integration variables by $u = (u_1, \ldots, u_K)$, assumed conditionally independent.

The cause-specific hazard function for the k-th cause, which is the hazard from failing from a given cause in the presence of the competing risks, is defined as

$$\lambda_k(u_k) = \lim_{h \to 0} \frac{\Pr(u_k < \tau_k \leq u_k + h; k|\tau_k > u_k)}{h}$$

The joint hazard from all causes is

$$\begin{aligned}
\lambda(u) &= \lim_{h \to 0} \frac{\Pr(u < \tau \leq u + h|\tau > u)}{h} \\
&= \sum_{k=1}^{K} \lambda_k(u_k)
\end{aligned}$$

where all inequalities are defined element-wise. The cause-specific integrated hazard is

$$\Lambda_k(\tau_k) = \int_0^{\tau_k} \lambda_k(u_k) du_k \tag{OA.2.1}$$

and the joint integrated hazard is

$$\Lambda(\tau) = \int_0^{\tau} \lambda(u) du = \int_0^{\tau} \sum_{k=1}^{K} \lambda_k(u_k) du = \sum_{k=1}^{K} \int_0^{\tau} \lambda_k(u_k) du_k = \sum_{k=1}^{K} \Lambda_k(\tau_k) \tag{OA.2.2}$$

The joint survival function is

$$\begin{aligned} S(u) &= \Pr(\tau > u) \\ &= \exp(-\Lambda(u)) \end{aligned} \tag{OA.2.3}$$

which is the complement of the probability of failure from any cause up to time τ given by the overall cumulative distribution function

$$\begin{aligned} F(u) &= \Pr(\tau \leq u) \\ &= 1 - S(u) \end{aligned}$$

For ease of exposition, we will focus on the case of two risk types with $K = 2$. The joint density of failure at time u is thus given by

$$\begin{aligned} f(u_1, u_2) &= \frac{\partial^2 F(u_1, u_2)}{\partial u_1 \partial u_2} \\ &= -\frac{\partial^2 S(u_1, u_2)}{\partial u_1 \partial u_2} \\ &= -\frac{\partial^2 \exp(-\Lambda_1(u_1) - \Lambda_2(u_2))}{\partial u_1 \partial u_2} \\ &= \exp(-\Lambda_1(u_1) - \Lambda_2(u_2)) \lambda_1(u_1)\lambda_2(u_2) \end{aligned} \tag{OA.2.4}$$

Equation (OA.2.4) links $f(u_1, u_2)$ with the risk-specific hazard functions. Parametrization of the latter in terms of covariates and unobserved heterogeneity (V_i, ζ) is given by Assumption B2. We will now invoke this Assumption and reintroduce (V_i, ζ), while suppressing notational conditioning on the covariates X without loss of generality.

Note that conditional on X the failure times u_1 and u_2 are dependent since ζ_{1t} and ζ_{2t} are correlated. However, conditional on X, V, ζ the failure times u_1 and u_2 are independent. Hence $f(u_1, u_2|V, \zeta)$ can be factorized into the product

$$f(u_1, u_2|V, \zeta) = f(u_1|V, \zeta) f(u_2|V, \zeta)$$

From (OA.2.4) it follows that

$$f(u_k|V, \zeta) = \exp(-\Lambda_k(u_k)) \lambda_k(u_k) \tag{OA.2.5}$$

Define the function

$$S_k(u_k) \equiv \exp(-\Lambda_k(u_k)) \tag{OA.2.6}$$

for $k \in \{1, 2\}$. From (OA.2.5) and (OA.2.6) we have,

$$f(u_k|V, \zeta) = S_k(u_k)\lambda_k(u_k) \tag{OA.2.7}$$

From (OA.2.1), (OA.2.6), and (OA.2.7) it follows that

$$\int_{\underline{\tau}_i}^{\overline{\tau}_i} f(u_k|V,\zeta)du_k = S_{k(t-1)} - S_{kt} \tag{OA.2.8}$$

The density (OA.2.7) should not be confused with the so-called subdensity function $f_j(u_j) = S(u = u_j)\lambda_j(u_j)$ that is sometimes used in CR analysis. Moreover, the function $S_k(u_k)$ defined in (OA.2.6) does not, in general, have the survival function interpretation for $K > 1$. Nonetheless, examining (OA.2.2), (OA.2.3), and (OA.2.6) reveals that the product of $S_k(u_k)$ over k equals the joint survival function:

$$S(u) = \prod_{k=1}^{K} S_k(u_k) \tag{OA.2.9}$$

(for further details of interpretation of functions with survival-like properties see e.g. Porta, Gómez, and Calle 2008). In general, the unconditional product form of (OA.2.9) characterizes independent risks. However, dependence among risks can be introduced by conditioning each $S_k(u_k)$ on variables correlated across the risk types.

2.2. Competing Risk Model: Conditional Likelihood

From (3.2),

$$P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}|V_i,\zeta) = A + B \tag{OA.2.10}$$

where

$$\begin{aligned} A &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{u_1}^{\overline{\tau}_i} f(u_1,u_2|V_i,\zeta)du_2du_1 \\ B &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{\overline{\tau}_i}^{\infty} f(u_1,u_2|V_i,\zeta)du_2du_1 \end{aligned}$$

The expression A is more difficult to evaluate than B since in A the lower bound u_1 of the inner integral is an argument of the outer integral. In contrast, the two integrals in B are independent of each other and hence can be factorized.

Thus,

$$\begin{aligned} A &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{u_1}^{\overline{\tau}_i} f_{it}(u_1|V_{1i},\zeta_1)f_{it}(u_2|V_{2i},\zeta_2)du_2du_1 \\ &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} \left[\int_{u_1}^{\overline{\tau}_i} f_{it}(u_2|V_{2i},\zeta_2)du_2\right] f_{it}(u_1|V_{1i},\zeta_1)du_1 \end{aligned} \tag{OA.2.11}$$

where $u_k \in [\underline{\tau}_i, \overline{\tau}_i)$ for $k \in \{1, 2\}$. For the inner integral in (OA.2.11), using (OA.2.8)

$$\int_{u_1}^{\overline{\tau}_i} f_{it}(u_2|V_{2i},\zeta_2)du_2 = S_{2i}(u_1) - S_{2it} \tag{OA.2.12}$$

Let $s_k = u_k - \underline{\tau}_i$ so that $s_k \in [0, 1)$. Then, from (OA.2.5), using piecewise constancy of the hazard function $\lambda_{ki}(\cdot)$ and hence piecewise linearity of the integrated hazard function $\Lambda_{ki}(\cdot)$

over time,

$$f_{it}(u_k|V_{ki},\zeta_k) = \exp\left(-\Lambda_{ki}\left(u_k\right)\right)\lambda_{ki}(u_k)$$
$$= \exp\left(-\Lambda_{ki(t-1)} - s_k\lambda_{kit}\right)\lambda_{kit}$$
$$= \exp\left(-\Lambda_{ki(t-1)}\right)\exp\left(-s_k\lambda_{kit}\right)\lambda_{kit} \quad \text{(OA.2.13)}$$

Similarly,

$$S_{ki}\left(u_j\right) = S_{ki(t-1)}\exp\left(-\left(s_j\lambda_{kit}\right)\right) \quad \text{(OA.2.14)}$$

for $k, j \in \{1, 2\}$. Using (OA.2.12), and integration by substitution with (OA.2.13) for $k = 1$ and with (OA.2.14) for $k = 2$, $j = 1$, in (OA.2.11) yields

$$A = \int_{\underline{\tau}_i}^{\overline{\tau}_i}\left[S_{2i}\left(u_1\right) - S_{2it}\right] f_{it}(u_1|V_{1i},\zeta_1)du_1$$
$$= S_{2i(t-1)}\int_0^1 \exp\left(-\left(s_1\lambda_{2it}\right)\right)\exp\left(-\Lambda_{1i(t-1)}\right)\exp\left(-s_1\lambda_{1it}\right)\lambda_{1it}ds_1$$
$$-S_{2it}\int_0^1 \exp\left(-\Lambda_{1i(t-1)}\right)\exp\left(-s_1\lambda_{1it}\right)\lambda_{1it}ds_1$$
$$= A_{11} + A_{12} \quad \text{(OA.2.15)}$$

where

$$A_{11} = S_{2i(t-1)}\int_0^1 \exp\left(-\left(s_1\lambda_{2it}\right)\right)\exp\left(-\Lambda_{1i(t-1)}\right)\exp\left(-s_1\lambda_{1it}\right)\lambda_{1it}ds_1$$
$$= S_{2i(t-1)}S_{1i(t-1)}\lambda_{1it}\left(\lambda_{2it} + \lambda_{1it}\right)^{-1}\int_0^1 \exp\left(-(s_1\left(\lambda_{2it} + \lambda_{1it}\right)\right)\left(\lambda_{2it} + \lambda_{1it}\right)ds_1$$
$$= -S_{2i(t-1)}S_{1i(t-1)}\lambda_{1it}\left(\lambda_{2it} + \lambda_{1it}\right)^{-1}\left[\exp\left(-\left(\lambda_{2it} + \lambda_{1it}\right)\right) - 1\right] \quad \text{(OA.2.16)}$$

and

$$A_{12} = -S_{2it}\int_0^1 \exp\left(-\Lambda_{1i(t-1)}\right)\exp\left(-s_1\lambda_{1it}\right)\lambda_{1it}ds_1$$
$$= -S_{2it}S_{1i(t-1)}\int_0^1 \exp\left(-s_1\lambda_{1it}\right)\lambda_{1it}ds_1$$
$$= S_{2it}S_{1i(t-1)}\left[\exp\left(-\lambda_{1it}\right) - 1\right] \quad \text{(OA.2.17)}$$

Using (OA.2.16) and (OA.2.17) in (OA.2.15) yields

$$A = S_{2it}S_{1it}\left\{1 - \exp(\lambda_{1it}) - \lambda_{1it}\left(\lambda_{2it} + \lambda_{1it}\right)^{-1}\left[1 - \exp\left(\lambda_{2it1} + \lambda_{1it}\right)\right]\right\} \quad \text{(OA.2.18)}$$

