Semi-Parametric Proportional Hazards Models with Crossed Random Effects for Psychometric Response Times

Abstract

The semi-parametric proportional hazards model with crossed random effects shares two important characteristics: it avoids explicit specification of the response time distribution by using semi-parametric models, and heterogeneity that is due to subjects and items is captured. The proposed model has a proportionality parameter for the speed of each test taker, for the time intensity of each item, and for subject or item characteristics of interest. It is shown how all these parameters can be estimated by Markov Chain Monte Carlo methods (Gibbs sampling). The performance of the estimation procedure is assessed with simulations and the model is further illustrated with the analysis of response times from a visual recognition task.

Keywords: Bayesian Estimation, Crossed Random Effects, Frailty Model, Response Time, Semi-Parametric Proportional Hazards Model

1 Introduction

Over the last 2 decades modeling of response time has seen rapid growth in the psychometric literature. Van der Linden (2009) classified existing response time models into two distinct categories based on the approaches those models have. The first category models response times in the framework of an item response theory (IRT) model, where responses are incorporated in the reaction time models or vice versa. The second category consists of distinct models for response time and the examinees responses. In this paper we build on the latter tradition, and focus more specifically on the model for the response time. A large variety of distributional assumptions for the response time have been proposed in the literature so far. An important class of parametric models for reaction times that is frequently used assumes the lognormal distribution, which was originally proposed in the literature by Furneaux (1952) and introduced in the psychometric literature by Thissen (1983). Van Breukelen (2005) and van der Linden (2006) amongst others further elaborated on this lognormal model. More recently a Box-Cox-normal model for response time has been proposed (Klein Entink, van der Linden & Fox, 2009). Rouder, Sun, Speckmanm, Lu and Zhou (2003) alternatively suggested the shifted 3-parameter Weibull distribution. Each of these distributions can be criticized for not exactly capturing the response time: the absence of a shift parameter in the lognormal models, the heavier tails than expected under the Weibull, ... (Rouder, Tuerlinckx, Speckman, Lu & Gomez, 2008). In this paper, we prefer to make progress within the *semi-parametric proportional hazards framework*, as the latter avoids the need to specify the distribution of the response times.

Bloxom (1985) was amongst the first to introduce the concept of *hazard function* in a psychometric context. To introduce the concept of the hazard, suppose that response times are observations of a random variable, which has probability density function $f(t)$. The hazard function $h(t)$ of this distribution can be defined as follows. Let the probability of response in a small (non-negative) interval of time Δt immediately following t denoted by $\Pr(t \leq T \leq t + \Delta t)$. The conditional probability of response in this interval, given that the response did not occur prior to t is $\Pr(t \leq T \leq t+\Delta t \mid T \geq t)$. If we divide this probability by Δt and pass it to the limit $\Delta t \to 0$, we obtain the ratio $h(t) = f(t)/S(t)$, where $S(t)$ is the survival function, indicating the probability that a response will not be given by time t. This is known as the hazard function. The cumulative hazard function $H(t) = \int_0^t h(u)du$ is linked to the survival function as follows, $S(t) = \exp(-H(t))$.

Loosely speaking, the hazard function expresses the likelihood of a participant providing a response in the next instant, given that he/she had not yet done so. Hence the hazard function is able to capture the instantaneous *capacity* or *speed* of the test taker to respond. A test taker with a high capacity has a higher conditional probability of responding (Wenger & Gibson, 2004). Similarly, we can view the hazard in terms of the *intensity* that an item requires to be responded. Easier items may need less processing time and have a larger hazard compared to more difficult items. With Wenger and Gibson (2004) we acknowledge the conceptual advantage of the hazard compared to other typical statistics on reaction times such as the mean for example. Indeed, the hazard provides information about the speed and intensity at any time t, in contrast to the mean response time that only provides an expectation on the response time.

The *proportional hazards model* assumes in its most general form that the specific hazard of each subject and item combination, i.e. the hazard $h_{ij}(t)$ for subject i and item j, is proportional to some unknown baseline hazard function $h_0(t)$. More specifically the hazard $h_{ij}(t)$ equals $\phi_{ij}h_0(t)$ where ϕ_{ij} is a positive scalar parameter. In *parametric* proportional hazards models, a distributional assumption is made on the response time, which leads to a fully parametric specification of $h_0(t)$ and thus of the model for $h_{ij}(t)$. Scheiblechner (1985) proposed in the early 80's the exponential density for response time. More specifically he assumed the baseline hazard to be constant and ϕ_{ij} to be an exponential function of the sum of a person parameter θ_i and an item parameter ϵ_j. A major drawback of such parametric models is that misspecification of the distribution of response times may lead to invalid inference; a misspecified baseline hazard causes all parameter estimates to be inconsistent (Ridder, 1987). Misspecification of the distribution is less of an issue in *semi-parametric* proportional hazards (also called the Cox PH-model). Indeed, in the latter model, the baseline hazard $h_0(t)$ is left unspecified.

The use of semi-parametric proportional hazards models for reaction times from psychological experiments is not entirely new. Van Breukelen (2005) and Wenger and Gibson (2004) independently proposed some variant on the Cox PH-model. Using a stratified partial likelihood approach (Allison, 1996) these authors allow for unobserved heterogeneity across participants and dependence among items within subjects. However, using their model potentially discards a considerable amount of information as no comparisons between subjects can be made and the item covariate estimates are based solely on within-subject comparisons. Moreover, as their model is stratified for subjects, it considers in a sense subject heterogeneity as a nuisance and hence does not allow to assess heterogeneity between subjects. Rather than treating subjects as strata and item effects as fixed like these authors, we would like to follow a recent trend in psychometry and view both subjects and items as random samples from a larger population. If both items and subjects are considered to be random, our target PH-model should therefore include both random subject and random item effects. Treating subjects as random effects is well established because it is reasonable to assume that no combination of observed subject covariates will be able to explain all the variance in response times between people. The same argument can be invoked for items; it is not likely that a set of item predictors can explain all the variance in response times that exists between test items. Since items are not nested within subjects and subjects are not nested within items (figure 1), such model with a random effect for subject and item is often called a *crossed random effects* model (Raudenbusch, 1993).

Semi-parametric PH-models with a random effect, commonly known as 'frailty models' in the field of survival analysis, were first introduced by Vaupel, Manton and Stallard (1979), and Clayton and Cuzick (1985) extended the model to allow for covariates. While the biostatistical literature has seen major advances over the last two decades, these models only recently received attention in the psychometric field. An early application of frailty models and notable exception in the psychometric literature was proposed by Douglas, Kosorok and Chewning (1999) when dealing with discrete response times. These authors consider a subject-specific latent psychological construct, assume independence between response times conditional on this unobservable frailty, and study the performance of the Metropolis-Hasting algorithm to estimate the ability of an item to distinguish among subjects with varying levels of that latent psychological construct. Ranger and Ortner (2012) proposed a profile likelihood approach that can circumvent the limitation of discrete response times for the latter model. While these authors used a marginalized maximum likelihood framework, Wang, Fan, Chang and Douglas (2013) recently described a more flexible Markov chain Monte Carlo framework.

The major difficulty with fitting frailty models is the often complicated integration of the likelihood over the random effect. While fitting (semi-)parametric PH-models with a single frailty has been extensively studied (for an overview see Ibrahim, Chen & Sinha, 2005; Duchateau & Janssen, 2008), simultaneous estimation of two or more random effects (one random effect for subject and one for item in our case) is a more challenging task, especially in the frequentist framework. Duchateau and Janssen (2008) for example distinguish two different cases for frailty models with more than one frailty: (i) different frailty terms occur within the same cluster (illustrated by Legrand, Ducrocq, Janssen, Sylvester & Duchateau, 2005), and (ii) the frailty terms are nested (Shih & Lu, 2009), but these authors focus their discussion to the fully parametric case. Estimation of crossed random effects in frailty models however has to our knowledge not been studied yet and will further be developed in this paper.

2 Bayesian estimation in a semi-parametric proportional hazards framework with crossed random effects

In this paper we elaborate on the following frailty model for response time T_{ij} from subject i ($i = 1, \ldots, N$) on item j ($j = 1, \ldots, k$):

$$h_{ij}(t) = h_0(t) \exp\left(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j}\right), \tag{1}$$

with the baseline hazard $h_0(t)$ left unspecified. In model (1), $\mathbf{x}_{ij}^t$ is a vector of subject- and item-specific covariates of interest. In the visual recognition study that we will discuss later in the illustration, one was interested in knowing whether neutral scenes are recognized faster then pleasant scenes. The item-specific covariate in model (1) then reflects each of these conditions. Alternatively, the effect of subject-specific characteristics like age might be of interest. Further, v_{1i} is a subject-specific random effect and v_{2j} an item-specific random effect in model (1). Conditional on these random effects, response times T_{ij} are assumed to be independent. In a crossed random effect setting where items are not nested within subjects or vice versa (figure 1), these random effects are assumed to be independent. Finally, the additional assumption is made that - conditional on measured covariates $\mathbf{x}_{ij}^t$ - there is a consistent item effect across subjects and vice versa.

