Goodness-of-fit tests in proportional hazards models with random effects

Wenceslao González-Manteiga

University of Santiago de Compostela, Spain

María Dolores Martínez-Miranda

University of Granada, Spain

Ingrid Van Keilegom

KU Leuven, Belgium

International Symposium on Recent Advances in Statistics,
in Honor of Ingrid Van Keilegom

A Coruña, 15 June 2022

Precedents: Regression models with random effects

- Mixed effects models assume a flexible covariance structure which allows for non-constant correlation among the observations (longitudinal data, repeated measurements, clustered data and small area estimation).
- A semiparametric mixed effects model:

$$g\left(E\left[Y_{ij}|X_{ij}, b_i\right]\right) = m(X_{ij}) + b_i' Z_{ij} \quad (j = 1, \ldots, n_i;\ i = 1, \ldots, q)$$

- González-Manteiga, Lombardía-Cortiña, Martínez-Miranda and Sperlich (2013) considered kernel estimation (bandwidth selection) and bootstrapping for the above model in the case of $g(x) = x$.
- González-Manteiga, Martínez-Miranda and Van Keilegom (2016) proposed a goodness-of-fit test for the function $m(\cdot)$, based on the empirical distribution of the residuals.

Survival regression: The Cox proportional hazard model

The hazard function of survival time Y given X, $\lambda(t|X)$, is:

$$\lambda(t|X) = \lambda_0(t)\exp(\beta' X),$$

- $\lambda_0(t)$ is the unspecified baseline hazard,
- X is a vector of covariates and β the regression coefficients.
- Assume independent survival times.

Survival regression: The Cox proportional hazard model

The hazard function of survival time Y given X, $\lambda(t|X)$, is:

$$\lambda(t|X) = \lambda_0(t)\exp(\beta' X),$$

- $\lambda_0(t)$ is the unspecified baseline hazard,
- X is a vector of covariates and β the regression coefficients.
- Assume independent survival times.

But correlation often arises because there are clusters in the data.

- Multicenter and large-scale medical studies, e.g., patients' survival rates may differ substantially across different hospitals but may be similar within the same hospital.
- Studies with repeated measurements, e.g., multiple car accidents caused by the same individuals in a given year.
- Recurrent event data. Each individual has several outcomes representing gap times between events, e.g. recurring infections.

Cox model with random effects

Assume that the conditional hazard of survival time Y is:

$$\lambda(t|X_{ij}, b_i) = \lambda_0(t)\exp(\beta' X_{ij} + b_i' Z_{ij}) \quad (j = 1, \ldots, n_i;\ i = 1, \ldots, q)$$

- b_i are (iid) s-dimensional random effects of mean zero and distribution depending on an unknown parameter θ.
- X_{ij} is a vector of covariates, Z_{ij} is a sub-vector of $(1, X_{ij}')'$.
- Assume random right censoring so we observe (T, δ), where $T = \min(Y, C)$ and $\delta = I(Y \leq C)$.
- Assume $b_i \perp X_{ij}$ and $(T_{ij}, \delta_{ij}) \perp (T_{ik}, \delta_{ik}) \,|\, (X_{ij}, X_{ik}, b_i)$.

Cox model with random effects

Assume that the conditional hazard of survival time Y is:

$$\lambda(t|X_{ij}, b_i) = \lambda_0(t)\exp(\beta' X_{ij} + b_i' Z_{ij}) \quad (j = 1, \ldots, n_i;\ i = 1, \ldots, q)$$

- b_i are (iid) s-dimensional random effects of mean zero and distribution depending on an unknown parameter θ.
- X_{ij} is a vector of covariates, Z_{ij} is a sub-vector of $(1, X_{ij}')'$.
- Assume random right censoring so we observe (T, δ), where $T = \min(Y, C)$ and $\delta = I(Y \leq C)$.
- Assume $b_i \perp X_{ij}$ and $(T_{ij}, \delta_{ij}) \perp (T_{ik}, \delta_{ik}) \,|\, (X_{ij}, X_{ik}, b_i)$.