The expression for B of (3.2) is given by

$$B = \left[F_{1it} - F_{1i(t-1)}\right]\left[1 - F_{2it}\right]$$
$$= \left[S_{1i(t-1)} - S_{1it}\right] S_{2it}$$
$$= S_{1i(t-1)}S_{2it} - S_{1it}S_{2it}$$
$$= S_{1it}S_{2it}\left[\exp(\lambda_{1it}) - 1\right] \quad \text{(OA.2.19)}$$

Combining (OA.2.18) and (OA.2.19) in (OA.2.10) yields

$$\begin{aligned} P(t_{1i} &= t,\ \tau_{2i} > \tau_{1i}|V_i,\zeta) = S_{2i(t-1)}S_{1i(t-1)}\lambda_{1it}\left(\lambda_{2it}+\lambda_{1it}\right)^{-1} \\ &\times\left[1-\exp\left(-\left(\lambda_{2it}+\lambda_{1it}\right)\right)\right] \end{aligned}$$

with the resulting log-likelihood

$$\begin{aligned} \ln P(t_{1i}=t,\ \tau_{2i}>\tau_{1i}|V_i,\zeta) &= -\Lambda_{2i(t-1)}-\Lambda_{1i(t-1)}+\log(\lambda_{1it})-\log\left(\lambda_{2it}+\lambda_{1it}\right) \\ &\quad +\log\left(1-\exp\left(-\left(\lambda_{2it}+\lambda_{1it}\right)\right)\right) \end{aligned}$$

2.3. **Competing Risk Model: Integrated Likelihood**

2.3.1. ***Quadrature Version***

Here we derive an expression for the expectation of the exit probability (3.2) with respect to unobserved heterogeneity for each risk type, based on a simple quadrature. Taking the expectation of (3.2) yields

$$\begin{aligned} E_vP(t_{1i}=t,\ \tau_{2i}>\tau_{1i}) &= E_v\int_{\underline{\tau}_i}^{\overline{\tau}_i}\int_{u_1}^{\overline{\tau}_i} f(u_1,u_2|V_i,\zeta)du_2du_1 \\ &\quad +E_v\int_{\underline{\tau}_i}^{\overline{\tau}_i}\int_{\overline{\tau}_i}^{\infty} f(u_1,u_2|V_i,\zeta)du_2du_1 \\ &= E_v\int_{\underline{\tau}_i}^{\overline{\tau}_i}\int_{u_1}^{\infty} f(u_1,u_2|V_i,\zeta)du_2du_1 \\ &= E_v\int_{\underline{\tau}_i}^{\overline{\tau}_i}\int_{u_1}^{\infty} f_{it}(u_1|V_{1i},\zeta_1)f_{it}(u_2|V_{2i},\zeta_2)du_2du_1 \\ &= E_v\int_{\underline{\tau}_i}^{\overline{\tau}_i}\int_{u_1}^{\infty} f_{it}(u_2|V_{2i},\zeta_2)du_2f_{it}(u_1|V_{1i},\zeta_1)du_1 \\ &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} E_{v_{2i}}\left[S_{2i}\left(u_1\right)\right]E_{v_{1i}}\left[f_{it}(u_1|V_{1i},\zeta_1)\right]du_1 \end{aligned} \tag{OA.2.20}$$

From (OA.2.6),

$$E_{v_{2i}}\left[S_{2i}\left(u_1\right)\right]=\mathcal{L}_2\left(\widetilde{\Lambda}_{2i}(u_1)\right) \tag{OA.2.21}$$

Using (OA.2.5),

$$\begin{aligned} E_{v_{1i}}\left[f_{it}(u_1|V_{1i},\zeta_1)\right] &= E_{v_{1i}}\left[\exp\left(-\Lambda_{1i}\left(u_1\right)\right)\lambda_{1i}(u_1)\right] \\ &= \widetilde{\lambda}_{1i}(u_1)E_{v_{1i}}\left[\exp\left(-v_{1i}\widetilde{\Lambda}_{1i}(u_1)\right)v_{1i}\right] \\ &= -\widetilde{\lambda}_{1i}(u_1)\mathcal{L}_1^{(1)}\left(\widetilde{\Lambda}_{1i}(u_1)\right) \end{aligned} \tag{OA.2.22}$$

where $\mathcal{L}^{(1)}(s)$ is the first derivative of the Laplace transform $\mathcal{L}(s)$ evaluated at s. Using (OA.2.21) and (OA.2.22) in (OA.2.20) yields

$$E_vP(t_{1i}=t,\ \tau_{2i}>\tau_{1i})=-\int_{\underline{\tau}_i}^{\overline{\tau}_i}\widetilde{\lambda}_{1i}(u_1)\mathcal{L}_2\left(\widetilde{\Lambda}_{2i}(u_1)\right)\mathcal{L}_1^{(1)}\left(\widetilde{\Lambda}_{1i}(u_1)\right)du_1 \tag{OA.2.23}$$

Letting again $s_k = u_k - (t-1)$, $s_k \in [0,1)$, $k \in \{1,2\}$, and using piecewise constancy of $\lambda_{ki}(\cdot)$ and piecewise linearity of $\Lambda_{ki}(\cdot)$, following a change of variables (OA.2.23) becomes
(OA.2.24)

$$E_v P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}) = -\widetilde{\lambda}_{1it} \int_0^1 \mathcal{L}_2\left(\widetilde{\Lambda}_{2i(t-1)} + \widetilde{\lambda}_{2it} s_1\right) \mathcal{L}_1^{(1)}\left(\widetilde{\Lambda}_{1i(t-1)} + \widetilde{\lambda}_{1it} s_1\right) ds_1$$

2.3.2. ***Series Expansion***

The series expansion expression for the expectation of (3.2) can be derived as follows. Using (OA.2.10) and taking expectations,

$$\begin{aligned} E_v P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}) &= E_v \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{u_1}^{\overline{\tau}_i} f(u_1, u_2 | V_i, \zeta) du_2 du_1 \\ &\quad + E_v \int_{\underline{\tau}_i}^{\overline{\tau}_i} \int_{\overline{\tau}_i}^{\infty} f(u_1, u_2 | V_i, \zeta) du_2 du_1 \\ &= E_v A + E_v B \end{aligned} \qquad \text{(OA.2.25)}$$

From (OA.2.11),

$$E_v A = \int_{\underline{\tau}_i}^{\overline{\tau}_i} E_{v_{2i}} \left[\int_{u_1}^{\overline{\tau}_i} f_{it}(u_2 | V_{2i}, \zeta_2) du_2\right] E_{v_{1i}} \left[f_{it}(u_1 | V_{1i}, \zeta_1)\right] du_1 \qquad \text{(OA.2.26)}$$

For the expectation of the inner integral,

$$\begin{aligned} E_{v_{2i}} \left[\int_{u_1}^{\overline{\tau}_i} f_{it}(u_2 | V_{2i}, \zeta_2) du_2\right] &= E_{v_{2i}} \left[S_{2i}(u_1) - S_{2it}\right] \\ &= E_{v_{2i}} \left[S_{2i}(u_1)\right] - \mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right) \end{aligned} \qquad \text{(OA.2.27)}$$

with the first right-hand side term intentionally not converted to the Laplace form in order to facilitate subsequent series expansion. Using (OA.2.27) in (OA.2.26),

$$\begin{aligned} E_v A &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} \left[E_{v_{2i}} \left[S_{2i}(u_1)\right] - \mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right)\right] E_{v_{1i}} \left[f_{it}(u_1 | V_{1i}, \zeta_1)\right] du_1 \\ &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} E_{v_{2i}} \left[S_{2i}(u_1)\right] E_{v_{1i}} \left[f_{it}(u_1 | V_{1i}, \zeta_1)\right] du_1 \\ &\quad - \mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right) E_{v_{1i}} \int_{\underline{\tau}_i}^{\overline{\tau}_i} f_{it}(u_1 | V_{1i}, \zeta_1) du_1 \\ &= E_v A_1 + E_v A_2 \end{aligned} \qquad \text{(OA.2.28)}$$

Substituting with (OA.2.6) and (OA.2.7),

$$\begin{aligned}
E_v A_1 &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} E_{v_{2i}}\left[S_{2i}\left(u_1\right)\right] E_{v_{1i}}\left[f_{it}(u_1|V_{1i},\zeta_1)\right] du_1 \\
&= E_{v_{2i}} E_{v_{1i}} \int_{\underline{\tau}_i}^{\overline{\tau}_i} S_{2i}(u_1)\left[f_{it}(u_1|V_{1i},\zeta_1)\right] du_1 \\
&= E_{v_{2i}} E_{v_{1i}} \int_{\underline{\tau}_i}^{\overline{\tau}_i} \exp\left(-\Lambda_{2i}\left(u_1\right)\right) \exp\left(-\Lambda_{1i}\left(u_1\right)\right) \lambda_{1i}(u_1) du_1 \\
&= E_{v_{2i}} E_{v_{1i}} \int_{\underline{\tau}_i}^{\overline{\tau}_i} \exp\left(-v_{2i}\widetilde{\Lambda}_{2i}\left(u_1\right)\right) \exp\left(-v_{1i}\widetilde{\Lambda}_{1i}\left(u_1\right)\right) v_{1i}\widetilde{\lambda}_{1i}(u_1) du_1 \qquad \text{(OA.2.29)}
\end{aligned}$$