We can rewrite (1) as

$$h_{ij}(t) = h_0(t)u_{1i}u_{2j}\exp\left(\mathbf{x}_{ij}^t\boldsymbol{\beta}\right) \tag{2}$$

where $u_{1i} = \exp(v_{1i})$ and $u_{2j} = \exp(v_{2j})$ subject- and item-specific frailties. The frailties u_{1i} and u_{2j} are interpreted as representing all unmeasured subject- and item-factors which affect the subject- and item-specific hazard, and assumed to capture the unobserved heterogeneity in response times between subjects and items respectively. As mentioned in the introduction, the hazard function can intuitively be viewed as the likelihood of a subject i completing item j, given that the subject has not yet completed the item. A fast responder has a high conditional probability of finishing the item, and will have a large value for v_{1i} (or u_{1i}) in model (1) (model (2) respectively). In other words, the value of v_{1i} allows to assess the speed of the test taker relative to its peers with the same measured characteristics. Similarly, an item that is more easily accomplished will have a high conditional probability of being finished, and hence have a large value for v_{2j} (or u_{2j}), reflecting the time intensity of the time relative to its alikes. Within the spirit of the proportional hazards framework, it is important to realize that - while we allow for different hazards for each item and subject - these hazards are restricted to be shifted proportionally from $h_0(t)$ by an item-specific and individual-specific component. Assuming such proportionality may not be realistic in all circumstances as item response distributions can differ considerably even in similar tests (Ranger & Kuhn, 2012). Other recent proposals in the semi-parametric PH-framework (Ranger and Ortner, 2012; Wang et al., 2013), specify the hazard as follows

$$h_{ij}(t) = h_{0j}(t) \exp(\beta_j v_{1i}),$$

where similar to model (2), v_{1i} can be viewed as a speed parameter. While the latter model allows for a different functional form for the hazard for each item and a discrimination parameter (β_j), the stratified baseline hazard would not allow to quantify the effect of an item-specific characteristic x_{ij} (neutral versus pleasant pictures for example) on the response time. In contrast, we aim to propose in this paper a model that allows to estimate such effects while acknowledging the heterogeneity that is both due to the subjects and items. To reach that goal, we require that the effect of each item j is to shift a common baseline hazard $h_0(t)$ with a factor $\exp(v_{2j})$. Such model is an important improvement over existing fully parametric counterparts. Indeed the semi-parametric PH-model (2) leaves the baseline hazard $h_0(t)$ unspecified, and hence allows for a wider range of true underlying response time distributions.

Distributional choices have to be made for the random effects. Typical choices are that v_{1i} (and v_{2j}) are an independent and identically distributed (i.i.d) sample from either a normal density with mean 0 and variance σ_1^2 (σ_2^2 respectively), or that u_{1i}

(and u_{2j}) are from a one-parameter gamma density with mean one and variance γ_1 (γ_2). The mean of the random effects (frailties) is set to zero (one, respectively) to allow for identifiability. In this paper, we will proceed with zero-mean normally distributed subject and item random effects. The parameters σ_1^2 and σ_2^2 can be viewed as heterogeneity parameters, and allow to assess the variability between subjects and items, respectively (Legrand et al., 2005).

We will use Bayesian techniques to fit the semi-parametric model (1) and follow the Markov Chain Monte Carlo-approach suggested by Kalbfleisch (1978) for semi-parametric Cox model and further extended for frailty models by Clayton (1991). These authors tackle the problem of an unspecified baseline hazard $h_0(t)$ by proposing an independent-increments gamma prior for the cumulative baseline hazard $H_0(t)$, with $H_0(t) = \int_0^t h_0(u)du$. Ibrahim, Chen & Sinha (2005, p.47) describe several alternative nonparametric prior processes for the baseline cumulative hazard (the beta process, correlated prior processes, the Dirichlet process, ...) but here the independent-increments gamma approach was choosen because of its ease of implementation in available software. Under this independent-increments approach, the hazard at each observed response time is considered a parameter, and so - with many different response times typically observed - we end up with a very high number of parameters to be estimated. To avoid high-dimensional sampling associated with the Metropolis algorithm, we will therefore rely on the more efficient Gibbs sampling, which is based on the posterior density of each parameter, conditional on all the other parameters. In this section we demonstrate how Gibbs sampling can be applied to obtain posterior densities for the parameters of interest in the semi-parametric frailty model (1) with zero-mean normal random effects. The approach outlined below builds further on the Bayesian estimation in the semi-parametric PH-model with one gamma frailty described in Duchateau and Janssen (2008, p.233-245).

A Bayesian estimation approach

We first order all response times t_{ij}, and partition the time axis in z (with z equal to the total number of *distinct* response times) disjoint intervals $(t_{(0)}, t_{(1)}], \ldots,$ $(t_{(z-1)}, t_{(z)}]$ with $t_{(0)} = 0$ and $t_{(m)}$ corresponding to the m-th ordered reaction time. Further denote the increase of the cumulative baseline hazard in interval $(t_{(m-1)}, t_{(m)}]$ by $h_{(m)}$. The probability that subject i gives a response to item j in the interval $(t_{(m-1)}, t_{(m)}]$ equals $\Pr(T_{ij} < t_{(m)}) - \Pr(T_{ij} < t_{(m-1)}) = \exp\left(-\int_0^{t_{(m-1)}} h_{ij}(u)du\right) - \exp\left(-\int_0^{t_{(m)}} h_{ij}(u)du\right)$, where the latter equality follows from $S(t) = \exp(-H(t))$. Viewing the reaction times as grouped within these intervals (where the number of intervals is equal to the number of observations if all reaction times are different), the grouped-data likelihood function corresponding to model (1) conditional on fixed and random effect equals

$$\prod_{i=1}^{N}\prod_{j=1}^{k} \exp\left(-\exp(\mathbf{x}_{ij}^t\boldsymbol{\beta}+v_{1i}+v_{2j})\sum_{m:t_{(m)}<t_{ij}} h_{(m)}\right)\times \left[1-\exp\left(-\exp(\mathbf{x}_{ij}^t\boldsymbol{\beta}+v_{1i}+v_{2j})h_{(s,ij)}\right)\right] \tag{3}$$

with $s,ij = \min\{m : t_{(m)} \geq t_{ij}\}$.

Using first-order Taylor expansion (see appendix A1), this expression can be approximated by

$$\prod_{i=1}^{N}\prod_{j=1}^{k} \exp\left(-\exp(\mathbf{x}_{ij}^t\boldsymbol{\beta}+v_{1i}+v_{2j})\sum_{m:t_{(m)}\leq t_{ij}} h_{(m)}\right)\left(\exp(\mathbf{x}_{ij}^t\boldsymbol{\beta}+v_{1i}+v_{2j})h_{(s,ij)}\right) \tag{4}$$

or alternatively

$$\prod_{i=1}^{N}\prod_{j=1}^{k}\prod_{m:t_{(m)}\leq t_{ij}} \left(h_{(m)}\exp(\mathbf{x}_{ij}^t\boldsymbol{\beta}+v_{1i}+v_{2j})\right)^{\delta_{ij}(t_{(m)})} \exp\left(-h_{(m)}\exp(\mathbf{x}_{ij}^t\boldsymbol{\beta}+v_{1i}+v_{2j})\right) \tag{5}$$

with $\delta_{ij}(t_{(m)})$ equal to 1 if $t_{ij} = t_{(m)}$, and else 0.

Expression (5) resembles the likelihood of a Poisson regression analysis with likelihood as if the indicators $\delta_{ij}(t_{(m)})$ were Poisson random variables with expectation $h_{(m)}\exp(\mathbf{x}_{ij}^t\boldsymbol{\beta}+v_{1i}+v_{2j})$. Since the conjugate prior for the Poisson mean is the gamma distribution, it is therefore convenient if $H_0(t)$ is assumed to follow a process in which the increments $h_{(m)}$ are distributed according to gamma distributions

(Kalbfleisch, 1978). The likelihood formulation (5) is also the one that is used in OPENBUGS (Lunn, Spiegelhalter, Thomas & Best, 2009), a free software package for performing Bayesian inference using Gibbs sampling, that we will use for the illustration too. Before we demonstrate how model (5) and specific prior choices lead to conditional posterior densities describing the Gibbs sampler which are either known distributions or shown to be log concave, we first specify these priors.

Prior distributions

As motivated above, we assume an independent gamma process prior for the cumulative baseline hazard. More specifically, we have that the increments $h_{(m)}$ are distributed as independent gamma variables with shape parameter $c\left(H_0^*(t_{(m)})-H_0^*(t_{(m-1)})\right)$ and scale parameter c respectively, with the Gamma distribution with shape parameter r and scale parameter μ defined in OPENBUGS as

$$f(x) = \frac{\mu^r x^{r-1} e^{-\mu x}}{\Gamma(r)} \quad \text{if } x > 0$$

We need to specify values for both H_0^* and c. The function H_0^* is taken to be based on a time-constant hazard h_0^* and c then reflects the degree of confidence in this prior guess h_0^*. Small values of c correspond to large variances on the increments and hence to weak prior beliefs. In the simulation study and illustrating example below we set c equal to 0.0001 and the increments in the cumulative baseline hazard

$H_0^*(t_{(m)}) - H_0^*(t_{(m-1)})$ were set to $r\Delta t_{(m)}$, where r is a guess at the event rate (assumed to be equal to 1 here) per time interval, and $\Delta t_{(m)}$ is the size of the interval, $t_{(m)} - t_{(m-1)}$.

For the fixed effect parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ we assume uniform priors $f(\beta_l) \propto 1$ and independence between all parameters, implying that $f(\boldsymbol{\beta}) = \prod_{l=1}^{p} f(\beta_l) \propto 1$. For the precision of the normally distributed random effects v_{1i} and v_{2j}, $1/\sigma_1$ and $1/\sigma_2$ respectively, non-informative gamma distributions are assumed.