The shared frailty model ($s = 1$ and $Z_{ij} = 1$):

$$\lambda(t|X_{ij}, b_i) = \lambda_0(t)\exp(\beta' X_{ij} + b_i) = \lambda_0(t) v_i \exp(\beta' X_{ij}),$$

where $v_i = \exp(b_i)$ is called frailty.

A good reference: Duchateau and Janssen (2008).

Three goodness-of-fit tests

Problem 1.

$$H_0 : \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left\{\beta' X_{ij} + b_i' Z_{ij}\right\}$$
$$H_1 : \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left\{m(X_{ij}) + b_i' Z_{ij}\right\}$$

Problem 2.

$$H_0 : \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left\{\beta' X_{ij} + b_i' Z_{ij}\right\}$$
$$H_1 : \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left\{\beta(t)' X_{ij} + b_i' Z_{ij}\right\}$$

Problem 3.

$$H_0 : \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left\{m_\theta(X_{ij}) + b_i' Z_{ij}\right\}$$
$$H_1 : \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left\{m(X_{ij}) + b_i' Z_{ij}\right\}$$

We are not aware of any significant contribution to these problems.

Some related literature

- Rich literature on testing linearity in the standard Cox model (Gray, 1994; Lin, Zhang and Davidian, 2008; among others), but with random effects the problem has not been considered so far.
- Xu, Vaida and Harrington (2009) use a profile-AIC and a profile-likelihood ratio test for model selection in the multivariate frailty model (testing for the significance of a specified subset of random or fixed effects).
- To capture the correct effect of the covariates on the conditional hazard Yu, Lin and Tu (2012) use smoothing splines. Yu and Lin (2008) use kernels (just one covariate, based on a marginal proportional hazard model).

Testing the linear covariate effects

- In this work we formulate a convenient version of the first testing problem:

$$H_0 \quad : \quad \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left(\sum_{k=1}^{d} \beta_k X_{ijk} + b_i' Z_{ij}\right)$$

$$H_1 \quad : \quad \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left(\sum_{k=1}^{p} \beta_k X_{ijk} + \sum_{k=p+1}^{d} m_k(X_{ijk}) + b_i' Z_{ij}\right),$$

  for some $0 \leq p \leq d-1$ given, where $m_k(\cdot)$ $(k = p+1, \ldots, d)$ are non-parametric, which are supposed to have mean zero.

- Our proposal[1] is a likelihood ratio test.
- Nonparametric estimation under the alternative is performed using orthogonal expansions.

[1] Just accepted in Biometrical Journal.

Next steps

1. Estimation under the null and the alternative
2. The likelihood ratio test
3. Simulations
4. Data application
5. Extensions

Estimation under the null model

- In the shared frailty model, with parametric baseline hazard and Gamma frailty, estimation can be performed maximizing the full marginal likelihood (the frailty is integrated out).
- In frailty models with unspecified baseline hazard direct maximization of the marginal likelihood is no longer possible.

- In the Cox model (without random effects) the regression coefficients are estimated using partial likelihood (PL).
- Ripatti and Palmgren (2000) suggest a penalized partial likelihood (PPL). This is much simpler but *some information might be lost*.

The full likelihood approach

We want to estimate $\xi = (\beta, \theta, \lambda_0)$

Suppose for the moment that the random effects b_i were observed.

Note that

$$f_{T,\delta,X,b} = f_{T,\delta|X,b}\, f_{b|X}\, f_X = f_{T,\delta|X,b}\, f_b\, f_X,$$

since b and X are independent. Also note that

- f_X does not depend on any of the parameters,
- f_b depends only on θ,
- $f_{T,\delta|X,b}$ gives rise to the classical partial likelihood of the Cox model.