Using integration by substitution with $s_k = u_k - \underline{\tau}_i$ in (OA.2.29) and piecewise constancy of $\widetilde{\lambda}_{ki}\left(s_k\right)$ for $s_k \in [0,1)$, $k \in \{1,2\}$,

$$\begin{aligned}
E_v A_1 &= E_{v_{2i}} E_{v_{1i}} \exp\left(-v_{2i}\widetilde{\Lambda}_{2i(t-1)}\right) \exp\left(-v_{1i}\widetilde{\Lambda}_{1i(t-1)}\right) \\
&\quad \times \int_0^1 \exp\left(-v_{2i}s_1\widetilde{\lambda}_{2it}\right) \exp\left(-v_{1i}s_1\widetilde{\lambda}_{1it}\right) v_{1i}\widetilde{\lambda}_{1it} ds_1 \\
&= E_{v_{2i}} E_{v_{1i}} \exp\left(-v_{2i}\widetilde{\Lambda}_{2i(t-1)}\right) \exp\left(-v_{1i}\widetilde{\Lambda}_{1i(t-1)}\right) \\
&\quad \times \int_0^1 \sum_{r_2=0}^{\infty} \frac{(-1)^{r_2}}{r_2!} \left(v_{2i}s_1\widetilde{\lambda}_{2it}\right)^{r_2} \sum_{r_1=0}^{\infty} \frac{(-1)^{r_1}}{r_1!} \left(v_{1i}s_1\widetilde{\lambda}_{1it}\right)^{r_1} v_{1i}\widetilde{\lambda}_{1it} ds_1 \\
&= \sum_{r_2=0}^{\infty} \frac{(-1)^{r_2}}{r_2!} \sum_{r_1=0}^{\infty} \frac{(-1)^{r_1}}{r_1!} E_{v_{1i}}\left[\exp\left(-v_{1i}\widetilde{\Lambda}_{1i(t-1)}\right)\left(v_{1i}\widetilde{\lambda}_{1it}\right)^{r_1+1}\right] \\
&\quad \times E_{v_{2i}}\left[\exp\left(-v_{2i}\widetilde{\Lambda}_{2i(t-1)}\right)\left(v_{2i}\widetilde{\lambda}_{2it}\right)^{r_2}\right] \int_0^1 s_1^{r_2+r_1} ds_1 \\
&= \sum_{r_2=0}^{\infty} \frac{(-1)^{r_2}}{r_2!} \sum_{r_1=0}^{\infty} \frac{(-1)^{r_1}}{r_1!} E_{v_1}\left[A_{11}\right] E_{v_2}\left[A_{12}\right] A_{13} \qquad \text{(OA.2.30)}
\end{aligned}$$

where

$$\begin{aligned}
E_{v_1}\left[A_{11}\right] &= \widetilde{\lambda}_{1it}^{r_1+1} E_{v_{1i}}\left[\exp\left(-v_{1i}\widetilde{\Lambda}_{1i(t-1)}\right) v_{1i}^{r_1+1}\right] && \text{(OA.2.31)} \\
&= (-1)^{r_1+1}\widetilde{\lambda}_{1it}^{r_1+1}\mathcal{L}_1^{(r_1+1)}\left(\widetilde{\Lambda}_{1i(t-1)}\right) && \text{(OA.2.32)} \\
E_{v_2}\left[A_{12}\right] &= \widetilde{\lambda}_{2it}^{r_2} E_{v_{2i}}\left[\exp\left(-v_{2i}\widetilde{\Lambda}_{2i(t-1)}\right) v_{2i}^{r_2}\right] && \text{(OA.2.33)} \\
&= (-1)^{r_2}\widetilde{\lambda}_{2it}^{r_2}\mathcal{L}_2^{(r_2)}\left(\widetilde{\Lambda}_{2i(t-1)}\right) && \text{(OA.2.34)} \\
A_{13} &= \int_0^1 s_1^{r_2+r_1} ds_1 \\
&= \frac{1}{r_2+r_1+1} && \text{(OA.2.35)}
\end{aligned}$$

whereby the time dimension of the previous quadrature has been parsed through following the series expansion linearization and integrated out in the remaining polynomial term in (OA.2.35).

Combining (OA.2.32) and (OA.2.34) and (OA.2.35) in (OA.2.30) results in

$$\begin{aligned} E_v A_1 &= \sum_{r_2=0}^{\infty} \frac{(-1)^{r_2}}{r_2!} \sum_{r_1=0}^{\infty} \frac{(-1)^{r_1}}{r_1!} \frac{(-1)^{r_1+r_2+1}}{r_2+r_1+1} \widetilde{\lambda}_{1it}^{r_1+1} \widetilde{\lambda}_{2it}^{r_2} \\ &\quad \times \mathcal{L}_1^{(r_1+1)}\left(\widetilde{\Lambda}_{1i(t-1)}\right) \mathcal{L}_2^{(r_2)}\left(\widetilde{\Lambda}_{2i(t-1)}\right) \end{aligned} \tag{OA.2.36}$$

For the second part of (OA.2.28),

$$\begin{aligned} E_v A_2 &= -\mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right) E_{v_{1i}} \int_{\underline{\tau}_i}^{\overline{\tau}_i} f_{it}(u_1|V_{1i}, \zeta_1) du_1 \\ &= -\mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right) E_{v_{1i}} \left[S_{1i(t-1)} - S_{1it}\right] \\ &= -\mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right) \left[\mathcal{L}_1\left(\widetilde{\Lambda}_{1i(t-1)}\right) - \mathcal{L}_1\left(\widetilde{\Lambda}_{1it}\right)\right] \end{aligned} \tag{OA.2.37}$$

Collecting (OA.2.36) and (OA.2.37) in (OA.2.28) yields

$$\begin{aligned} E_v A &= \sum_{r_2=0}^{\infty} \frac{(-1)^{r_2}}{r_2!} \sum_{r_1=0}^{\infty} \frac{(-1)^{r_1}}{r_1!} \frac{(-1)^{r_1+r_2+1}}{r_2+r_1+1} \widetilde{\lambda}_{1it}^{r_1+1} \widetilde{\lambda}_{2it}^{r_2} \\ &\quad \times \mathcal{L}_1^{(r_1+1)}\left(\widetilde{\Lambda}_{1i(t-1)}\right) \mathcal{L}_2^{(r_2)}\left(\widetilde{\Lambda}_{2i(t-1)}\right) \\ &\quad -\mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right) \left[\mathcal{L}_1\left(\widetilde{\Lambda}_{1i(t-1)}\right) - \mathcal{L}_1\left(\widetilde{\Lambda}_{1it}\right)\right] \end{aligned} \tag{OA.2.38}$$

The expectation expression for B in (OA.2.25) is

$$\begin{aligned} E_v B &= \int_{\underline{\tau}_i}^{\overline{\tau}_i} E_{v_{2i}} \left[\int_{\overline{\tau}_i}^{\infty} f_{it}(u_2|V_{2i}, \zeta_2) du_2\right] E_{v_{1i}} \left[f_{it}(u_1|V_{1i}, \zeta_1)\right] du_1 \\ &= E_{v_{2i}} \left[S_{2it}\right] E_{v_{1i}} \left[S_{1i(t-1)} - S_{1it}\right] \\ &= \mathcal{L}_2\left(\widetilde{\Lambda}_{2it}\right) \left[\mathcal{L}_1\left(\widetilde{\Lambda}_{1i(t-1)}\right) - \mathcal{L}_1\left(\widetilde{\Lambda}_{1it}\right)\right] \end{aligned} \tag{OA.2.39}$$

Substituting (OA.2.38) and (OA.2.39) into (OA.2.25) yields

$$\begin{aligned} E_v P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}) &= \sum_{r_2=0}^{\infty} \sum_{r_1=0}^{\infty} \frac{(-1)^{2r_1+2r_2+1}}{r_2! r_1! \left(r_2+r_1+1\right)} \widetilde{\lambda}_{1it}^{r_1+1} \widetilde{\lambda}_{2it}^{r_2} \\ &\quad \times \mathcal{L}_1^{(r_1+1)}\left(\widetilde{\Lambda}_{1i(t-1)}\right) \mathcal{L}_2^{(r_2)}\left(\widetilde{\Lambda}_{2i(t-1)}\right) \end{aligned} \tag{OA.2.40}$$

2.3.3. ***Derivatives of the Laplace transform***

In general,

$$\mathcal{L}^{(r)}(s) = (-1)^r \int v^r \exp(-sv) g(v) dv \tag{OA.2.41}$$

(see e.g. Hougaard 2000, p. 498) and $\mathcal{L}^{(r)}(s)$ exists for each $r > c$ such that $|g(v)| \leq K \exp(cv)$ if $g(v)$ is piecewise continuous over its domain.

In the GIG density function (2.16), replace θ with $\theta/2$, then let $\chi = \varphi^2/\theta$, and then substitute the resulting expression into (OA.2.41) to obtain

$$\begin{aligned}
\mathcal{L}^{(r)GIG}(s) &= (-1)^r \int v^r \exp(-sv)\, g^{GIG}(v) dv \\
&= (-1)^r \int v^r \exp(-sv) \frac{(\theta/\chi)^{\kappa/2}}{2K_\kappa\left((\theta\chi)^{1/2}\right)} v^{\kappa-1} \exp\left\{-\frac{1}{2}\left(\theta v + \frac{\chi}{v}\right)\right\} dv \\
&= (-1)^r \int \frac{(\theta/\chi)^{\kappa/2}}{2K_\kappa\left((\theta\chi)^{1/2}\right)} v^{\kappa+r-1} \exp\left\{-\frac{1}{2}\left((\theta+2s)\, v + \frac{\chi}{v}\right)\right\} dv \\
&= (-1)^r \frac{2K_{\kappa+r}\left(((\theta+2s)\,\chi)^{1/2}\right)}{2K_{\kappa+r}\left(((\theta+2s)\,\chi)^{1/2}\right)} \frac{((\theta+2s)/\chi)^{(\kappa+r)/2}}{((\theta+2s)/\chi)^{(\kappa+r)/2}} \\
&\quad \times \int \frac{(\theta/\chi)^{\kappa/2}}{2K_\kappa\left((\theta\chi)^{1/2}\right)} v^{\kappa+r-1} \exp\left\{-\frac{1}{2}\left((\theta+2s)\, v + \frac{\chi}{v}\right)\right\} dv \\
&= (-1)^r \frac{K_{\kappa+r}\left(((\theta+2s)\,\chi)^{1/2}\right)}{K_\kappa\left((\theta\chi)^{1/2}\right)} \frac{(\theta/\chi)^{\kappa/2}}{((\theta+2s)/\chi)^{(\kappa+r)/2}} \\
&\quad \times \int \frac{((\theta+2s)/\chi)^{(\kappa+r)/2}}{2K_{\kappa+r}\left(((\theta+2s)\,\chi)^{1/2}\right)} v^{\kappa+r-1} \exp\left\{-\frac{1}{2}\left((\theta+2s)\, v + \frac{\chi}{v}\right)\right\} dv \\
&= (-1)^r \frac{K_{\kappa+r}\left(((\theta+2s)\,\chi)^{1/2}\right)}{K_\kappa\left((\theta\chi)^{1/2}\right)} \frac{(\theta/\chi)^{\kappa/2}}{((\theta+2s)/\chi)^{(\kappa+r)/2}}
\end{aligned}$$