Posterior distribution

Following Bayes theorem, the joint posterior density function of the parameter vector $\boldsymbol{\omega}$ given the observed data $\mathbf{t}$, equals

$$f(\boldsymbol{\omega} \mid \mathbf{t}) = \frac{f(\mathbf{t} \mid \boldsymbol{\omega}) f(\boldsymbol{\omega})}{f(\mathbf{t})}$$

with $\boldsymbol{\omega} = (\mathbf{h}^t, \sigma_1, \sigma_2, \boldsymbol{\beta}^t, \mathbf{v}_1^t, \mathbf{v}_2^t)$, a $(z + 1 + 1 + p + N + k) \times 1$ vector with $\mathbf{h} = (h_{(1)}, \ldots, h_{(z)})^t$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, $\mathbf{v}_1 = (v_{11}, \ldots, v_{1N})^t$ and $\mathbf{v}_2 = (v_{21}, \ldots, v_{2k})^t$. To use Gibbs sampling, Bayes theorem can also be applied to one specific parameter ω_i while conditioning on the other parameters (denoted by $\boldsymbol{\omega}_{(-i)}$),

$$f(\omega_i \mid \mathbf{t}, \boldsymbol{\omega}_{(-i)}) = \frac{f(\mathbf{t} \mid \boldsymbol{\omega}) f(\omega_i \mid \boldsymbol{\omega}_{(-i)})}{f(\mathbf{t} \mid \boldsymbol{\omega}_{(-i)})}$$

Given the independence between prior densities of all parameters, we have

$$f(\omega_i \mid \mathbf{t}, \boldsymbol{\omega}_{(-i)}) \propto f(\mathbf{t} \mid \boldsymbol{\omega}) f(\omega_i), \tag{6}$$

where we further dropped the normalizing factor as it is difficult to obtain in this setting.

In the following we derive the conditional posterior densities for all the parameters of interest. We first consider the conditional posterior density of one component of $\boldsymbol{\beta}$, say β_l $(l = 1, \ldots, p)$. From (6) and the uniform prior for β_l, it follows that

$$\begin{aligned} f \quad & (\beta_l \mid \mathbf{t}, \mathbf{h}, \sigma_1, \sigma_2, \boldsymbol{\beta}_{(-l)}, \mathbf{v}_1, \mathbf{v}_2) \propto f(\mathbf{t} \mid \mathbf{h}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2) \\ & = \prod_{i=1}^{N} \prod_{j=1}^{k} \prod_{m: t_{(m)} \leq t_{ij}} \left(h_{(m)} \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j}) \right)^{\delta_{ij}(t_{(m)})} \\ & \qquad \exp\left(-h_{(m)} \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j}) \right) \end{aligned} \tag{7}$$

As this conditional posterior density is logconcave (see Appendix A1), it allows to make use of the adaptive rejection sampling algorithm (Gilks & Wild, 1992), a fact that is also used in OPENBUGS to generate a sample.

Next we consider $f(h_{(l)} \mid \mathbf{t}_{(l)}, \mathbf{h}_{(-l)}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2)$, the conditional posterior density of $h_{(l)}$ $(l = 1, \ldots, z)$, which equals

$$\frac{f(\mathbf{t}_{(l)} \mid \mathbf{h}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2) f(h_{(l)} \mid \mathbf{h}_{(-l)}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2)}{f(\mathbf{t}_{(l)} \mid \mathbf{h}_{(-l)}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2)}$$

with $\mathbf{t}_{(l)}$ referring to all reaction times equals to $t_{(l)}$. Given the assumed independence between all parameters and the independent increments in cumulative hazard, this expression simplifies to

$$f(h_{(l)} \mid \mathbf{t}_{(l)}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2) = \frac{f(\mathbf{t}_{(l)} \mid h_{(l)}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2) f(h_{(l)})}{f(\mathbf{t}_{(l)} \mid \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2)} \tag{8}$$

The conditional likelihood expression of $\mathbf{t}_{(l)}$ in the numerator of (8) is given by

$$\begin{aligned} f(\mathbf{t}_{(l)} \mid h_{(l)}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2) &= \prod_{i=1}^{N} \prod_{j=1}^{k} \left(h_{(l)} \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j})\right)^{\delta_{ij}(t_{(l)})} \exp\left(-h_{(l)} B_{(l)}\right) \\ &= \prod_{i=1}^{N} \prod_{j=1}^{k} \left(\exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j})\right)^{\delta_{ij}(t_{(l)})} \times h_{(l)}^{D_{(l)}} \\ &\quad \times \exp\left(-D_{(l)} h_{(l)} B_{(l)}\right) \end{aligned} \tag{9}$$

with $B_{(l)} = \sum_{i'j' \in R(t_{(l)})} \exp\left(\mathbf{x}_{i'j'}^t \boldsymbol{\beta} + v_{1i'} + v_{2j'}\right)$, where $R\left(t_{(l)}\right)$ is the risk set at time $t_{(l)}$ and $D_{(l)}$ the number of observed reaction times equal to $t_{(l)}$.

The conditional likelihood expression of $\mathbf{t}_{(l)}$ in the denominator of (8) can be found by integrating out $h_{(l)}$ from (9)

$$\begin{aligned} f(\mathbf{t}_{(l)} \mid \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2) &= \prod_{i=1}^{N} \prod_{j=1}^{k} \left(\exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j})\right)^{\delta_{ij}(t_{(l)})} \\ &\quad \times \int_0^\infty h_{(l)}^{D_{(l)}} \exp\left(-D_{(l)} h_{(l)} B_{(l)}\right) dh_{(l)} \\ &= \prod_{i=1}^{N} \prod_{j=1}^{k} \left(\exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j})\right)^{\delta_{ij}(t_{(l)})} \\ &\quad \times \frac{c^{ch_{(l)}^*}}{\Gamma(ch_{(l)}^*)} \left(c + D_{(l)} B_{(l)}\right)^{-ch_{(l)}^* - D_{(l)}} \Gamma\left(ch_{(l)}^* + D_{(l)}\right) \end{aligned}$$

with $h_{(l)}^* = H_0^*(t_{(l)}) - H_0^*(t_{(l-1)})$. Therefore, we end up with

$$\begin{aligned} f(h_{(l)} \mid \mathbf{t}_{(l)}, \sigma_1, \sigma_2, \boldsymbol{\beta}, \mathbf{v}_1, \mathbf{v}_2) &= \left(c + D_{(l)} B_{(l)}\right)^{ch_{(l)}^* + D_{(l)}} h_{(l)}^{ch_{(l)}^* + D_{(l)} - 1} \\ &\quad \times \exp\left(-h_{(l)}(c + D_{(l)} B_{(l)})\right) \left(\Gamma(ch_{(l)}^* + D_{(l)})\right)^{-1}, \end{aligned}$$

which corresponds to a gamma density with parameters $ch_{(l)}^* + D_{(l)}$ and $c + D_{(l)} B_{(l)}$, and hence we can sample from the gamma density.

For the random effect v_{1i} $(i = 1, \ldots, N)$, the conditional posterior density is proportional to

$$\begin{aligned}
f(v_{1i} \mid \mathbf{t}, \mathbf{h}, \sigma_1, \sigma_2, \mathbf{v}_2) &\propto f(\mathbf{t} \mid \mathbf{h}, \sigma_1, \sigma_2, \mathbf{v}_2) f(v_{1i}) \\
&= \prod_{j=1}^{k} \prod_{l:t_{(l)} \leq t_{ij}} \left(h_{(l)} \exp\left(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j} \right) \right)^{\delta_{ij}(t_{(l)})} \\
&\quad \times \exp\left\{ - \left(h_{(l)} \exp\left(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j} \right) \right) \right\} \times \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{v_{1i}^2}{2\sigma_1^2} \right)
\end{aligned}$$

Given the logconcavity of this conditional posterior density, we can again rely on the adaptive rejection resampling algorithm. Similar arguments can be used for the random effect v_{2j} $(j = 1, \ldots, K)$.

Finally, as the gamma distribution is the conjugate prior for the precision of the normal distribution, sampling from its posterior density is easily obtained.

3 Simulation study

Competitive models

To assess the finite sample properties of the proposed estimation procedure for the parameters of interest in model (1), we performed a simulation study. We contrasted the performance of the newly proposed semi-parametric PH-model with crossed random effects with 3 existing PH-models for reaction times: (a) a fully parametric PH-model, assuming a shifted Weibull distribution, with crossed random effects (b) a more standard semi-parametric PH-model with a single random effect for subject, and (c) a discrete PH-model with crossed random effects. Under the fully parametric approach we assumed the following hazards model :

$$h_{ij}(t) = \lambda_0 \gamma (t-\psi)^{\gamma-1} \exp\left(\mathbf{x}_{ij}^t \boldsymbol{\beta}^{*} + v_{1i}^{*} + v_{2j}^{*}\right) \quad \text{when } t \geq \psi \tag{10}$$

with $\psi \in \Re^+$ the shift parameter, $\gamma \in \Re^+$ the shape parameter, and $\lambda_0 \in \Re^+$ the (baseline) rate parameter of the shifted 3-parameter Weibull (Rouder, Tuerlinckx, Speckman, Lu, & Gomez, 2008). When ψ equals zero and γ equals one, the shifted Weibull distribution reduces to an exponential distribution. The roles of each of these parameters and their estimation through Markov Chain Monte Carlo integration is further discussed by Rouder and colleagues (2008). Similar as in model (1) v_{1i}^{*} and v_{2j}^{*} are crossed random effects, reflecting the speed of subject i and intensity of item j (Loeys, Rosseel & Baten, 2011), where here too the assumption is made that $v_{1i}^{*} \sim N(0, \sigma_1^{*2})$ and $v_{2j}^{*} \sim N(0, \sigma_2^{*2})$. As noted by one of the referees, the shift parameters are often hard to estimate, but results presented below did not drastically differ for the Weibull model with or without shift parameter.