The full likelihood approach

Hence, the likelihood is given by

$$\begin{aligned} L(\beta, \theta, \lambda_0) &= \left[\prod_{i=1}^{q} f_{T_{i1},\ldots,T_{in_i},\delta_{i1},\ldots,\delta_{in_i}|X_{i1},\ldots,X_{in_i},b_i}\right]\left[\prod_{i=1}^{q} f_{b_i}\right] \\ &= \left[\prod_{i=1}^{q}\prod_{j=1}^{n_i} f_{T_{ij},\delta_{ij}|X_{ij},b_i}\right]\left[\prod_{i=1}^{q} f_{b_i}\right]. \end{aligned}$$

since $(T_{ij}, \delta_{ij}) \perp (T_{ik}, \delta_{ik}) \,|\, (X_{ij}, X_{ik}, b_i)$, and the log-likelihood is

$$\begin{aligned} \log L &= \sum_{i=1}^{q}\sum_{j=1}^{n_i}\left\{\delta_{ij}\log\lambda_0(T_{ij}) + \delta_{ij}(\beta' X_{ij} + b_i' Z_{ij}) - \Lambda_0(T_{ij})\exp(\beta' X_{ij} + b_i' Z_{ij})\right\} \\ &\quad + \sum_{i=1}^{q}\log f(b_i|\theta) \end{aligned}$$

As the random effects b_i are not observed, this is an infeasible likelihood!

Full likelihood and the EM algorithm

We can use the EM algorithm to maximize the log-likelihood:

$$\begin{aligned}
\log L(\beta, \theta, \lambda_0) &= S_1(\beta, \lambda_0) + S_2(\theta) \\
S_1(\beta, \lambda_0) &= \sum\sum \Big\{ \delta_{ij} \log \lambda_0(T_{ij}) + \delta_{ij}(\beta' X_{ij} + b_i' Z_{ij}) \\
&\qquad - \Lambda_0(T_{ij}) \exp(\beta' X_{ij} + b_i' Z_{ij}) \Big\} \\
S_2(\theta) &= \sum \log f(b_i|\theta)
\end{aligned}$$

- The λ_0-function that maximizes the likelihood is concentrated at the uncensored failures times $t_1, \ldots, t_h$.
- Thus we can equivalently maximize the parametric log-likelihood where the unknown parameters are:

$$(\beta, \theta, \lambda_0(t_1), \ldots, \lambda_0(t_h))$$

- Start with initial parameter values: $\tilde{\xi} = (\tilde{\beta}, \tilde{\theta}, \tilde{\lambda}_0(t_1), \ldots, \tilde{\lambda}_0(t_h))$

The E-step

Calculation of

$$
\begin{aligned}
& E[\log L(\beta, \theta, \lambda_0) \mid \tilde{\xi}, D] \\
&= E[S_1(\beta, \lambda_0) \mid \tilde{\xi}, D] + E[S_2(\theta) \mid \tilde{\xi}, D] \\
&= \sum_{i=1}^{q} \sum_{j=1}^{n_i} \Big\{ \delta_{ij} \log \lambda_0(T_{ij}) + \delta_{ij}(\beta' X_{ij} + E[b_i \mid \tilde{\xi}, D]' Z_{ij}) \\
&\qquad\qquad -\Lambda_0(T_{ij}) \exp(\beta' X_{ij}) E[\exp(b_i' Z_{ij}) \mid \tilde{\xi}, D] \Big\} \\
&\quad + \sum_{i=1}^{q} E[\log f(b_i|\theta) \mid \tilde{\xi}, D] \\
&= Q_1(\beta, \lambda_0) + Q_2(\theta),
\end{aligned}
$$

conditional on the current parameter value $\tilde{\xi}$ and the observed data D.

The M-step

- Maximization of $Q_1(\beta, \lambda_0)$:
  Profile likelihood approach, as in the usual Cox model (with offsets $\log E[\exp(b_i' Z_{ij}) \mid \tilde{\xi}, D]$).

- Maximization of $Q_2(\theta)$:
  $Q_2(\theta) =$ log-likelihood of q independent observations with density $\exp\{E[\log f(b_i|\theta) \mid \tilde{\xi}, D]\}$

  $\Rightarrow$ $Q_2(\theta)$ can be maximized either explicitly or numerically depending on the density of the random effects

The E and M-steps should be iterated until convergence.

Feasibility of the EM algorithm

- The usefulness of the EM algorithm depends on two conditions: (1) *it should be easy to obtain expected values*, (2) *maximisation of the likelihood conditional on the expected values should be straightforward*.
- The conditional expectations in the E-step are in general not available in closed-form and s-dimensional numerical integration would be required[2]. An exception is the shared frailty model with Gamma frailty (E-step can be performed using closed-form expressions).
- In the M-step maximization is performed using partial likelihood ideas.