Reversing the substitution with $\varphi = \sqrt{\theta\chi}$ and then replacing θ with 2θ yields

(OA.2.42) $$\mathcal{L}^{(r)GIG}(s) = (-1)^r \frac{K_{\kappa+r}\left(\varphi\,(1+s/\theta)^{1/2}\right)}{K_\kappa(\varphi)} \left(\frac{\varphi}{2\theta}\right)^r (1+s/\theta)^{-(\kappa+r)/2}$$

The quadrature version for the GIG then follows from (2.17), (OA.2.24), and (OA.2.42).

$$E_v^{GIG} P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}) = \frac{\widetilde{\lambda}_{1it}\varphi_1}{2\theta_1 K_{\kappa_1}(\varphi_1) K_{\kappa_2}(\varphi_2)}$$
$$\times \int_0^1 \left(1 + \frac{1}{\theta_1}\widetilde{\Lambda}_{1i(t-1)} + \frac{1}{\theta_1}\widetilde{\lambda}_{1it}s_1\right)^{-(\kappa_1+1)/2}$$
$$\times \left(1 + \frac{1}{\theta_2}\widetilde{\Lambda}_{2i(t-1)} + \frac{1}{\theta_2}\widetilde{\lambda}_{2it}s_1\right)^{-\kappa_2/2}$$
$$\times K_{\kappa_1+1}\left(\varphi_1\left(1 + \frac{1}{\theta_1}\widetilde{\Lambda}_{1i(t-1)} + \frac{1}{\theta_1}\widetilde{\lambda}_{1it}s_1\right)^{1/2}\right)$$
$$\times K_{\kappa_2}\left(\varphi_2\left(1 + \frac{1}{\theta_2}\widetilde{\Lambda}_{2i(t-1)} + \frac{1}{\theta_2}\widetilde{\lambda}_{2it}s_1\right)^{1/2}\right) ds_1 \tag{OA.2.43}$$

The series version follows from (OA.2.40) and (OA.2.42).

$$E_v^{GIG} P(t_{1i} = t,\ \tau_{2i} > \tau_{1i}) = \sum_{r_2=0}^{\infty}\sum_{r_1=0}^{\infty} \frac{(-1)^{3r_1+3r_2}}{r_2! r_1! (r_2 + r_1 + 1)} \left(\frac{\widetilde{\lambda}_{1it}\varphi_1}{2\theta_1}\right)^{r_1+1} \left(\frac{\widetilde{\lambda}_{2it}\varphi_2}{2\theta_2}\right)^{r_2}$$
$$\times \left(1 + \frac{1}{\theta_1}\widetilde{\Lambda}_{1i(t-1)}\right)^{-(\kappa+r_1+1)/2} \left(1 + \frac{1}{\theta_2}\widetilde{\Lambda}_{2i(t-1)}\right)^{-(\kappa_2+r_2)/2}$$
$$\times K_{\kappa_1+r_1+1}\left(\varphi_1\left(1 + \frac{1}{\theta_1}\widetilde{\Lambda}_{1i(t-1)}\right)^{1/2}\right) [K_{\kappa_1}(\varphi_1)]^{-1}$$
$$\times K_{\kappa_2+r_2}\left(\varphi_2\left(1 + \frac{1}{\theta_2}\widetilde{\Lambda}_{2i(t-1)}\right)^{1/2}\right) [K_{\kappa_2}(\varphi_2)]^{-1} \tag{OA.2.44}$$

The censored case,

$$E_v^{GIG} P(t_{1i} > T,\ t_{2i} > T) = \left(1 + \frac{1}{\theta_1}\widetilde{\Lambda}_{1iT}\right)^{-\kappa_1/2} \left(1 + \frac{1}{\theta_2}\widetilde{\Lambda}_{2iT}\right)^{-\kappa_2/2}$$
$$\times K_{\kappa_1}\left(\varphi_1\left(1 + \frac{1}{\theta_1}\widetilde{\Lambda}_{1iT}\right)^{1/2}\right) [K_{\kappa_1}(\varphi_1)]^{-1}$$
$$\times K_{\kappa_2}\left(\varphi_2\left(1 + \frac{1}{\theta_2}\widetilde{\Lambda}_{2iT}\right)^{1/2}\right) [K_{\kappa_2}(\varphi_2)]^{-1} \tag{OA.2.45}$$

Expressions (OA.2.43), (OA.2.44), and (OA.2.45) are referenced in Corollary 1 to Theorem 2.

For the gamma density function (2.20),

$$\begin{aligned}
\mathcal{L}^{(r)G}(s) &= (-1)^r \int v^r \exp(-sv)\, g^G(v)dv \\
&= (-1)^r \int v^r \exp(-sv)\, \theta^\gamma \frac{1}{\Gamma(\gamma)} v^{\gamma-1} \exp(-\theta v)dv \\
&= (-1)^r \int \theta^\gamma \frac{1}{\Gamma(\gamma)} v^{\gamma+r-1} \exp(-(\theta+s)\, v)dv \\
&= (-1)^r \frac{(\theta+s)^{\gamma+r}}{(\theta+s)^{\gamma+r}} \frac{\Gamma(\gamma+r)}{\Gamma(\gamma+r)} \\
&\quad \times \int \theta^\gamma \frac{1}{\Gamma(\gamma)} v^{\gamma+r-1} \exp(-(\theta+s)\, v)dv \\
&= (-1)^r \frac{\theta^\gamma}{(\theta+s)^{\gamma+r}} \frac{\Gamma(\gamma+r)}{\Gamma(\gamma)} \\
&\quad \times \int (\theta+s)^{\gamma+r} \frac{1}{\Gamma(\gamma+r)} v^{\gamma+r-1} \exp(-(\theta+s)\, v)dv \\
&= (-1)^r \frac{\theta^\gamma}{(\theta+s)^{\gamma+r}} \frac{\Gamma(\gamma+r)}{\Gamma(\gamma)} \qquad \text{(OA.2.46)}
\end{aligned}$$

The quadrature version for the gamma then follows from (2.21), (OA.2.24), and (OA.2.46).

$$\begin{aligned}
E_v^G P(t_{1i}=t,\ \tau_{2i}>\tau_{1i}) &= \gamma_1 \frac{\widetilde{\lambda}_{1it}}{\theta_1} \int_0^1 \left(1+\frac{1}{\theta_2}\widetilde{\Lambda}_{2i(t-1)} + \frac{1}{\theta_2}\widetilde{\lambda}_{2it} s_1\right)^{-\gamma_2} \\
&\quad \times \left(1+\frac{1}{\theta_1}\widetilde{\Lambda}_{1i(t-1)} + \frac{1}{\theta_1}\widetilde{\lambda}_{1it} s_1\right)^{-(\gamma_1+1)} ds_1 \qquad \text{(OA.2.47)}
\end{aligned}$$

The series version follows from (OA.2.40) and (OA.2.46).

$$\begin{aligned}
E_v^G P(t_{1i}=t,\ \tau_{2i}>\tau_{1i}) &= \sum_{r_2=0}^{\infty}\sum_{r_1=0}^{\infty} \frac{(-1)^{r_1+r_2}}{r_1! r_2! (r_1+r_2+1)} \left(\frac{\widetilde{\lambda}_{1it}}{\theta_1}\right)^{r_1+1} \left(\frac{\widetilde{\lambda}_{2it}}{\theta_2}\right)^{r_2} \\
&\quad \times \left(1+\frac{1}{\theta_1}\widetilde{\Lambda}_{1i(t-1)}\right)^{-(\gamma_1+r_1+1)} \left(1+\frac{1}{\theta_2}\widetilde{\Lambda}_{2i(t-1)}\right)^{-(\gamma_2+r_2)} \\
&\quad \times \Gamma(\gamma_1+r_1+1)\,[\Gamma(\gamma_1)]^{-1}\, \Gamma(\gamma_2+r_2)\,[\Gamma(\gamma_2)]^{-1} \qquad \text{(OA.2.48)}
\end{aligned}$$

For the censored case,

$$E_v^G P(t_{1i}>T,\ t_{2i}>T) = \left(1+\frac{1}{\theta_1}\widetilde{\Lambda}_{1iT}\right)^{-\gamma_1} \left(1+\frac{1}{\theta_2}\widetilde{\Lambda}_{2iT}\right)^{-\gamma_2} \qquad \text{(OA.2.49)}$$

Expressions (OA.2.47), (OA.2.48), and (OA.2.45) are referenced in Corollary 2 to Theorem 2.

3. Model Identification

Cox (1962) and Tsiatis (1975) state that the simple competing risks model with no regressors is not identified. In particular, any competing risk model with correlated risks is observationally equivalent to some other competing risks model with independent risks. Heckman and Honoré (1989), henceforth HH, establish an identification theorem for a general class of competing risks models with regressors. This class includes models with marginal distributions that follow proportional hazards, mixed proportional hazards, and accelerated hazards. The results are presented for two competing risks but generalize to any arbitrary finite number of risks. HH assume that the exact time of exit is observed.

Our competing risk (CR) model is based on continuous latent times of exit $\tau_{1i}, \ldots, \tau_{Ki}$ with a minimum $\tau_i = \min(\tau_{1i}, \ldots, \tau_{Ki})$. We observe the time interval $[\underline{\tau}_i, \overline{\tau}_i)$, labeled as t_i, which contains τ_i. Nonetheless, our model assumptions impose more structure that would typically be implied by interval outcome data, which allows us to identify the structural model components. Assumption B2 explicitly parametrizes the time-varying model components as functions of the continuous time τ. In Assumptions B3 and B4, the values of these components are assumed constant within each time period t. Assumptions B2 and B4 thus allow us to adhere to a counterpart of the HH identification approach in our model setting.