The second model considered for comparison, the semi-parametric PH-model with a single random effect for subject can be expressed as

$$h_{ij}(t) = h_0(t) \exp\left(\mathbf{x}_{ij}^t \boldsymbol{\beta}^{**} + v_{1i}^{**}\right) \tag{11}$$

In model (11) it is assumed that the correlation between measurements on the same subject i is captured by the random effect v_{1i}^{**}, while the heterogeneity due to items is ignored as compared to model (1).

Finally, we compare the proposed estimation approach based on continuous time to an approach based on discrete time. More specifically, we define the discrete-time hazard rate λ_{ijk} for subject i on item j in interval k as

$$\lambda_{ijk} = \Pr\left(T_{ij} = k \mid T_{ij} \geq k, \mathbf{x}_{ij}, w_{1i}, w_{2j}\right), \qquad (12)$$

Assuming model (1), it can be shown (Prentice & Gloecker, 1978) that

$$\lambda_{ijk} = 1 - \exp\left(-\exp(\gamma_k + \mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j})\right)$$

which can be rewritten as

$$\log\left[-\log(1 - \lambda_{ijk})\right] = \gamma_k + \mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j} \qquad (13)$$

where the coefficients $\boldsymbol{\beta}$ are identical to that of model (1) and α_k is a constant related to the conditional survival probability in the interval k. The grouped data survival model is therefore equivalent to the binary response model with the complementary log-log link function. To fit this model with generalized linear mixed model software that allows for crossed random effects (for e.g. proc Glimmix in SAS, or the glmer-function in the R-library lme4), one must treat each discrete time unit for each subject as a separate observation. For each of these observations, the response is then dichotomous, corresponding to whether or not the subject gave an answer in the time unit.

Data generation

For a fixed sample size of 15 subjects and 20 items (approximately reflecting the size of the illustrating example), we generated a single response time for each subject-item combination. Conditional on covariates and random effects, the continuous response times were assumed to follow (i) an exponential distribution ('exp') with rate parameter $\lambda = 1/500$ (simulation setting (S1a) through (S1e) in table 1), (ii) a Weibull distribution ('Wei') with rate parameter $\lambda = 1/500$, shape parameter $\gamma = 2$ and $\psi = 0$ (setting (S2a) through (S2c)), or (iii) a piecewise exponential distribution ('p-e') with rate 1/1000, 1/400 and 1/700 and change points at 400 and 600 (setting (S3a) through (S3c)). A single observed item-specific covariate X_j is assumed that is either Bernoulli with probability 0.5 ('B(0.5)') or standard normal ('N(0,1)'), except for setting (S1e) which simultaneously has an item-specific fixed effect, a subject-specific fixed effect and its interaction. All settings assumed subject- and item-specific random effects v_{1i} and v_{2j} that are normally distributed. Also note that in setting (S1d) the heterogeneity due to items was set to zero. For the discrete PH-approach, 10 intervals of length 100 were considered.

The semi-parametric PH-model with crossed random effects (1) ('Cox PH-2') and its discrete counterpart (13) ('Discrete PH-2') hold under all simulation settings except (S1d); the shifted Weibull model (14) ('Weibull-2') additionally does not hold under the piece-wise exponential baseline hazard; while the single frailty model (11) ('Cox PH-1') only holds under scenario (S1d).

Simulation Results

Table 1 presents for each of these 4 models summaries of the estimated fixed effects and the estimated standard deviation of the subject random effect (estimated standard deviations of the item random effect performed similar and are not shown) that are based on 200 repetitions in each setting. All estimates are derived from the posterior means for these parameters obtained from 2 independent chains of length 2500, except for the estimates from model (13) where a frequentist approach was taken.

The semi-parametric PH-model does not show any evidence of biased fixed estimates. It is interesting to note that - while making less distributional assumptions - the cost in terms of efficiency for the fixed effect estimates as compared to its parametric Weibull counterpart - is very mild, if any. There is some indication of undercoverage of the 95% posterior interval for the fixed effect estimate, both in the semi-parametric and parametric PH-models with crossed random effects. The variability of the subject random effect is well recovered in all scenarios by all models with crossed random effects, also when no item variability was generated (S1d) but assumed. Of note, the continuous and discrete PH-model with crossed random effect yield rather similar results.

While the semi-parametric approach treats the baseline hazard as a nuisance, it can still be of interest to look at the performance of its estimation. The left panel of figure 2 presents as an example the estimated cumulative baseline hazard for each of 200 simulations under scenario (S3a), and the true cumulative baseline hazard is on average relatively well approximated. The shifted Weibull family does not include the piecewise exponential distribution (S3a), but does perform relatively well from this perspective too (right panel of figure 2).

When random variance due to both subjects and items is present (all settings except for S1d) but only one source is included in the model, potential problems arise. In the context of linear mixed effect models, Baayen, Davidson and Bates (2008) demonstrated that deflated standard errors and biased fixed effect estimates results when item or person variance is not taken into account. On comparing the results obtained from model (1) with model (11), we find similar findings in the semi-parametric PH-framework. Moreover there is evidence of a deflated variance component for the subject random effect when ignoring the item variability.

Based on the settings explored, we can conclude that while the semi-parametric approach does not force the practitioner to make specific distributional choices for the response time distributions, it does not come at the price of much efficiency loss. Furthermore, ignoring one level of heterogeneity as in the standard frailty model with a single random effect may seriously compromise the inference for the fixed effect parameters. Finally, it is worth noting that while one could have expected a loss in precision or power by discretizing response times, the PH-approach with crossed random effects for discrete response times performs surprisingly well.

4 Example

In a recently performed visual object recognition study, Schettino, Loeys, Delplanque and Pourtois (2011) explored the effect of emotional content on early recognition. More precisely, an initial blurred visual scene was first shown, before the actual content of the stimulus was gradually revealed. The first (blurred) image level of a given picture was presented for 500ms, followed by a 250ms blank screen. Next, the second image level of the same picture (containing more HSF information) was immediately presented for 500 ms, plus 250 ms blank screen, and the same procedure was repeated until the presentation of the sixth image level (i.e. unfiltered picture). Participants were asked to press the spacebar key on a keyboard as soon as they felt they could decide with sufficient confidence, whether the presented scene contained a living object or not. Pressing the spacebar key immediately interrupted the presentation of the stimuli. Standard neutral and emotional scenes were selected from the International Affective Picture System, a standardized database containing emotionally-evocative pictures that depict objects and scenes across a wide range of categories and situations. The pictures were divided into three emotion categories, according to their pre-defined valence scores: neutral, unpleasant and pleasant. Since there are gender differences in both valence and arousal ratings, two sets of 138 pictures were selected in order to balance the arousal levels of the emotional pictures across male versus female participants. Among these selected pictures, only 42 were shared between male and female participants.

In total, 19 psychology students participated to the study. For this illustration, we further selected the 11 neutral and 16 pleasant pictures that were shared between male and female participants. The main purpose of the analysis below is to explore the effect of pleasant versus neutral pictures on the time to recognition (defined as the time between initiation of the trial and pressing the spacebar). We assume a semi-parametric proportional hazards model for the reaction time T_{ij} (participants $i = 1, \ldots, 19$ and pictures $j = 1, \ldots, 27$) with a fixed effect for emotion of the picture (denoted x_{1j} being equal to 1 if neutral and 0 if pleasant) and crossed random effects v_{1i} for participants and v_{2j} for pictures, i.e.

$$h_{ij}(t) = h_0(t) \exp\left(\beta_1 x_{1j} + v_{1i} + v_{2j}\right) \tag{14}$$

with $h_0(t)$ further left unspecified, $\beta_1, v_{1i}, v_{2j} \in \Re$ and $v_{1i} \sim N(0, \sigma_1^2)$ and $v_{2j} \sim N(0, \sigma_2^2)$. In model (14) a positive (negative) β_1 implies that neutral pictures are recognized faster (slower, respectively) than pleasant pictures. Furthermore, participants with a positive v_{1i} tend to respond faster, i.e. they have a higher speed, while pictures with a positive v_{2j} need less processing time conditional on whether the picture is neutral or pleasant.

Bayesian Analysis

We fitted model (14) in OPENBUGS (detailed code can be found in appendix A2), and ran 2 independent chains of length 15000, where the first half's were used as a burn-in period. The total run time for the two chains of length 15000 was about 30 minutes on a Windows PC. When the MCMC algorithm is used, it is important to assess first its convergence to ensure that the random draws are actually coming from the posterior distribution of interest. An informal approach is to visually inspect the plot of the Gibbs sampler run, the so-called trace plot, and the autocorrelation plot, with high autocorrelation indicative for slow mixing and possibly nonconvergence to the limiting distribution (because the chain will tend to explore less space in finite time). Gelman and Rubin (1992) further proposed a formal diagnostic method when running several chains, the so-called scale reduction factor $\hat{R}$, which is based on the estimated within-chain variance and between-chain variance. If the value of $\hat{R}$ is close to 1, for example less than 1.2, we may conclude that the MCMC algorithm reasonably converges; otherwise the algorithm may fail to converge. Figure 3 presents the trace plots after the burn-in period for both chains and a corresponding autocorrelation plots for the fixed effect parameter β_1 and the random effect variance components σ_1 and σ_2. While for the latter two there is no indication of a problem in convergence or in high autocorrelation, the mixing of the 2 chains for β_1 is very slow and the autocorrelation pretty high. The Gelman-Rubin statistic equals 1.09. Although this is smaller than the threshold typically used, we further explored this phenomenon in the earlier described simulation settings and observed similar behavior throughout.