[2] For Normal random effects: Vaida and Xu (2000) suggest a MCMC method with Gibbs sampling; Abrahantes and Burzykowsky (2006) suggest a Laplace approximation (clusters need to be large).

A pseudo marginal likelihood approach

- In the shared frailty model Gorfine, Zucker and Hsu (2006) suggest an alternative algorithm.
- β and θ are estimated by maximizing the marginal likelihood:

$$IL(\beta, \theta, \lambda_0) = \prod_{i=1}^{q} \int \prod_{j=1}^{n_i} f_{T_{ij},\delta_{ij}|X_{ij},b_i} f_{b_i} \, \mathrm{d}b_i$$

- A step-function estimate of Λ_0 (integrated baseline hazard) is plugged-in at each iteration to simplify the maximization problem.
- The approach works for any frailty distribution with finite moments.
- Estimates are shown to be very close to those derived by the EM algorithm.

Software available for estimation under the null

There are several available R packages:

- Multivariate frailty model ($s \geq 1$):
  - coxme::`coxme` and survival::`coxph`. Estimation by PPL.
  - phmm::`phmm`. Full likelihood and MC-EM algorithm (Xu and Vaida, 2000).
- Only shared frailty model ($s = 1$)
  - frailtyEM::`emfrail`. Full likelihood and EM algorithm (Balan and Putter, 2017). Several frailty distributions. Right censoring and truncation.
  - frailtySurv::`fitfrail`. Pseudo-marginal likelihood (Gorfine et al., 2016). Several frailty distributions.
- Other approaches: frailtypack::`frailtyPenal` (multivariate frailty model, splines), frailtyHL::`frailtyHL` (hierarchical-likelihood), parfm::`parfm` (parametric baseline), survBayes::`survBayes`.

Estimation under the alternative model

$$\lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left(\sum_{k=1}^{p} \beta_k X_{ijk} + \sum_{k=p+1}^{d} m_k(X_{ijk}) + b_i' Z_{ij}\right)$$

- We use orthogonal expansions to estimate the m_k-functions.
- We approximate $m_k(x)$ by an expansion of the form

$$\sum_{\ell=1}^{r} \gamma_\ell u_\ell(x)$$

  for some known orthogonal basis functions $u_1, \ldots, u_r$.
- The same estimation approach as under the null model can be used, except that the model now contains more coefficients. We can use the same software.
- Examples of common basis functions are orthogonal polynomials or trigonometric functions.

Justification

- Orthogonal expansions can approximate arbitrarily well any continuous function with respect to a certain distance, as long as the number of basis functions r is taken sufficiently large.
- How to choose the number of basis functions r_k for the function m_k ?
- We use AIC:
  - Fit P^{d-p} models (take at most P basis functions for each k).
  - Select the model with the lowest AIC among these P^{d-p} candidate models.

The likelihood ratio test

- We consider the test statistic:

$$LR = -2\left\{\log L(\widehat{\beta}_{H_0}, \widehat{\theta}_{H_0}, \widehat{\lambda}_{0,H_0}|H_0) - \log L(\widehat{\beta}_{H_1}, \widehat{\gamma}_{H_1}, \widehat{\theta}_{H_1}, \widehat{\lambda}_{0,H_1}|H_1)\right\}$$

- To calibrate the test we use a model based bootstrap procedure that creates bootstrap samples satisfying the null hypothesis (resampling scheme extending Massonnet, Burzykowski and Janssen, 2006).
- For each bootstrap sample we recalculate the optimal number of basis functions $r^*_{p+1}, \ldots, r^*_d$ using the AIC.

  $\Rightarrow$ This leads to the bootstrap test statistic LR^*.

  This procedure is repeated B times leading to bootstrapped test statistics $LR^*_1, \ldots, LR^*_B$, and the critical value of the test at level α is then approximated by the $[(1-\alpha)B]$-th order statistic of these B values.