As in the main text, we assume $K = 2$ risk types. HH identify the single-index structural parameters β_k up to scale from the ratio of the derivatives of the survival function with respect to a time increment of each risk type, evaluated at the time origin. Our counterpart is the ratio of the survival functions integrated over the first time period ($t = 1$). From Theorem 1,

$$\begin{aligned}
\frac{P(t_{1i}=1,\ \tau_{2i}>\tau_{1i}|V,\zeta)}{P(t_{2i}=1,\ \tau_{1i}>\tau_{2i}|V,\zeta)} &= \frac{S_{2i0}S_{1i0}\lambda_{1i1}\left(\lambda_{2i1}+\lambda_{1i1}\right)^{-1}\left[1-\exp\left(-\left(\lambda_{2i1}+\lambda_{1i1}\right)\right)\right]}{S_{1i0}S_{2i0}\lambda_{2i1}\left(\lambda_{1i1}+\lambda_{2i1}\right)^{-1}\left[1-\exp\left(-\left(\lambda_{1i1}+\lambda_{2i1}\right)\right)\right]} \\
&= \frac{\lambda_{1i1}}{\lambda_{2i1}} \\
&= \frac{v_{i1}\exp(X_{i1}\beta_1+\delta_{011})}{v_{i2}\exp(X_{i1}\beta_2+\delta_{021})}
\end{aligned}$$

Taking expectations with respect to $v_{ik} = \exp(V_{ik})$ and using the normalization restrictions $E[v_{ik}] = 1$ from Assumption B6, the absence of a constant term in X_{it}, and the support condition for X_{it} in Assumption B2 identifies the ratio of β_1 and β_2.

Conditional on $X = x$, HH assume the survival function structure

(OA.3.1) $$S(\tau|x) = \mathcal{K}\left[U_1(\tau|x), U_2(\tau|x)\right]$$

where $U_k(\tau|x) = \exp\left[-Z_k(\tau)\phi_k(x)\right]$ and $\mathcal{K}$ is a joint distribution function on $[0,1]^2$. In our case, the counterpart of (**X**) for period t outcomes, the joint expected survival function, can be expressed as

(OA.3.2) $$E_v S_t(x) = \mathcal{K}\left[U_{1t}(x), U_{2t}(x)\right]$$

where $U_{kt}(x) = \exp\left(-\sum_{j=1}^{t} z_{kt}\phi_{kt}(x)\right)$ with $\phi_{kt}(x) = \exp\left(x_{kt}\beta_k\right)$ and $z_{kt} = \exp(\delta_{0kt})$, or equivalently $U_{kt}(x) = \widetilde{\Lambda}_{kt}(x)$. Our model assumptions uniquely determine the function $\mathcal{K}$ which is given in Theorem 2: for censored observations $\mathcal{K}$ is a product of the Laplace transforms of

$\widetilde{\Lambda}_{1t}(x)$ and $\widetilde{\Lambda}_{2t}(x)$, and for non-censored observations an expression involving the derivatives of the respective Laplace transforms.

Let $\phi_{2t}(x) \to 0$ while holding $\phi_{1t}(x)$ fixed, which is feasible by the full support condition for the covariates. Then

$$E_v S_t(x) = \mathcal{K}\left[\exp\left(-\sum_{j=1}^{t} z_{1t} \exp\left(x_{1t}\beta_1\right)\right), 1\right]$$

Since $\mathcal{K}$ and $\phi_{1t}(x)$ are known and $\mathcal{K}$ is increasing in both arguments, z_{1t} can be identified for any t, and similarly for z_{2t}. Identification of ρ, σ_1^2, and σ_2^2 follow directly from identification of z_{kt} and Assumptions B2 and B4.

Using (OA.2.9) the joint expected survivor function can be expressed explicitly in terms of v_1 and v_2 as

(OA.3.3) $$E_v S_t(x) = \int_\Omega \exp\left[-v_1 U_{1t}(x)\right] \exp\left[-v_2 U_{2t}(x)\right] dG(v_1, v_2)$$

HH show nonparametric identification of G for a special case with $v_1 = v_2 = \exp(c_2\omega)$. Honoré (1993) provides the proof in full generality, albeit in that paper (OA.3.3) was obtained from a multi-spell background. The argument is that if the marginal distributions of G along with other model components are identified, then G is nonparametrically identified by the uniqueness of the multivariate Laplace transform. The same argument can be used when (OA.3.3) is obtained from a multiple risk background as we consider here.

The marginal distributions of G are in turn identified under the following Elbers and Ridder (1982) assumptions:

ER1: v_k is non-negative, with $E[v_k] = 1$.

ER2: The function $z_k(\tau)$ defined on $[0, \infty)$ can be written as the integral of a non-negative function ψ.

ER3': There are two points in the support of X, x_0 and x_1, such that $\phi(x_0) \neq \phi(x_1)$. Furthermore, $\phi(x_0) = 1$.

Assumption ER1 is satisfied by our Assumption B6, and Assumptions ER2 and ER3' are satisfied by our Assumptions B2, B6 and the definition of integrated baseline hazard.

Identification of ϕ_k in HH relies on a limit result with the time variable approaching zero. An alternative proof of nonparametric identification of a general class of CR models that does not rely on a time limit at zero is provided in a recent paper by Lee and Lewbel (2013), henceforth LL. Their approach also does not depend on exclusion restrictions and allows for discrete regressors as long as some are continuously distributed.

LL define mappings $B_k(s|x)$ and $C(s|x)$ that are identified directly from data and whose unique solution is the accelerated failure time nonparametric regression function $g(x)$. Both $B_k(s|x)$ and $C(s|x)$ are expressed as integrals over the continuous time domain which we can evaluate as well under our assumptions using the formula for the density of the continuous latent time of exit.

Other than regularity conditions that are satisfied in our model, LL rely on a key rank assumption stating that the columns of the Fréchet derivative of $C^*(s,h) = C(s|x)$ with respect to its functional argument h are linearly independent and that C^* is a proper mapping preserving compactness under inverse image. These conditions generally require that X contain at least K continuously distributed elements and also that no one element of $g(x)$ can be expressed as a function of the other elements of $g(x)$. LL show that the conditions can be met under parametric assumptions preventing non-degeneracy of the correlation structure between the competing risks, and hence we conclude that these will hold in our model.

Figure OA.1. Posterior Density of the Dirichlet Process Concentration Parameter α, GIG mixture

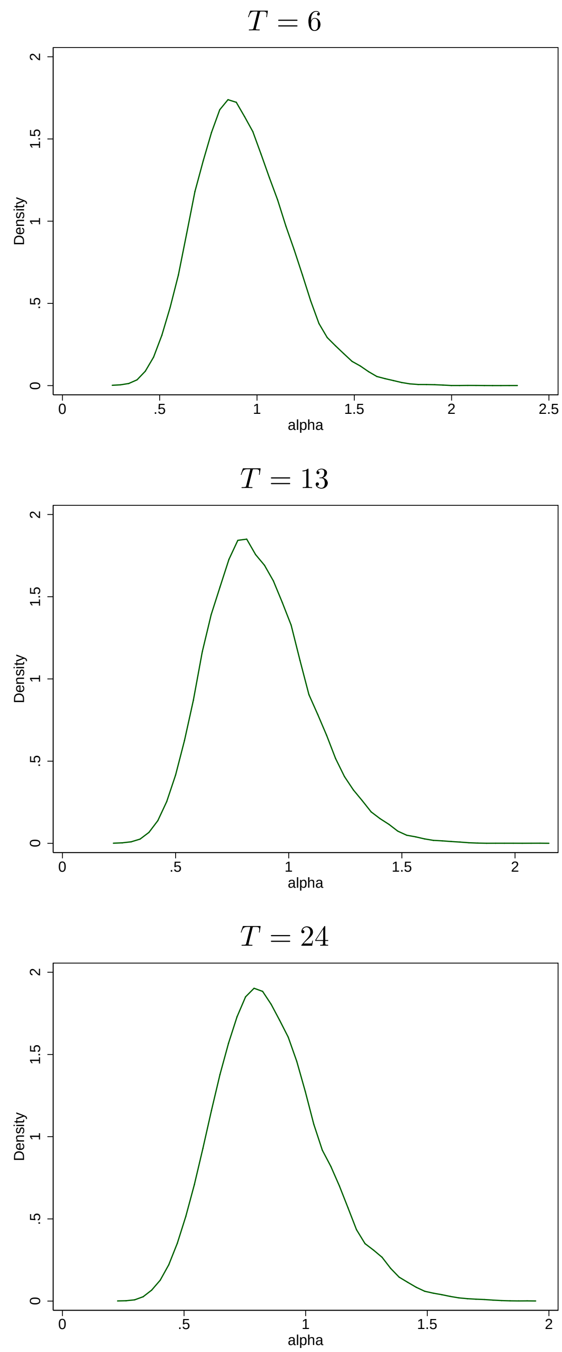


Table OA.1. Duration model with parametric gamma heterogeneity (Han and Hausman, 1990)

		6 periods		13 periods		24 periods	
		Mean	s.e.	Mean	s.e.	Mean	s.e.
γ		0.210	0.034	0.242	0.017	0.316	0.020
Urate		-0.480	0.036	-0.457	0.021	-0.402	0.012
Race		-0.184	0.070	-0.186	0.060	-0.195	0.055
Age		-0.403	0.077	-0.325	0.072	-0.284	0.048
Rrate		-1.449	0.104	-0.941	0.074	-0.383	0.052
t	1	-0.346	0.215	-0.774	0.111	-1.374	0.058
	2	0.305	0.237	-0.168	0.123	-0.819	0.058
	3	0.194	0.263	-0.303	0.137	-1.012	0.059
	4	0.688	0.289	0.153	0.139	-0.588	0.062
	5	0.525	0.322	-0.046	0.154	-0.826	0.077
	6	1.095	0.337	0.499	0.144	-0.324	0.072
	7			0.131	0.162	-0.721	0.078
	8			0.571	0.161	-0.322	0.080
	9			0.351	0.195	-0.566	0.100
	10			0.700	0.179	-0.262	0.088
	11			0.590	0.169	-0.389	0.100
	12			0.945	0.193	-0.063	0.102
	13			1.007	0.212	-0.015	0.101
	14					0.163	0.120
	15					0.305	0.096
	16					0.463	0.125
	17					0.307	0.132
	18					0.712	0.142
	19					0.658	0.153
	20					0.916	0.152
	21					0.853	0.165
	22					1.063	0.166
	23					0.995	0.202
	24					1.283	0.176