Interestingly, we find marginal evidence for prolonged response times for pleasant versus neutral images (posterior mean for β_1=0.45 with 95% posterior density ranging from -0.02 to 0.86). Furthermore, the variability in speed amongst participants (posterior mean for σ_1=0.56 with 95% posterior density ranging from 0.39 to 0.80) is somewhat smaller than the variability in intensity amongst items (posterior mean for σ_1=0.73 with 95% posterior density ranging from 0.54 to 0.99). Figure 4 shows the posterior mean of the parameter v_{2j} with its 95% posterior interval for all 27 images, and allows to compare the intensity between the presented items.

Model Validity

The validity of the model is assessed here using posterior predictive checks, which are based on the response time for participant i on item j predicted from the posterior distribution of the model parameters (van der Linden, Breithaupt, Chuah, & Yang, 2007), denoted by $\tilde{t}_{ij}$. For each observation t_{ij}, one can calculate the left-sided probability of exceedance of the observation under its predictive density,

$$\Pr(\tilde{t}_{ij} < t_{ij}) \quad i = 1, \ldots, N; j = 1, \ldots, k$$

It can be shown that - if the model fits well - the cumulative distribution of these probabilities follows the identity line. The results from this check of the global fit of the model is shown in the left panel of Figure 5, which provides evidence that the model fits the data well. Indeed the cumulative distribution plot almost coincides with the identity line. This same analysis was repeated with the cumulative distributions of the predictive probabilities for the individual items (right panel of Figure 5). Because these distributions are based on smaller samples, we see more variability but still a very reasonable fit for this random selection of items.

Alternative modeling approach: the log-normal model.

To contrast the newly proposed semi-parametric approach with existing parametric approaches for reaction time modeling, we also present the results from the most popular model, the lognormal model (Van der Linden, 2006) distribution, assuming crossed random effects for subjects and items too.

The density of the lognormal distribution is given by

$$f(t_{ij} \mid \mu^*_{ij}, \sigma^*) = \frac{1}{\sqrt{2\pi}\sigma^* t_{ij}} \exp\left[-\frac{1}{2}\left(\frac{\log t_{ij} - \mu^*_{ij}}{\sigma^*}\right)^2\right] \tag{15}$$

where we restricted the variance σ^{*2} not to depend on subject or item, while the mean μ^*_{ij} equals,

$$\mu^*_{ij} = \beta^*_0 + \beta^*_1 x_{1j} + v^*_{1i} + v^*_{2j}, \tag{16}$$

with $\beta^*_0, \beta^*_1, v^*_{1i}, v^*_{2j} \in \Re$ and $\sigma^* \in \Re^+$. We further assume that random intercepts for subject and item are normally distributed, i.e. $v^*_{1i} \sim N(0, \sigma_1^{*2})$ and $v_{2j} \sim N(0, \sigma_2^{*2})$. The parameters in (16) express effects on the mean of the (log) response time and can not directly be related to the parameters in the PH-model (14). Moreover, their signs have opposite interpretations. Estimated parameters are shown in table 2, and similar conclusions are reached as before. However the posterior predictive checks based on the predicted response time distribution provide evidence of a poorer global fit as compared to the PH-model (left panel of Figure 5).

Accuracy

As participants in the above discussed visual recognition study were under time pressure to give a correct answer as soon as possible, not all answers to the question whether the presented scene was gradually revealing a living object or not were correct (overall, nearly 95% of the answers were correct). In contrast to the traditional item-response framework where responses are incorporated in the reaction time models or vice versa, we focused here on a distinct model for the response time. It is still possible though to explore the speed-accuracy trade-off. Indeed, following van der Linden (2007), one can propose a response model, a model for response time and a higher-level structure accounting for the dependencies between the item and subject parameters in these models. While van der Linden (2007) used a log-normal model for the response times, and Loeys, Baten and Rosseel (2011) used a shifted Weibull distribution with crossed random effects for the response time for the analysis of psycholinguistic data, this joint modeling framework is flexible enough to include the newly proposed semi-parametric PH-model too. More specfically, we continue to assume model (14) for the response time, while we assume the following logistic regression model with crossed random effects for the response

$$\text{logit}(P(Y_{ij} = 1)) = \alpha_0 + \alpha_1 x_{1j} + w_{1i} + w_{2j} \tag{17}$$

with Y_{ij} denoting the response of subject i to item j (1 if correct, else 0), and x_{1j} a dummy variable for the neutrality of the scene as before. In model (17) a positive value of α_1 implies a higher accuracy rate associated with neutral scenes as compared to pleasant scenes. Further, w_{1j} and w_{2j} are zero-mean normal random effects, capturing the heterogeneity between subjects and between items, respectively.

A joint modeling approach is invoked by imposing a joint multivariate distribution on the vector of all random effects for subject and item from models (14) and (17). More specifically, we assume that both the subject parameters v_{1i} and w_{1i} and the item parameters v_{2i} and w_{2i} follow a bivariate normal distribution with mean 0 and a covariance structure specified by

$$\Sigma_S = \begin{pmatrix} \sigma_{v1}^2 & \rho_1\sigma_{v1}\sigma_{w1} \\ \rho_1\sigma_{v1}\sigma_{w1} & \sigma_{w1}^2 \end{pmatrix} \text{ and } \Sigma_I = \begin{pmatrix} \sigma_{v2}^2 & \rho_2\sigma_{v2}\sigma_{w2} \\ \rho_2\sigma_{v2}\sigma_{w2} & \sigma_{w2}^2 \end{pmatrix}. \quad (18)$$

In (18), ρ_1 measures the correlation between speed (as captured by v_1) and ability (as captured by w_1) at the subject level, while ρ_2 measures the correlation between time intensity (as captured by the opposite of v_2) and difficulty at the item level (as captured by the opposite of w_2).

Setting the inverse Wishart distribution as prior for the covariances Σ_S and Σ_I for example, one can proceed with joint estimation of models (14) and (17) within the Bayesian framework using Gibbs sampling (for further details, see van der Linden, 2007; and Loeys et al., 2011). For the visual recognition study, we find under this joint modeling approach similar estimated effects (not shown) of the emotional context of the scene, and of the variability in response time due to participants and items as under the response time modeling alone. Moreover, assuming (18), a negative correlation between speed and ability (left panel of figure 6), and a positive correlation between the difficulties and time intensities of the items (right panel of figure 6) is found, the posterior mean for ρ_1 and ρ_2 equal -0.54 and 0.55, respectively (with 95% posterior interval from -0.83 to 0.01 and 0.13 tot 0.81, respectively). The latter finding is in line with van der Linden (2009) who also observed a strong tendency to a substantial positive correlation between difficulty and time intensity of the items across several educational tests, while for ρ_1 he found varying correlations. One possible explanation of the negative correlation between ability and speed, might be the better time management among more able participants. When there is ample time, these participants may slow down to maximally profit from it.

5 Discussion

The newly proposed semi-parametric PH-model nicely complements the increasingly popular crossed random effects model based on the lognormal distribution from van der Linden (2006). While the latter fits within the accelerated failure time framework, we made progress here within the PH-framework. As compared to modeling the effect of an item condition change from A to B on the mean reaction time, the hazard ratio is directly interpretable as a ratio of the instantaneous capacity of the test-taker for completing an item under condition A versus condition B at any time. While processing capacity is a critical construct in cognitive psychology (Wenger & Gibson, 2004) the use of the hazard function to assess changes in such processing capacity may get reinforced by the proposed model. A second advantage of the proposed semi-parametric model is its greater flexibility than fully parametric alternatives like the lognormal or Weibull model. The main drawback of the latter models is indeed the need to specify the distribution that most appropriately mirrors that of the actual response times. This is an important requirement that needs to be verified and an appropriate distribution may be difficult to identify (Rouder et al., 2008; Klein Entink et al., 2009). On the other hand, when a suitable distribution can be found, the parametric model might be more informative than the Cox model. Simulation studies presented in this paper do not show evidence of loss in efficiency when using semi-parametric versus parametric proportional hazards models. Third, the proposed PH-model can easily deal with censored observations. Many response time researchers truncate their data, excluding as spurious all response times falling outside a specified range (Ulrich & Miller, 1994). While censoring instead of truncating may greatly reduce the biasing effects of outliers, the psychometric literature has seen few developments of estimating response time distributions in the presence of censoring (Dolan, van der Maas, & Molenaar, 2002). In Appendix A3 we demonstrate how censoring can easily be incorporated in the proposed PH-framework. Fourth, treating both subject and items as random makes most sense from a theoretical perspective. It allows to generalize findings to the population of subjects and items (De Boeck, 2008), respectively, and to explain the person's variation in speed and the item's variation in intensity. As clearly demonstrated in the simulation study, ignoring the variability due to items may dramatically impact the validity of the statistical inference. Both the newly proposed model and van der Linden's log-normal models have crossed random effects and share the strength of acknowledging heterogeneity due to subjects and items. Fifth, in contrast to the traditional item-response framework where responses are incorporated in the reaction time models or vice versa, we focused here on a distinct model for the response time, and showed how it can easily be embedded within a joint modeling framework for response time and accuracy.