Resampling algorithm

1. Under H_0 fit the model and get the estimators $\widehat{\beta}_{H_0}, \widehat{\theta}_{H_0}$ and $\widehat{\lambda}_{0,H_0}$.
2. Draw i.i.d. random effects b_i^*, $i = 1, \ldots, q$, from their distribution with θ replaced by $\widehat{\theta}_{H_0}$.
3. Generate survival times Y_{ij}^* $(j = 1, \ldots, n_i,\ i = 1, \ldots, q)$ from the estimated survival function

$$\widehat{S}(\cdot|X_{ij}) = \widehat{S}_0(\cdot)^{\exp(\widehat{\beta}'_{H_0} X_{ij} + b_i^{*\prime} Z_{ij})},$$

   with $\widehat{S}_0(\cdot)$ the baseline survival obtained from $\widehat{\lambda}_{0,H_0}$ in step 1.
4. Generate censoring times C_{ij}^* $(j = 1, \ldots, n_i,\ i = 1, \ldots, q)$ from the Cox-regression estimator of the censoring distribution:

$$\widehat{G}(\cdot|X_{ij}) = \widehat{G}_0(\cdot)^{\exp(\widehat{\delta}' X_{ij})},$$

5. Set $T_{ij}^* = \min(Y_{ij}^*, C_{ij}^*)$ and $\delta_{ij}^* = I(T_{ij}^* \leq C_{ij}^*)$. The bootstrap sample is then $\{(T_{ij}^*, X_{ij}, \delta_{ij}^*);\ j = 1, \ldots, n_i,\ i = 1, \ldots, q\}$.

Simulations: Aims of the study

(i) To evaluate the type I error and power of our likelihood ratio test.

(ii) To compare our test with two possible competitors in terms of type I error and power.

(iii) To evaluate the sensitivity of our test to: misspecification of the frailty distribution, varying cluster sizes, and the dimension of the parameters.

(iv) To evaluate the performance of our estimator of the nonparametric covariate effect under the alternative, including a comparison with an estimator based on splines.

Scenario 1: Shared frailty model

Consider the following model under H_0 :

$$\lambda(t|X_{ij1}, X_{ij2}, b_i) = \lambda_0(t)\exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + b_i),$$

where

- $\exp(b_i) \sim Gamma(\text{mean}= 1, \text{variance}= \theta)$ with $\theta = 0.5$ or 2
- $X_1 \sim Be(0.5)$, $\beta_1 = 0.5$, $X_2 \sim Un[0, 1]$, $\beta_2 = 1$
- Total sample size $n = 300, 600$ or 1200
- Samples with q clusters and n_i observations per cluster, with $n_i = 5$ or 20
- Censoring distribution: (40-70% censoring)
  $\lambda_{cen}(c|X_{ij1}, X_{ij2}) = 0.4\exp(0.2X_{ij1} + 0.5X_{ij2})$
  and maximum follow-up time $= 5$