$N = 15,491$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

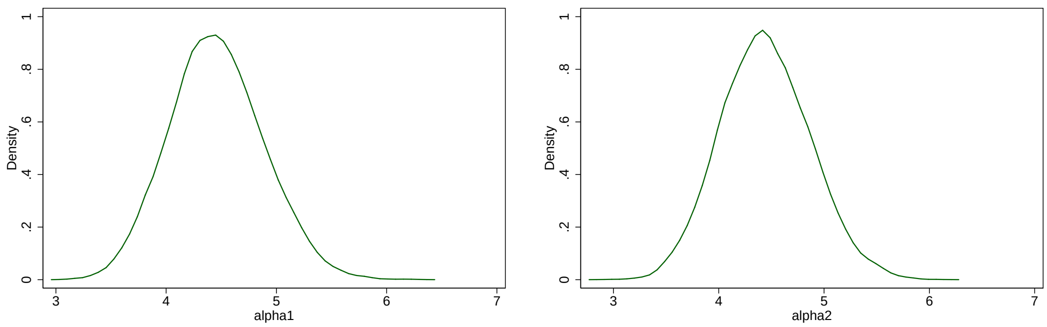
TABLE OA.2. Duration Model with Parametric GIG Heterogeneity

		6 periods		13 periods		24 periods	
		Mean	s.e.	Mean	s.e.	Mean	s.e.
κ		-0.825	0.054	-1.176	0.048	-1.501	0.072
φ		1.900	0.151	2.882	0.136	3.792	0.203
Urate		-0.307	0.018	-0.285	0.014	-0.259	0.017
Race		-0.108	0.056	-0.152	0.043	-0.111	0.044
Age		-0.273	0.058	-0.223	0.041	-0.178	0.038
Rrate		-1.190	0.069	-0.708	0.043	-0.039	0.059
t	1	0.115	0.150	0.938	0.102	1.345	0.068
	2	0.584	0.139	1.397	0.092	1.734	0.095
	3	0.309	0.145	1.124	0.099	1.510	0.069
	4	0.640	0.138	1.442	0.098	1.800	0.090
	5	0.334	0.139	1.131	0.095	1.510	0.073
	6	0.734	0.104	1.538	0.103	1.891	0.117
	7			1.059	0.117	1.429	0.071
	8			1.400	0.098	1.718	0.090
	9			1.066	0.101	1.438	0.068
	10			1.311	0.100	1.623	0.082
	11			1.096	0.103	1.462	0.054
	12			1.356	0.098	1.665	0.084
	13			1.334	0.074	1.676	0.072
	14					1.669	0.093
	15					1.788	0.097
	16					1.819	0.100
	17					1.677	0.066
	18					1.931	0.125
	19					1.775	0.097
	20					1.939	0.142
	21					1.798	0.089
	22					1.922	0.113
	23					1.752	0.074
	24					1.790	0.096

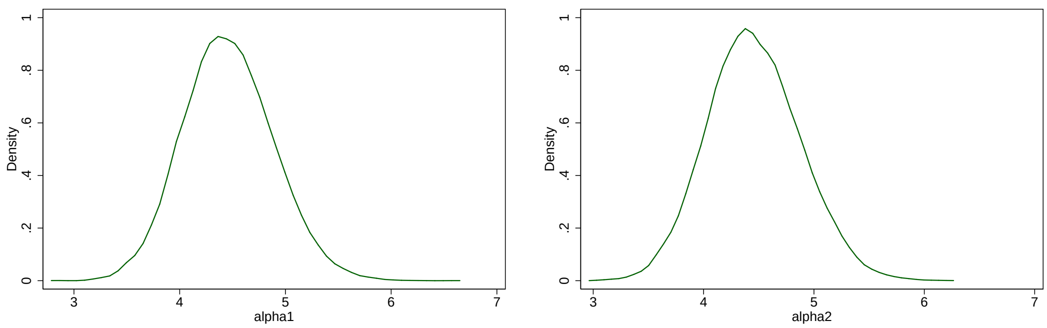
$N = 15,491$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

FIGURE OA.2. Competing Risk Model, Posterior Density of the Dirichlet Process Concentration Parameter α, Type 1 Risk (left) and Type 2 Risk (right), GIG mixture

$T = 6$



$T = 13$



$T = 24$

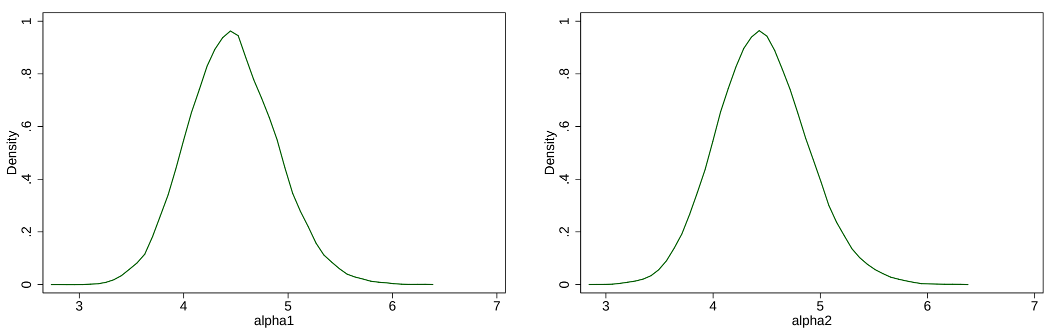


TABLE OA.3. Competing Risk Model without Individual Heterogeneity

		6 periods				13 periods				24 periods			
		Risk 1		Risk 2		Risk 1		Risk 2		Risk 1		Risk 2	
		Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
Urate		-0.295	0.061	-0.305	0.081	-0.199	0.045	-0.116	0.067	-0.199	0.037	-0.095	0.055
Race		-0.003	0.220	-0.044	0.422	-0.083	0.181	-0.329	0.312	-0.035	0.136	-0.289	0.215
Age		-0.155	0.175	-0.807	0.429	-0.097	0.135	-0.596	0.271	-0.011	0.119	-0.437	0.200
Rrate		-1.461	0.265	-0.865	0.506	-1.092	0.222	-0.407	0.302	-0.330	0.138	-0.374	0.226
t	1	-1.312	0.343	-3.434	0.644	-1.928	0.302	-4.495	0.578	-2.402	0.255	-4.643	0.494
	2	-0.759	0.322	-3.210	0.609	-1.389	0.300	-4.288	0.540	-1.865	0.230	-4.426	0.462
	3	-1.118	0.339	-3.538	0.669	-1.739	0.323	-4.583	0.601	-2.210	0.250	-4.736	0.524
	4	-0.829	0.335	-2.236	0.549	-1.448	0.303	-3.320	0.459	-1.946	0.243	-3.442	0.354
	5	-1.285	0.363	-2.226	0.547	-1.898	0.329	-3.335	0.482	-2.387	0.263	-3.444	0.360
	6	-1.763	0.384	-2.058	0.527	-2.407	0.369	-3.146	0.445	-2.891	0.299	-3.294	0.352
	7					-2.297	0.358	-4.205	0.571	-2.779	0.303	-4.342	0.498
	8					-1.854	0.326	-2.922	0.452	-2.349	0.278	-3.031	0.336
	9					-3.146	0.443	-3.813	0.529	-3.587	0.405	-3.947	0.468
	10					-2.004	0.362	-3.792	0.538	-2.503	0.289	-3.951	0.445
	11					-2.671	0.384	-3.798	0.533	-3.147	0.332	-3.940	0.448
	12					-2.187	0.360	-3.741	0.536	-2.715	0.308	-3.875	0.421
	13					-2.287	0.357	-3.721	0.521	-2.753	0.318	-3.832	0.426
	14									-2.833	0.329	-3.683	0.438
	15									-2.592	0.329	-3.439	0.405
	16									-3.169	0.381	-3.764	0.434
	17									-2.241	0.277	-3.713	0.414
	18									-2.695	0.316	-4.388	0.571
	19									-2.681	0.318	-3.145	0.393
	20									-2.831	0.341	-3.763	0.460
	21									-2.020	0.275	-3.467	0.400
	22									-4.002	0.576	-4.664	0.660
	23									-2.307	0.308	-3.651	0.455
	24									-3.277	0.433	-3.600	0.449

$N = 1,243$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

Table OA.4. Competing Risk Model with Parametric Gamma Heterogeneity

		6 periods				13 periods				24 periods			
		Risk 1		Risk 2		Risk 1		Risk 2		Risk 1		Risk 2	
		Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
γ		0.480	0.112	0.240	0.091	0.420	0.061	0.128	0.028	0.350	0.051	0.207	0.045
Urate		-0.361	0.060	-0.327	0.106	-0.298	0.058	-0.199	0.080	-0.342	0.054	-0.167	0.071
Race		0.024	0.272	-0.160	0.440	-0.130	0.236	-0.541	0.416	-0.067	0.217	-0.330	0.296
Age		-0.052	0.212	-0.766	0.386	0.002	0.197	-1.072	0.398	0.089	0.192	-0.723	0.291
Rrate		-1.763	0.257	-1.141	0.536	-1.434	0.235	-0.737	0.401	-0.957	0.223	-0.507	0.333
t	1	-0.844	0.321	-3.151	0.647	-1.278	0.306	-3.846	0.539	-1.381	0.296	-4.172	0.466
	2	-0.171	0.287	-2.907	0.604	-0.605	0.295	-3.578	0.490	-0.678	0.269	-3.973	0.474
	3	-0.401	0.285	-3.171	0.577	-0.825	0.304	-3.782	0.537	-0.876	0.319	-4.190	0.682
	4	-0.009	0.265	-1.792	0.487	-0.431	0.282	-2.427	0.426	-0.442	0.297	-2.891	0.373
	5	-0.390	0.280	-1.719	0.463	-0.798	0.321	-2.309	0.435	-0.791	0.315	-2.799	0.399
	6	-0.759	0.227	-1.500	0.403	-1.205	0.340	-1.967	0.417	-1.192	0.371	-2.562	0.357
	7					-1.061	0.338	-2.909	0.531	-1.056	0.368	-3.507	0.420
	8					-0.566	0.292	-1.490	0.376	-0.526	0.321	-2.141	0.357
	9					-1.762	0.406	-2.309	0.443	-1.756	0.419	-3.000	0.418
	10					-0.607	0.318	-2.242	0.435	-0.585	0.362	-2.946	0.397
	11					-1.186	0.359	-2.065	0.465	-1.142	0.385	-2.746	0.390
	12					-0.693	0.328	-1.962	0.472	-0.677	0.346	-2.759	0.434
	13					-0.876	0.345	-2.073	0.325	-0.709	0.393	-2.719	0.415
	14									-0.667	0.397	-2.433	0.400
	15									-0.333	0.379	-2.105	0.393
	16									-0.913	0.445	-2.385	0.440
	17									0.109	0.361	-2.345	0.442
	18									-0.263	0.402	-2.869	0.508
	19									-0.204	0.415	-1.525	0.358
	20									-0.230	0.401	-2.216	0.456
	21									0.679	0.384	-1.709	0.389
	22									-1.313	0.748	-2.859	0.576
	23									0.469	0.408	-1.835	0.421
	24									-1.060	0.346	-1.594	0.318