Weaknesses of the semi-parametric PH-model include the lack of diagnostic tools for checking the PH-assumption for the covariates of interest and the random effects, and for checking the distributional assumptions for the random effects. The assumed shift in hazards for each item in the proposed model may be questionable in some settings. Indeed, Ranger and Kuhn (2012) recently demonstrated that the reaction time distribution may differ substantially across items within a test. These authors therefor unified the proportional hazards models and accelerated failure time models in latent trait models for discrete response times allowing for such item-specific distributions, but their models should rather be viewed within the item-response theory framework with focus on item discrimination, i.e. the ability of an item to distinguish between subjects with unequal speed. Within that same framework, Wang, Fan, Chang & Douglas (2013) proposed the linear transformation with frailty model for continuous response times, a generalization that encompasses the lognormal model, the Weibull model and the Cox PH-model amongst others. In contrast to the approach presented in this paper, both the model of Ranger and Kuhn (2012) and the model of Wang et al. (2013) offer the flexibility of allowing for differential distributional assumptions between items, but their models do not allow to estimate the effect of observed item-specific characteristics as in our illustration.

Finally the proposed estimation procedure is computationally intensive, and improving the speed in estimating the parameters is currently under investigation. Several estimation procedures for (semi-)parametric PH-model with a single random effect or nested random effects exist. The penalized likelihood approach, the EM-algorithm (with several variants) or Bayesian techniques based on Gibbs sampling amongst others have different advantages and drawbacks (Abrahantes, Legrand, Burzykowski, Janssen, Ducrocq & Duchateau, 2006). In this first application of cross-classified semi-parametric PH-models, we opted to use Gibbs sampling. Computational speed may be improved by avoiding Gibbs sampling and working along the lines of Cho and Rabe-Hesketh (2011) who discuss the alternating imputation posterior (AIP) algorithm for crossed random effects in (generalized) linear mixed models. The AIP-algorithm alternates between an item wing in which item intensity is sampled for given person speed, and a subject wing in which person speed is sampled for given item intensity, and our hope is that the implementation of Laplacian approximations in these alternations may largely improve speed.

References

Abrahantes, J.C., Legrand, C., Burzykowski, T., Janssen, P., Ducrocq, V., & Duchateau, L. (2006). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational Statistics and Data Analysis, 51,* 3913-3930.

Allison, P.D. (1996). Fixed-effects partial likelihood for repeated events. *Sociological Methods and Research, 25,* 207-222.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language, 59,* 390-412.

Bloxom, B. (1985) Considerations in psychometric modeling of response time. *Psychometrika, 50,* 383-397.

Cho, S.-J. & Rabe-Hesketh, S. (2011). Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics and Data Analysis, 55,* 12-25.

Clayton, D.G. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society A, 148,* 82-117.

Clayton, D.G. (1991). A Monte Carlo method for Bayesian Inference in frailty models. *Biometrics, 47,* 467-485.

De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73 (4),* 533-559.

Dolan, C.V., van der Maas, H.L.J., & Molenaar, P.C.M. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods, Instruments & Computers, 34 (3),* 304-323.

Douglas, J.A., Kosorok, M.R., & Chewning, B. (1999). A latent variable model for discrete multivariate psychometric waiting times. *Psychometrika, 64,* 69-82.

Duchateau, L., & Janssen, P. (2008). *The frailty model.* New York: Springer.

Furneaux, W. (1952). Some speed, error and difficulty relationship within a problem solving situation. *Nature, 170,* 37-38.

Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7,* 457472.

Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics, 41,* 337-348.

Ibrahim, J.G., Chen, M.-H., & Sinha, D. (2005). *Bayesian survival analysis.* New York: Springer.

Kalbfleisch, J.D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B, 40,* 214-221.

Klein Entink, R.H., van der Linden, W.J., & Fox, J.P. (2009). A Box-Cox normal model for response times. *British Journal of Mathematical & Statistical Psychology, 62,* 621-640.

Legrand, C., Ducrocq, V., Janssen, P., Sylvester, R., & Duchateau, L. (2005). A Bayesian approach to jointly estimate center and treatment by center heterogeneity in a proportional hazards model. *Statistics in Medicine, 24 (24),* 3789-3804.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine, 28,* 3049-3082.

Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction times and accuracy in psycholinguistic experiments. *Psychometrika, 3,* 487-503. *Journal of the American Statistical Association, 81*, pp. 635644

Prentice, R.L., & Gloecker, L.A. (1978) Regression analysis of grouped survival data with application to breast cancer data. *Biometrics, 34*, 57-67.

Ranger, J., & Kuhn, J.-T. (2012) A flexible latent trait model for response times in tests. *Psychometrika, 77 (1),* 31-47.

Ranger, J., & Ortner, T. (2012) A latent trait model for response times on tests employing the proportional hazards model. *British Journal of Mathematical and Statistical Psychology,* 65, 334-349.

Raudenbush. S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics, 18,* 321-349.

Ridder, G. (1987). *The sensitivity of duration models to misspecified unobserved heterogeneity and duration dependence.* University of Amsterdam, Amsterdam.

Rouder, J., Sun, D., Speckmanm P.L., Lu, J. & Zhou, D. (2003). A Hierarchical Bayesian Statistical Framework for Response Time Distributions. *Psychometrika, 68,* 589-606.

Rouder, J., Tuerlinckx, F., Speckman, P., Lu, J., & Gomez, P. (2008). A Hierarchical Approach for Fitting Curves to Response Time Measurements. *Psychonomic Bulletin & Review, 15 (6),* 1201-1208.

Scheiblechner, H. (1985). Psychometric models for speed-test construction: the linear exponential model. In S.E. Embretson (Ed.) *Test design, developments in psychology and psychometrics* London Academic Press.

Schettino, A., Loeys, T., Delplanque, S., & Pourtois, G. (2011). Brain dynamics of upstream perceptual processes leading to visual object recognition: a high density ERP topographic mapping study. *Neuroimage, 55 (3),* 1227-1241.

Shih, J.H. & Lu, S.E. (2009). Semiparametric estimation of a nested random effects model for the analysis of multi-level clustered failure time data. *Computational statistics and data analysis, 53 (11),* 3864-3871.

Thissen, D. (1983). Timed testing: an approach using item response theory. In D.J. Weiss (ed.) *New horizons in testing: Latent trait theory and computerized adaptive testing.* New York: Academic Press.

Ulrich, R., & Miller, J. (1994). Effects of Truncation on Reaction Time Analysis. *Journal of Experimental Psychology: General, 123 (1),* 34-80.

Van Breukelen, G. (2005). Psychometric modeling of response speed and accuracy with mixed and conditional regression. *Psychometrika, 70 (2),* 359-376.

van der Linden, W. (2006). A Lognormal Model for Response Times on Test Items. *Journal of Educational and Behavioral Statistics, 31 (2),* 181-204.

van der Linden, W. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72 (3),* 297-308.

van der Linden, W., Breithaupt, K., Chuah, S.C., & Zhang, Y. (2007). Detecting differential speededness in multistage testing. *Journal of Educational Measurement,44*, 117-130.

van der Linden, W. (2009). Conceptual Issues in Response-Time Modeling. *Journal of Educational Measurement, 46 (3),* 247-272.

Vaupel, J. , Manton, K.G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography, 16,* 439-454.

Wang, C., Chang, H.-H., & Douglas, J.A. (2012). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology,* 66 (1), 144-168.

Wang, C., Fan, Z., Chang, H.-H., & Douglas, J.A. (2013). A semiparametric model for jointly analyzing response times and accuracy in computerized testing. *Journal of Educational and Behavioral Statistics.* doi:10.3102/1076998612461831.

Wenger, M.J. & Gibson, B.S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology, 30 (4),* 708-719.

Appendix

Appendix A1

In order to derive the approximation for the likelihood (4), consider the contribution of the second ordered response time (assume this belongs to subject k on item l):

$$\begin{aligned} L(t_{kl}) &= \exp\left(-\exp(\mathbf{x}_{kl}^t\boldsymbol{\beta} + v_{1k} + v_{2l})h_{(1)}\right) - \left[1 - \exp\left(-\exp(\mathbf{x}_{kl}^t\boldsymbol{\beta} + v_{1k} + v_{2l})h_{(1)} + h_{(2)}\right)\right] \\ &= \phi(h_{(1)}) - \phi(h_{(1)} + h_{(2)}) \end{aligned}$$

Using first-order Taylor series approximation around $h_{(1)}$, this can be approximated by

$$\begin{aligned} L(t_{kl}) &\approx \left(h_{(1)} - (h_{(1)} + h_{(2)})\right)\phi^{'}(h_{(1)} + h_{(2)}) \\ &= h_{(2)}\exp(\mathbf{x}_{kl}^t\boldsymbol{\beta} + v_{1k} + v_{2l})\exp\left(-\exp(\mathbf{x}_{kl}^t\boldsymbol{\beta} + v_{1k} + v_{2l})h_{(1)} + h_{(2)}\right) \end{aligned}$$

By doing so for every observed response time, likelihood approximation (2) is obtained.

In order to demonstrate the logconcavity of the conditional posterior density $f(\beta_l \mid \mathbf{t}, \mathbf{h}, \sigma_1, \sigma_2, \boldsymbol{\beta}_{(-l)}, \mathbf{v}_1, \mathbf{v}_2)$, we need to prove that its second derivative is nonpositive.

$$\frac{\partial^2}{\partial \beta_l^2} \log f(\beta_l \mid \mathbf{t}, \mathbf{h}, \sigma_1, \sigma_2, \boldsymbol{\beta}_{(-l)}, \mathbf{v}_1, \mathbf{v}_2)$$

$$= \frac{\partial^2}{\partial \beta_l^2} \sum_{i=1}^{N} \sum_{j=1}^{k} \sum_{m: t_{(m)} \leq t_{ij}} \big[\delta_{ij}(t_{(m)}) \log \big(h_{(m)} \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j})\big)$$

$$- h_{(m)} \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j})\big]$$

$$= - \sum_{i=1}^{N} \sum_{j=1}^{k} \sum_{m: t_{(m)} \leq t_{ij}} h_{(m)} \exp(\mathbf{x}_{ij}^t \boldsymbol{\beta} + v_{1i} + v_{2j}) x_{ijl}^2$$

This last expression is always less than or equal to zero.