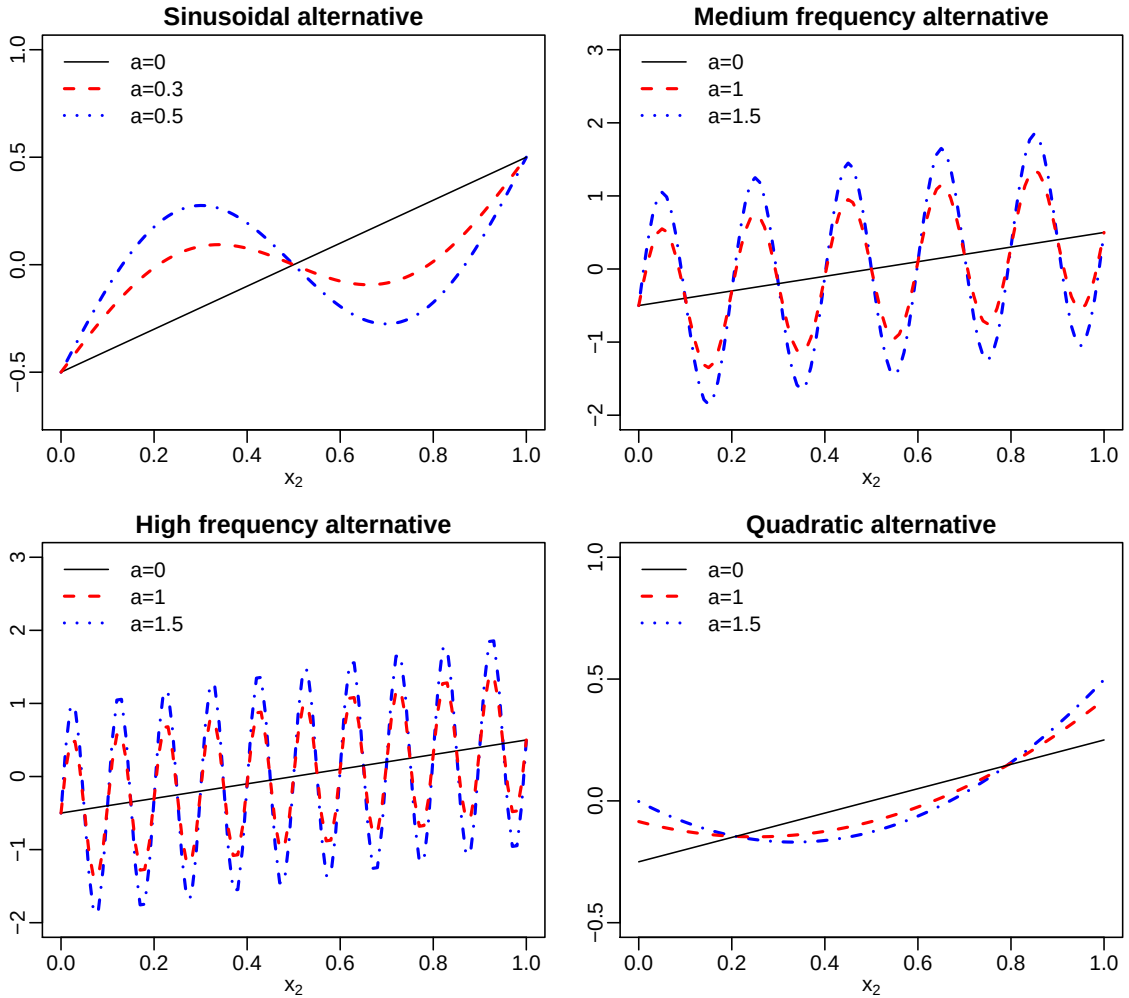
Empirical level of the likelihood ratio test

n	n_i	$\theta = 0.5$	$\theta = 2$
300	5	5.6 (0.52)	5.8 (0.51)
300	20	5.2 (0.51)	6.0 (0.51)
600	5	4.4 (0.50)	4.8 (0.49)
600	20	4.6 (0.50)	5.4 (0.47)
1200	5	5.1 (0.49)	5.7 (0.50)
	20	5.4 (0.50)	5.5 (0.50)

Table: Empirical level (%) of the test and average p-value (between brackets) under the shared frailty model with Gamma frailty. The nominal level is 5%.

Evaluation of the power under several alternatives

- Sinusoidal: $m(x_2) = \beta_2 x_2 + a\sin(b\pi x_2)$, with $b = 2, 10$ or 20
- Quadratic: $m(x_2) = (\beta_2 - a)x_2 + ax_2^2$



Empirical power of our test

		Alternative hypothesis							
		Sinusoidal		Medium freq.		High freq.		Quadratic	
n	n_i	$a=0.3$	$a=0.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$
					$\theta = 0.5$				
300	5	24.0	69.2	40.4	67.2	15.6	22.0	10.8	24.8
	20	38.4	76.8	44.4	62.4	14.8	22.8	17.6	29.2
600	5	48.0	92.8	72.0	95.6	28.4	48.8	21.6	38.4
	20	57.6	96.4	79.2	94.0	27.6	45.2	27.6	46.8
1200	5	89.4	98.6	95.7	100.0	50.0	63.4	41.2	73.5
	20	87.6	100.0	97.2	99.6	51.2	78.0	43.2	78.4
					$\theta = 2$				
300	5	27.6	54.0	32.4	54.0	12.4	19.2	9.5	16.4
	20	26.0	66.4	30.4	54.8	12.0	18.4	11.2	21.2
600	5	42.0	81.6	58.0	87.6	24.4	34.8	17.6	32.8
	20	45.6	90.0	70.0	91.2	27.6	42.0	17.1	40.0
1200	5	71.0	98.4	87.5	99.1	41.2	63.3	29.5	57.9
	20	78.8	99.6	96.4	99.6	49.2	69.4	41.2	70.8

Table: Percentage of rejections under the alternative.