$N = 1,243$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

Table OA.5. Competing Risk Model with Parametric GIG Heterogeneity

		6 periods				13 periods				24 periods			
		Risk 1		Risk 2		Risk 1		Risk 2		Risk 1		Risk 2	
		Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
κ		-1.117	0.043	-1.145	0.081	-1.121	0.036	-1.157	0.056	-1.134	0.029	-1.153	0.047
φ		2.715	0.122	2.794	0.227	2.726	0.101	2.827	0.159	2.762	0.083	2.817	0.132
Urate		-0.215	0.049	-0.259	0.077	-0.182	0.043	-0.265	0.057	-0.295	0.044	-0.198	0.052
Race		0.059	0.259	-0.104	0.401	-0.110	0.227	-0.472	0.342	-0.073	0.190	-0.346	0.226
Age		-0.026	0.192	-0.733	0.418	-0.056	0.162	-0.760	0.269	-0.039	0.142	-0.553	0.224
Rrate		-1.323	0.241	-0.797	0.423	-1.150	0.236	-0.631	0.318	-0.568	0.180	-0.473	0.202
t	1	-1.737	0.272	-3.399	0.351	-1.949	0.255	-3.695	0.504	-1.798	0.276	-4.022	0.452
	2	-1.101	0.255	-3.096	0.332	-1.316	0.249	-3.485	0.539	-1.158	0.261	-3.834	0.450
	3	-1.365	0.272	-3.271	0.356	-1.588	0.264	-3.657	0.491	-1.440	0.276	-4.107	0.622
	4	-1.030	0.264	-2.257	0.265	-1.243	0.259	-2.395	0.373	-1.093	0.265	-2.844	0.372
	5	-1.453	0.281	-2.228	0.281	-1.639	0.267	-2.397	0.359	-1.495	0.294	-2.849	0.346
	6	-2.208	0.286	-2.195	0.215	-2.104	0.302	-2.185	0.353	-1.931	0.303	-2.623	0.356
	7					-1.982	0.317	-3.306	0.546	-1.821	0.341	-3.631	0.479
	8					-1.527	0.280	-1.868	0.343	-1.354	0.284	-2.325	0.334
	9					-2.751	0.445	-2.821	0.418	-2.651	0.448	-3.230	0.414
	10					-1.627	0.297	-2.772	0.446	-1.464	0.312	-3.188	0.396
	11					-2.232	0.383	-2.711	0.505	-2.105	0.373	-3.102	0.473
	12					-1.768	0.320	-2.597	0.443	-1.632	0.335	-3.016	0.435
	13					-2.215	0.271	-2.285	0.309	-1.696	0.331	-3.053	0.396
	14									-1.694	0.360	-2.848	0.389
	15									-1.474	0.325	-2.640	0.410
	16									-1.998	0.372	-2.924	0.396
	17									-1.103	0.304	-2.870	0.482
	18									-1.508	0.369	-3.572	0.513
	19									-1.461	0.351	-2.223	0.367
	20									-1.638	0.352	-2.919	0.469
	21									-0.772	0.316	-2.481	0.394
	22									-2.598	0.526	-3.746	0.687
	23									-1.068	0.333	-2.695	0.404
	24									-2.241	0.358	-2.279	0.391

$N = 1,243$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

Table OA.6. Competing Risk Model with Independent Risks, GIG Mixture

		6 periods				13 periods				24 periods			
		Risk 1		Risk 2		Risk 1		Risk 2		Risk 1		Risk 2	
		Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
κ		-1.321	0.046	-1.015	0.058	-1.444	0.045	-1.314	0.056	-1.550	0.041	-1.466	0.056
φ		3.288	0.129	2.430	0.163	3.634	0.126	3.270	0.159	3.933	0.115	3.698	0.160
Urate		-0.153	0.048	-0.210	0.072	-0.120	0.042	-0.014	0.059	-0.170	0.037	-0.160	0.052
Race		0.011	0.220	-0.031	0.406	-0.085	0.177	-0.203	0.329	-0.066	0.145	-0.305	0.243
Age		-0.187	0.174	-0.487	0.398	-0.208	0.134	-0.255	0.277	-0.246	0.113	-0.174	0.204
Rrate		-1.252	0.233	-0.517	0.371	-1.067	0.189	-0.226	0.261	-0.494	0.145	-0.151	0.197
t	1	-1.526	0.369	-4.443	0.383	-1.759	0.257	-3.392	0.484	-1.873	0.246	-3.656	0.487
	2	-0.951	0.355	-4.176	0.356	-1.185	0.236	-3.212	0.458	-1.302	0.223	-3.478	0.461
	3	-1.263	0.370	-4.408	0.387	-1.499	0.253	-3.556	0.520	-1.621	0.242	-3.829	0.528
	4	-0.947	0.368	-3.453	0.302	-1.187	0.245	-2.309	0.357	-1.313	0.233	-2.572	0.354
	5	-1.384	0.392	-3.486	0.307	-1.624	0.273	-2.339	0.365	-1.751	0.259	-2.603	0.357
	6	-2.524	0.352	-3.846	0.319	-2.085	0.309	-2.168	0.350	-2.220	0.305	-2.435	0.346
	7					-1.988	0.310	-3.229	0.489	-2.121	0.297	-3.509	0.493
	8					-1.543	0.276	-1.938	0.335	-1.673	0.266	-2.207	0.338
	9					-2.795	0.412	-2.857	0.435	-2.935	0.404	-3.128	0.437
	10					-1.665	0.289	-2.861	0.449	-1.806	0.278	-3.114	0.435
	11					-2.297	0.355	-2.825	0.440	-2.435	0.349	-3.086	0.439
	12					-1.835	0.306	-2.801	0.436	-1.982	0.301	-3.067	0.438
	13					-2.661	0.349	-2.522	0.366	-2.060	0.314	-3.055	0.441
	14									-2.088	0.322	-2.904	0.420
	15									-1.855	0.300	-2.666	0.385
	16									-2.424	0.370	-2.976	0.437
	17									-1.469	0.275	-2.954	0.434
	18									-1.893	0.317	-3.703	0.586
	19									-1.842	0.320	-2.400	0.371
	20									-2.031	0.346	-3.018	0.455
	21									-1.170	0.267	-2.725	0.417
	22									-3.087	0.566	-3.929	0.661
	23									-1.434	0.299	-2.952	0.456
	24									-2.777	0.242	-2.690	0.461

$N = 1,243$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

Table OA.7. Single Risk Model with Competing Risk Data, GIG mixture

		6 periods		13 periods		24 periods	
		Mean	s.e.	Mean	s.e.	Mean	s.e.
κ		-1.230	0.035	-1.417	0.035	-1.534	0.037
φ		3.032	0.101	3.560	0.099	3.887	0.105
Urate		-0.150	0.038	-0.136	0.034	-0.167	0.031
Race		0.035	0.192	-0.139	0.157	-0.117	0.122
Age		-0.153	0.157	-0.183	0.125	-0.136	0.099
Rrate		-0.998	0.204	-0.765	0.155	-0.303	0.116
t	1	-2.075	0.225	-2.242	0.220	-2.395	0.211
	2	-1.567	0.206	-1.740	0.203	-1.894	0.190
	3	-1.907	0.224	-2.084	0.221	-2.240	0.213
	4	-1.463	0.208	-1.639	0.203	-1.798	0.194
	5	-1.789	0.224	-1.966	0.222	-2.127	0.219
	6	-2.416	0.239	-2.176	0.236	-2.335	0.226
	7			-2.477	0.258	-2.640	0.251
	8			-1.775	0.218	-1.937	0.209
	9			-2.862	0.299	-3.024	0.293
	10			-2.157	0.243	-2.324	0.234
	11			-2.571	0.276	-2.734	0.269
	12			-2.274	0.255	-2.444	0.249
	13			-2.636	0.237	-2.493	0.254
	14					-2.464	0.252
	15					-2.238	0.239
	16					-2.693	0.279
	17					-2.068	0.230
	18					-2.561	0.274
	19					-2.128	0.239
	20					-2.487	0.272
	21					-1.818	0.224
	22					-3.448	0.408
	23					-2.086	0.244
	24					-2.761	0.238

$N = 1,243$, *Urate* denotes the state unemployment rate, *Rrate* denotes the replacement rate.

4. Extended Counterfactual Policy Experiment

In the main text we have reported the results of a counterfactual policy experiment whereby we simulated a change in the replacement rate and estimated its impact on the probability of exit from unemployment as captured by the survival function. Here we further provide the details on two extensions of the counterfactual experiment: first estimating potential differences between the policy impact on individuals with different unobserved heterogeneities, and second estimating the impact of varying the changes of the replacement rate over time. A summary of the findings is presented in the main text.

4.1. Counterfactuals for Split Samples Based on Unobserved Heterogeneity

The unobserved heterogeneity term v_i can be interpreted as a factor which also contributes to the variation in the hazard rates but is not included among the observed explanatory variables and instead inferred indirectly from the model. One of the key advantages of estimating v_i is that it enables us to differentiate among various groups of individuals based on their unobserved qualities. In our model specification, increasing v_i increases the cumulative hazard function and hence decreases the survival function of unemployment. Thus, individuals with higher v_i have better chances exiting unemployment faster, while individuals with lower v_i are more likely to be long term unemployed. It is difficult to interpret the exact meaning of the unobserved individual component. Nonetheless, given the way it influences the hazard function, v_i can perhaps be thought of as individual ability or quality of labor market characteristics.