The logconcavity of the conditional posterior distribution of the random effects can be shown similarly.

Appendix A2

```
# N: total number of observations (obs.t contains all response times)
# T: number of distinct response times (t contains ordered distinct response times)
# subject: label for the participant (Nsubj is total number participants)
# item: label for the item (Nitem is total number of items)
# X: predictor
# eps: epsilon (small number like 0.0001 for eg.)
model
{
for (i in 1:N){
    for (k in 1:T){
         Y[i,k]<-step(obs.t[i]-t[k]+eps)
         dN[i,k]<-Y[i,k]*step(t[k+1]-obs.t[i]-eps)
                 }
             }

for (k in 1:T){
  for (i in 1:N){
        dN[i,k]~dpois(Idt[i,k])
        Idt[i,k]<-Y[i,k]*exp(beta1*X[i]+b1[subject[i]]+b2[item[i]])*dL0[k]
}
   dL0[k]~dgamma(mu[k],c)
   mu[k]<-dL0.star[k]*c
   S0[k]<-exp(-sum(dL0[1:k]))
   S1[k]<-pow(exp(-sum(dL0[1:k])),exp(beta1))
}
for (i in 1:Nsubj){
   b1[i]~dnorm(0.0,tau1)
 }
for (j in 1:Nitem){
   b2[j]~dnorm(0.0,tau2)
 }
tau1~dgamma(0.001,0.001)
tau2~dgamma(0.001,0.001)
sigma1<-sqrt(1/tau1)
sigma2<-sqrt(1/tau2)
c<-0.0001
r<-1
for  (k in 1:T){
   dL0.star[k]<-r*(t[k+1]-t[k])
}
beta1~dnorm(0,0.0000001)
}
```

Appendix A3

Response times can be right censored - that is response times are known for only a portion of the subject/item combinations under study, and the remainder of the response times are known only to exceed certain values. Specifically, an observation is said to be right censored at c if the exact value of the observation is not known but only that it is greater than or equal to c.

Suppose that there are N subjects and k items under study, and associated with subject i and item j is a response time T_{ij} and a censoring time C_{ij}. The T_{ij} are assumed to be distributed with density $f(t)$ and survivor function $S(t)$. The response time T_{ij} will be observed only if $T_{ij} \leq C_{ij}$. Censored data can be represented by pairs of random variables (Y_{ij}, δ_{ij}), where $Y_{ij} = \min(T_{ij}, C_{ij})$, and δ_{ij} equals 1 if $T_{ij} \leq C_{ij}$ and 0 if $T_{ij} > C_{ij}$.

In the presence of such censoring, the grouped-data likelihood function (4) corresponding to model (1) equals

$$\prod_{i=1}^{N}\prod_{j=1}^{k} \exp\left(-\exp(\mathbf{x}_{ij}^{t}\boldsymbol{\beta} + v_{1i} + v_{2j}) \textstyle\sum_{m:t_{(m)}<y_{ij}} h_{(m)}\right) \times \left[1 - \exp\left(-\exp(\mathbf{x}_{ij}^{t}\boldsymbol{\beta} + v_{1i} + v_{2j}) h_{(s,ij)}\right)\right]^{\delta_{ij}}$$

with $s, ij = \min\{m : t_{(m)} \geq y_{ij}\}$ and $t_{(m)}$ is the m-th ordered *observed* response time.

Using the same Taylor expansion as before, this expression can be approximated by

$$\prod_{i=1}^{N}\prod_{j=1}^{k}\prod_{m:t_{(m)}\leq y_{ij}} \left(h_{(m)}\exp(\mathbf{x}_{ij}^{t}\boldsymbol{\beta}+v_{1i}+v_{2j})\right)^{\delta_{ij}(t_{(m)})} \exp\left(-h_{(m)}\exp(\mathbf{x}_{ij}^{t}\boldsymbol{\beta}+v_{1i}+v_{2j})\right) \tag{19}$$

with $\delta_{ij}(t_{(m)})$ equal to 1 if $y_{ij} = t_{(m)}$ and $\delta_{ij} = 1$, and else 0.

The only change that needs to be made to the OPENBUGS code in Appendix A2 is to replace

```
dN[i,k]<-Y[i,k]*step(t[k+1]-obs.t[i]-eps)
```

by

```
dN[i,k]<-Y[i,k]*step(t[k+1]-obs.t[i]-eps)*delta[i,k]
```

where δ is the above defined status indicator.

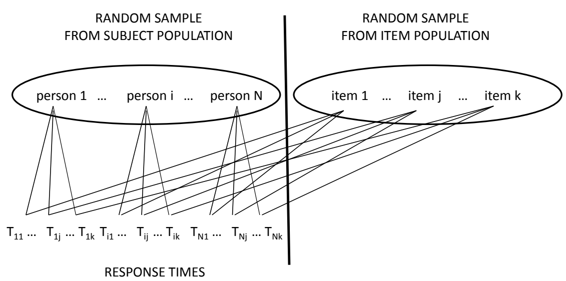

Figure 1: Graphical presentation of a crossed random effects model

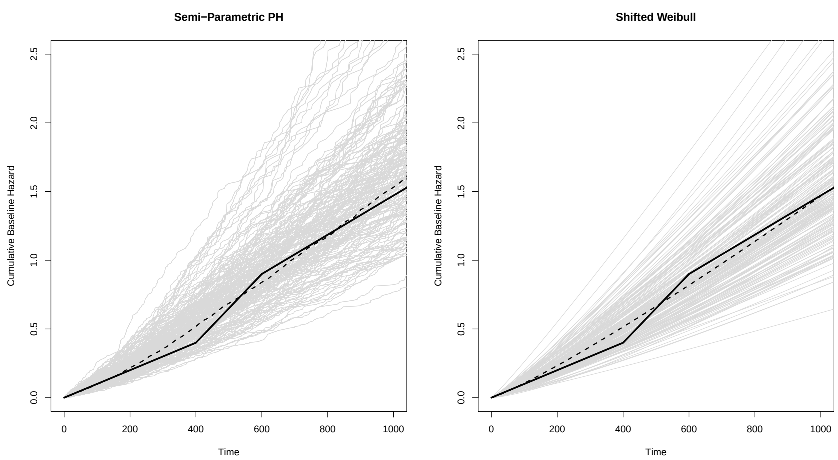

Figure 2: Estimation of the cumulative baseline hazard function under the semi-parametric PH-model with crossed random effects (left panel) and the shifted Weibull model with crossed random effects (right panel) in simulation setting (S3a). The black solid line represents the true cumulative baseline hazard (shifted exponential distribution), the dashed line represents the average of the estimated cumulative hazards over the 200 simulations (with the estimated hazard from each simulation run represented by the gray lines)

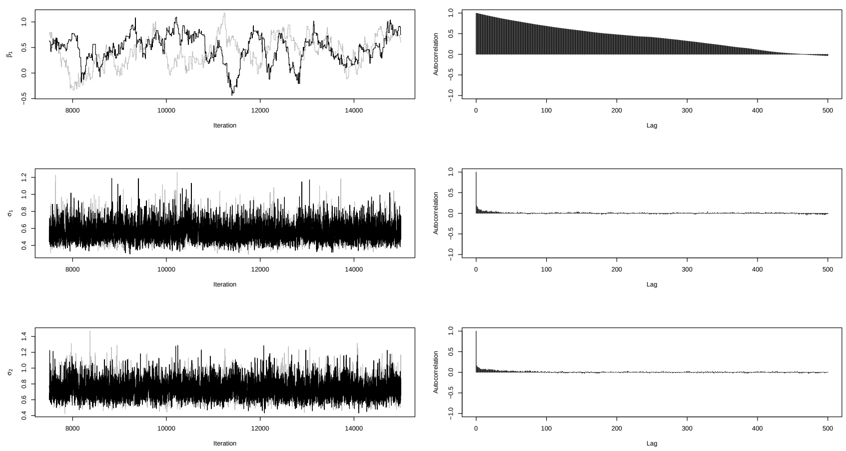

Figure 3: Traceplots and autocorrelation plots for the fixed effect parameter β_1 and random effect variance components σ_1 and σ_2 in the visual recognition study

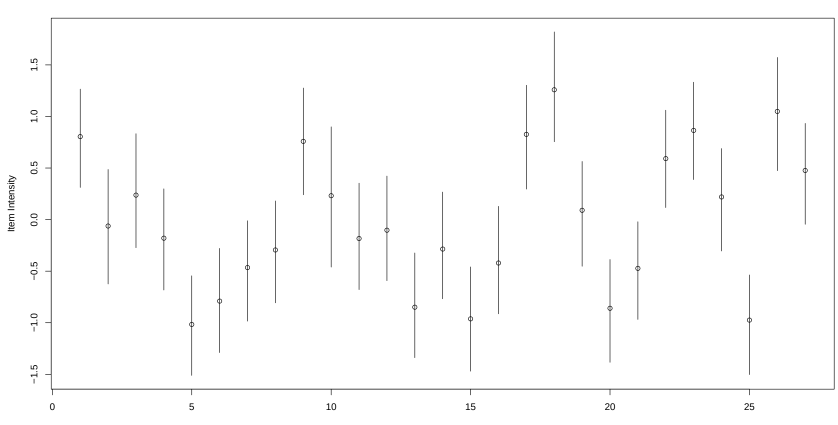

Figure 4: Estimated random item effects with 95% posterior intervals from the semi-parametric PH-model for the visual recognition study