Scenario 2: Multivariate frailty model

The second model is a frailty model with two independent Gaussian random effects.

$$\lambda(t|X_{ij1}, X_{ij2}, b_{i1}, b_{i2}) = \lambda_0(t)\exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + b_{i1} + b_{i2} X_{ij1}),$$

where

- $b_{i1}, b_{i2} \sim N(0, 0.25)$, b_{i1} and b_{i2} are independent
- $X_1 \sim Be(0.5)$, $\beta_1 = 0.5$
- $X_2 \sim Un[0, 1]$, $\beta_2 = 1$
- $n = 300$ and 600, with clusters of size $n_i = 5$ (computations more intense than before)
- $C \sim \text{Exp}(\lambda = 0.4)$, maximum follow-up time $= 5$

Empirical level and power of our test

Alternatives :

- $m(x_2) = \beta_2 x_2 + a\sin(b\pi x_2)$, with $b = 10$ or 20
- $m(x_2) = (\beta_2 - a)x_2 + a x_2^2$

		Medium freq.	High freq.	Quadratic	
n	Null hypothesis	$a = 1.5$	$a = 1.5$	$a = 1$	$a = 1.5$
300 ($n_i = 5$)	5.9 (0.51)	59.7	16.5	9.3	19.3
600 ($n_i = 5$)	4.0 (0.54)	92.8	39.7	15.6	36.4

Table: Empirical level and power of our test under a multivariate frailty model with two independent Normal random effects. The nominal level is 5%.

Two possible competitors for our test

- **Competitor 1**: Same likelihood ratio test but using standard Cox regression estimates (ignoring the correlation).
- **Competitor 2**: Same likelihood ratio test but with parametric alternative. For linear null hypothesis estimate the alternative using an orthogonal expansion with e.g. three basis functions.

We perform a comparison between our test and each of these competitors under the shared frailty model (scenario 1).

Competitor 1: Power comparison

		Alternative hypothesis							
		Sinusoidal		Medium freq.		High freq.		Quadratic	
n	n_i	$a=0.3$	$a=0.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$
					$\theta = 0.5$				
300	5	0.85	0.86	0.82	0.90	0.77	0.93	0.85	0.82
	20	0.64	0.76	0.82	0.92	0.89	0.77	0.73	0.75
600	5	0.80	0.88	0.89	0.94	1.00	0.95	0.85	0.95
	20	0.64	0.85	0.82	0.97	0.78	0.88	0.70	0.81
1200	5	0.84	1.00	0.96	1.00	0.90	1.16	0.88	0.88
	20	0.83	1.00	0.95	1.00	0.88	0.94	0.75	0.78
					$\theta = 2$				
300	5	0.51	0.44	0.51	0.55	0.77	0.50	0.84	0.63
	20	0.54	0.49	0.57	0.53	0.63	0.50	0.61	0.53
600	5	0.49	0.54	0.51	0.63	0.64	0.71	0.66	0.57
	20	0.46	0.45	0.49	0.66	0.45	0.50	0.75	0.51
1200	5	0.45	0.75	0.61	0.88	0.45	0.59	0.47	0.48
	20	0.43	0.77	0.62	0.88	0.52	0.61	0.38	0.44

Table: Power of the competitor divided by the power of our proposal.

Competitor 1: Empirical level

n	n_i	$\theta = 0.5$	$\theta = 2$
300	5	5.2 (0.51)	3.8 (0.51)
300	20	5.6 (0.51)	6.6 (0.49)
600	5	4.4 (0.52)	6.2 (0.47)
600	20	4.8 (0.49)	6.2 (0.49)
1200	5	4.6 (0.53)	5.2 (0.50)
1200	20	8.2 (0.48)	7.2 (0.49)

Table: Empirical level (%) of the test and average p-value (between brackets) under the shared frailty model. Nominal level is 5%.