As the MCMC output we obtained a Markov chain of draws for each v_i. Denote its mean by $\overline{v}_i$ and the median of the individual means by v_{med}. For both single risk and competing risk model, we split the sample into two parts: one for individuals with $\overline{v}_i \leq v_{med}$ (label these as "low type") and individuals with $\overline{v}_i > v_{med}$ (label these as "high type"). We then ran the counterfactual experiment changing the replacement rate by 10% for each subsample separately, for the case $T = 24$. The resulting % change of the survival function are reported in Table OA.8 (single risk) and Table OA.9 (competing risks) below. In each model, high type individuals react more to the replacement rate changes than low type individuals, for either direction of the change. In the single risk model, the relative ratio of the survival function changes of high type to low type individuals is just over 20%, while in the competing risk case the corresponding figure is approximately 15%. This finding is consistent with the literature estimating the policy effect of training and job placement effects.[10]

[10] We would like to thank an anonymous referee for pointing this out.

Table OA.8. % Change in Survival Function, Single Risk, GIG mixture, T=24

	Pooled		High Type		Low Type	
t	down	up	down	up	down	up
1	-0.052	0.045	-0.057	0.050	-0.047	0.041
2	-0.128	0.111	-0.140	0.121	-0.117	0.101
3	-0.182	0.158	-0.200	0.174	-0.167	0.145
4	-0.255	0.221	-0.282	0.245	-0.234	0.203
5	-0.305	0.265	-0.340	0.295	-0.281	0.244
6	-0.377	0.328	-0.421	0.366	-0.348	0.302
7	-0.421	0.366	-0.477	0.415	-0.388	0.338
8	-0.480	0.418	-0.548	0.477	-0.443	0.386
9	-0.521	0.453	-0.601	0.523	-0.481	0.419
10	-0.570	0.497	-0.662	0.576	-0.528	0.460
11	-0.609	0.531	-0.713	0.621	-0.566	0.493
12	-0.657	0.573	-0.771	0.671	-0.613	0.534
13	-0.701	0.612	-0.828	0.722	-0.657	0.574
14	-0.746	0.651	-0.879	0.766	-0.705	0.615
15	-0.794	0.693	-0.940	0.820	-0.754	0.658
16	-0.838	0.732	-0.987	0.861	-0.802	0.701
17	-0.869	0.760	-1.018	0.889	-0.840	0.734
18	-0.919	0.803	-1.076	0.940	-0.892	0.780
19	-0.953	0.834	-1.107	0.967	-0.932	0.816
20	-0.998	0.874	-1.154	1.009	-0.981	0.859
21	-1.033	0.905	-1.203	1.052	-1.019	0.893
22	-1.071	0.938	-1.225	1.072	-1.062	0.930
23	-1.102	0.966	-1.292	1.132	-1.096	0.960
24	-1.187	1.041	-1.387	1.215	-1.183	1.038

”down” denotes counterfactual decrease of the replacement rate by 10%, and ”up” denotes increase by 10%.

Table OA.9. % Change in Survival Function, Competing Risks, GIG mixture, T=24

	Pooled		High Type		Low Type	
t	down	up	down	up	down	up
1	-0.141	0.120	-0.154	0.131	-0.109	0.093
2	-0.370	0.316	-0.406	0.347	-0.286	0.244
3	-0.522	0.446	-0.567	0.485	-0.411	0.351
4	-0.750	0.642	-0.817	0.701	-0.601	0.514
5	-0.896	0.768	-0.988	0.848	-0.735	0.630
6	-1.015	0.870	-1.126	0.967	-0.842	0.721
7	-1.103	0.946	-1.228	1.055	-0.924	0.792
8	-1.274	1.095	-1.417	1.220	-1.081	0.929
9	-1.327	1.141	-1.481	1.275	-1.135	0.975
10	-1.450	1.248	-1.633	1.408	-1.246	1.071
11	-1.526	1.314	-1.722	1.485	-1.317	1.134
12	-1.630	1.404	-1.835	1.584	-1.414	1.218
13	-1.721	1.484	-1.936	1.672	-1.506	1.298
14	-1.816	1.568	-2.084	1.802	-1.600	1.380
15	-1.936	1.673	-2.245	1.943	-1.717	1.483
16	-1.997	1.726	-2.318	2.007	-1.792	1.548
17	-2.130	1.844	-2.472	2.143	-1.933	1.672
18	-2.195	1.901	-2.506	2.173	-2.020	1.749
19	-2.299	1.994	-2.614	2.269	-2.148	1.862
20	-2.371	2.058	-2.735	2.378	-2.239	1.943
21	-2.538	2.206	-2.990	2.604	-2.417	2.101
22	-2.526	2.196	-2.874	2.499	-2.453	2.132
23	-2.664	2.318	-3.076	2.678	-2.588	2.253
24	-2.686	2.338	-3.101	2.717	-2.641	2.300

"down" denotes counterfactual decrease of the replacement rate by 10%, and "up" denotes increase by 10%.

4.2. Counterfactuals for Time-varying Changes in the Replacement Rate

In this section we further explore two additional scenarios for the counterfactual policy change: first, a sharply declining replacement rate at the beginning of the spell, and second a scenario where the rate declines sharply only at the end of the spell. For this purpose we construct a decreasing function $r : [0,1] \to [0,1]$ defining the factor $r(t/T)$ by which we multiply the original replacement rate in each time period t with T being the final period. Let $u \in [0,1]$. In the first scenario,

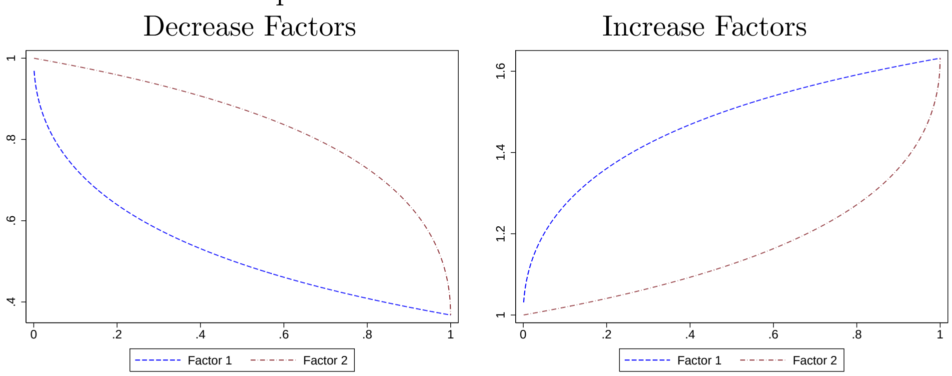
$$r_1 = \exp\left(-u^{\delta}\right)$$

and in the second scenario,

$$r_2 = 1 + \exp(-1) - \exp(-(1-u)^{\delta})$$

A mirror image of r_1 and r_2 increasing from 1 is also used for a counterfactual increase of the replacement rate. It is important to note that at the end of the observation time window both factors become equal, $r_1(1) = r_2(1)$. The constant $\delta \in (0,1)$ controls the degree of curvature within the exponential change of r_1 and r_2, which smaller δ yielding shaper curvature. We set $\delta = 1/2$, resulting in r_1 and r_2 as shown in Figure OA.3.

FIGURE OA.3. Replacement Rate Counterfactual Decrease Factors



The results are presented in Table OA.10. In both SR and CR models, factor 1 (sharp initial change of the replacement rate) leads to an overall larger change in the survival function than factor 2 (sharp change towards the end of the observation time period). In the SR model the change for factor 1 relative to factor 2 at $T = 24$ is more than twofold, and in the CR model it is close to threefold. This indicates that on average individuals respond more to incentives provided early in their unemployment spells relative to ones provided later, even if the response is still overall inelastic.

Table OA.10. % Change in Survival Function, Time-varying Replacement Rate, GIG mixture, T=24

	Single Risk				Competing Risk			
	Factor 1		Factor 2		Factor 1		Factor 2	
t	down	up	down	up	down	up	down	up
1	-0.055	0.100	-0.006	0.029	-0.264	0.219	-0.010	0.009
2	-0.253	0.220	-0.011	0.014	-0.855	0.702	-0.047	0.041
3	-0.428	0.319	-0.052	0.022	-1.330	1.085	-0.085	0.075
4	-0.648	0.505	-0.074	0.018	-2.133	1.732	-0.164	0.143
5	-0.816	0.645	-0.094	0.034	-2.708	2.194	-0.231	0.201
6	-1.117	0.823	-0.173	0.025	-3.210	2.594	-0.297	0.259
7	-1.256	0.987	-0.171	0.070	-3.608	2.913	-0.358	0.311
8	-1.454	1.215	-0.175	0.136	-4.419	3.565	-0.495	0.430
9	-1.603	1.375	-0.187	0.181	-4.685	3.777	-0.546	0.474
10	-1.855	1.527	-0.268	0.189	-5.318	4.288	-0.678	0.586
11	-2.025	1.686	-0.301	0.232	-5.725	4.612	-0.771	0.666
12	-2.239	1.885	-0.348	0.291	-6.297	5.073	-0.916	0.790
13	-2.442	2.082	-0.395	0.356	-6.823	5.500	-1.065	0.916
14	-2.659	2.279	-0.459	0.422	-7.385	5.957	-1.238	1.063
15	-2.853	2.534	-0.491	0.538	-8.104	6.546	-1.479	1.266
16	-3.067	2.759	-0.558	0.638	-8.502	6.869	-1.636	1.398
17	-3.263	2.898	-0.653	0.687	-9.342	7.574	-1.986	1.691
18	-3.476	3.184	-0.717	0.844	-9.795	7.953	-2.209	1.877
19	-3.616	3.411	-0.762	0.985	-10.499	8.551	-2.578	2.184
20	-3.783	3.714	-0.815	1.187	-10.989	8.973	-2.870	2.425
21	-3.959	3.926	-0.916	1.330	-12.090	9.921	-3.557	2.989
22	-4.175	4.136	-1.070	1.481	-12.099	9.930	-3.645	3.059
23	-4.423	4.253	-1.290	1.568	-12.999	10.713	-4.368	3.637
24	-5.311	4.369	-2.260	1.725	-13.205	10.880	-4.720	3.895

"down" denotes counterfactual time-varying decrease of the replacement rate, and "up" denotes time-varying increase.