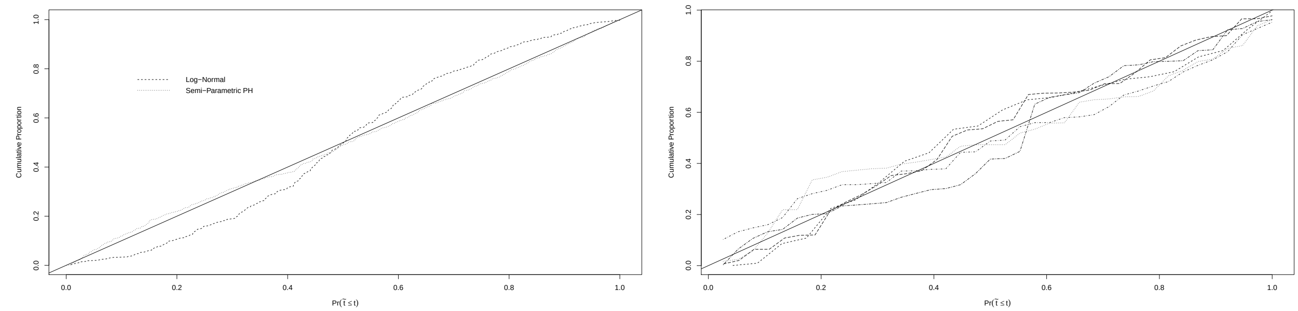

Figure 5: Left panel: overall fit of the semiparametric PH-model (dotted line) and lognormal model (dashed line) to the visual recognition data. The better the fit, the closer the empirical distribution to the identity line (solid line). Right panel: Item fit of the semi-parametric PH-model for a sample of 5 items of the 27 items in the visual recognition data.

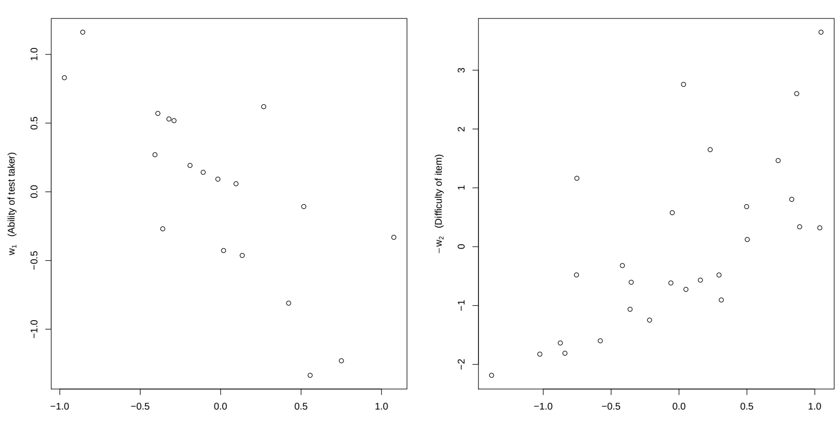

Figure 6: The association between speed and ability (left panel) and between intensity and difficulty (right panel) for the visual recognition study

						Fixed effect												Random effect SD			
						Cox PH - 2			Weibull - 2			Cox PH - 1			Discrete PH - 2			Cox PH - 2	Weibull - 2	Cox PH - 1	Discrete PH - 2
	base	X_{ij}	v_{1i}	v_{2j}	β	M (SD)	COV	POW	M (SD)	COV	POW	M (SD)	COV	POW	M (SD)	COV	POW	M (SD)	M (SD)	M (SD)	M (SD)
(S1a)	exp	$X_j \sim B(0.5)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$\log 1$	-0.03(0.29)	92.2	-	0.01 (0.27)	89.9	-	-0.00 (0.26)	64.0	-	0.00 (0.28)	92.0	-	0.49 (0.13)	0.49 (0.12)	0.39 (0.10)	0.49 (0.12)
(S1b)	exp	$X_j \sim B(0.5)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$\log 2$	0.66 (0.30)	91.3	64.6	0.65 (0.30)	88.9	65.9	0.58 (0.26)	57.0	92.5	0.69 (0.28)	91.5	72.0	0.49 (0.13)	0.49 (0.12)	0.39 (0.10)	0.49 (0.12)
(S1c)	exp	$X_j \sim B(0.5)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$\log 0.5$	-0.75 (0.30)	90.7	77.8	-0.64 (0.29)	87.5	64.5	-0.60 (0.26)	63.0	91.0	-0.69 (0.29)	93.0	71.5	0.49 (0.13)	0.49 (0.13)	0.39 (0.10)	0.49 (0.13)
(S1d)	exp	$X_j \sim B(0.5)$	$N(0,0.5^2)$	-	$\log 1$	-0.01 (0.13)	95.8	-	-0.01 (0.13)	95.2	-	-0.01 (0.12)	96.5	-	0.01 (0.13)	96.5	-	0.50 (0.13)	0.50 (0.13)	0.47 (0.11)	0.50 (0.13)
(S1e)	exp	$X_{1j} \sim N(0,1)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$\log 1.5$	0.43 (0.16)	93.8	76.4	0.30 (0.20)	81.2	58.4	0.36 (0.13)	69.0	94.5	0.41 (0.14)	95.0	80.0				
		$X_{2i} \sim N(0,1)$			$\log 1.25$	0.21 (0.21)	89.7	30.8	0.09 (0.22)	78.2	25.3	0.20 (0.16)	90.6	33.2	0.21 (0.16)	97.0	34.0	0.51 (0.16)	0.54 (0.17)	0.43 (0.27)	0.46 (0.13)
		$X_{1j} \times X_{2i}$			$\log 2$	0.72 (0.10)	89.2	100	0.69 (0.08)	96.0	100	0.62 (0.08)	83.0	100	0.69 (0.11)	92.5	100				
(S2a)	Wei	$X_j \sim N(0,1)$	$N(0,1)$	$N(0,1)$	$\log 1$	0.00 (0.26)	91.0	-	0.00 (0.20)	91.5	-	-0.00 (0.18)	52.5	-	0.00 (0.26)	92.0	-	1.00 (0.20)	1.00 (0.19)	0.61 (0.14)	0.98 (0.21)
(S2b)	Wei	$X_j \sim N(0,1)$	$N(0,1)$	$N(0,1)$	$2 * \log 2$	1.41 (0.27)	93.0	99.0	1.40 (0.35)	89.0	68.0	0.97 (0.22)	13.5	100	1.36 (0.27)	93.0	99.0	1.02 (0.21)	1.00 (0.19)	0.63 (0.14)	0.97 (0.21)
(S2c)	Wei	$X_j \sim N(0,1)$	$N(0,1)$	$N(0,1)$	$2 * \log 0.5$	-1.40 (0.28)	91.1	99.5	-1.40 (0.34)	87.5	60.5	-0.97 (0.22)	13.0	100	-1.36 (0.27)	91.0	100	1.02 (0.20)	1.01 (0.20)	0.63 (0.14)	1.00 (0.21)
(S3a)	p-e	$X_j \sim N(0,1)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$\log 1$	-0.01 (0.33)	90.0	-	0.01 (0.29)	91.0	-	0.01 (0.28)	58.5	-	0.00 (0.28)	92.0	-	0.55 (0.15)	0.55 (0.14)	0.42 (0.11)	0.48 (0.14)
(S3b)	p-e	$X_j \sim N(0,1)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$\log 2$	0.73 (0.30)	88.0	68.0	0.71 (0.33)	92.0	94.0	0.67 (0.30)	59.0	95.0	0.65 (0.28)	92.5	71.5	0.56 (0.13)	0.55 (0.14)	0.42 (0.11)	0.49 (0.13)
(S3c)	p-e	$X_j \sim N(0,1)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$\log 0.5$	-0.72 (0.34)	89.0	75.0	-0.64 (0.30)	87.0	51.0	-0.64 (0.29)	59.0	92.0	-0.69 (0.30)	90.5	68.0	0.52 (0.14)	0.55 (0.14)	0.42 (0.11)	0.48 (0.15)

Table 1: Simulation study: estimation of the fixed effect parameters and subject random effect standard deviation. (M=Mean, SD=Standard Deviation, COV=Coverage of the 95% posterior interval for β, POW=power of the 2-sided test $\beta = 0$ at the 5% significance level). Performance under the semi-parametric Cox PH-model with 2 random effects (Cox PH-2) and 1 random effect (Cox PH-2), the parametric shifted Weibull PH-model with 2 random effects (Weibull-2), and the discrete PH-model with 2 random effects (Discrete PH-2).

.

Semi-Parametric	β_1	σ_1	σ_2
	0.45	0.56	0.73
	(-0.02,0.86)	(0.39,0.80)	(0.54,0.99)
Log-Normal	β_1^*	σ_1^*	σ_2^*
	-0.07	0.09	0.11
	(-0.15,0.01)	(0.06,0.11)	(0.08,0.13)

Table 2: Estimated fixed effect of emotion, and standard deviation of random subject and item effects (with 95% posterior intervals) from the semi-parametric PH-model and the log-normal model with crossed random effects