$\rightarrow$ For large θ the empirical level is above the nominal level.

Competitor 2: Power comparison

		Alternative hypothesis							
		Sinusoidal		Medium freq.		High freq.		Quadratic	
n	n_i	$a=0.3$	$a=0.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$
					$\theta = 0.5$				
300	5	1.03	1.00	0.75	0.77	0.64	0.67	0.89	0.77
	20	0.95	0.99	0.71	0.78	0.92	0.82	0.89	0.84
600	5	0.95	1.01	0.83	0.89	0.68	0.61	0.85	0.88
	20	0.97	1.01	0.86	0.92	0.67	0.64	0.88	0.90
1200	5	0.99	1.01	0.95	0.99	0.71	0.75	0.82	0.90
	20	1.00	1.00	0.94	1.00	0.62	0.71	0.94	0.88
					$\theta = 2$				
300	5	0.96	1.03	0.75	0.81	0.65	0.69	1.05	0.88
	20	1.02	0.99	0.78	0.77	0.87	0.74	0.82	0.89
600	5	1.05	1.02	0.82	0.87	0.77	0.76	0.73	0.76
	20	0.94	1.02	0.84	0.85	0.86	0.75	1.03	0.81
1200	5	0.99	1.01	0.84	0.95	0.68	0.72	0.80	0.91
	20	1.01	1.00	0.87	0.99	0.74	0.73	0.82	0.90

Table: Power of the competitor divided by the power of our proposal.

Sensitivity to frailty distribution misspecification

		Alternative hypothesis					
		Sinusoidal		Medium freq.		High freq.	
n	n_i	$a=0.3$	$a=0.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$
				$\theta = 0.5$			
300	5	1.03	1.01	1.01	0.98	1.10	0.91
300	20	1.00	1.00	0.92	0.93	1.10	0.91
600	5	0.98	0.99	0.99	0.99	1.01	1.00
600	20	0.94	0.98	0.96	0.97	0.98	0.93
1200	5	0.99	1.01	0.99	1.00	0.90	0.94
1200	20	1.01	1.00	0.99	0.99	0.94	0.91
				$\theta = 2$			
300	5	1.10	0.89	0.86	0.97	1.00	0.98
300	20	0.90	0.91	0.89	0.86	0.99	1.00
600	5	0.88	0.92	0.94	0.97	0.93	1.01
600	20	0.85	0.90	0.94	1.01	0.93	1.01
1200	5	0.90	0.97	0.85	0.99	0.89	0.87
1200	20	0.98	1.00	0.94	0.99	0.84	0.90

Table: Power of our test with misspecified frailty distribution divided by the power of the test with correctly specified frailty distribution.

Clusters with varying sizes

		Alternative hypothesis					
		Sinusoidal		Medium freq.		High freq.	
n	n_i	$a=0.3$	$a=0.5$	$a=1$	$a=1.5$	$a=1$	$a=1.5$
				$\theta = 0.5$			
300	5	1.28	0.96	1.04	1.01	1.03	1.15
	20	1.01	0.98	0.95	1.06	1.00	0.98
600	5	0.93	1.00	1.03	0.99	1.00	1.01
	20	0.99	1.01	0.99	1.00	1.13	1.01
1200	5	0.96	1.01	0.95	1.00	0.97	1.16
	20	0.99	1.00	1.00	1.00	1.07	1.09
				$\theta = 2$			
300	5	0.81	0.97	0.93	0.93	1.32	1.10
	20	1.20	0.99	1.14	1.03	1.03	1.15
600	5	0.89	1.01	1.01	1.00	0.85	1.06
	20	1.04	0.98	0.98	1.00	1.07	1.07
1200	5	1.19	0.99	0.98	1.00	1.07	0.93
	20	0.95	1.00	1.00	1.00	0.98	1.01

Table: Power of our test with varying cluster sizes divided by the power with fixed cluster sizes.

Effect of the dimension

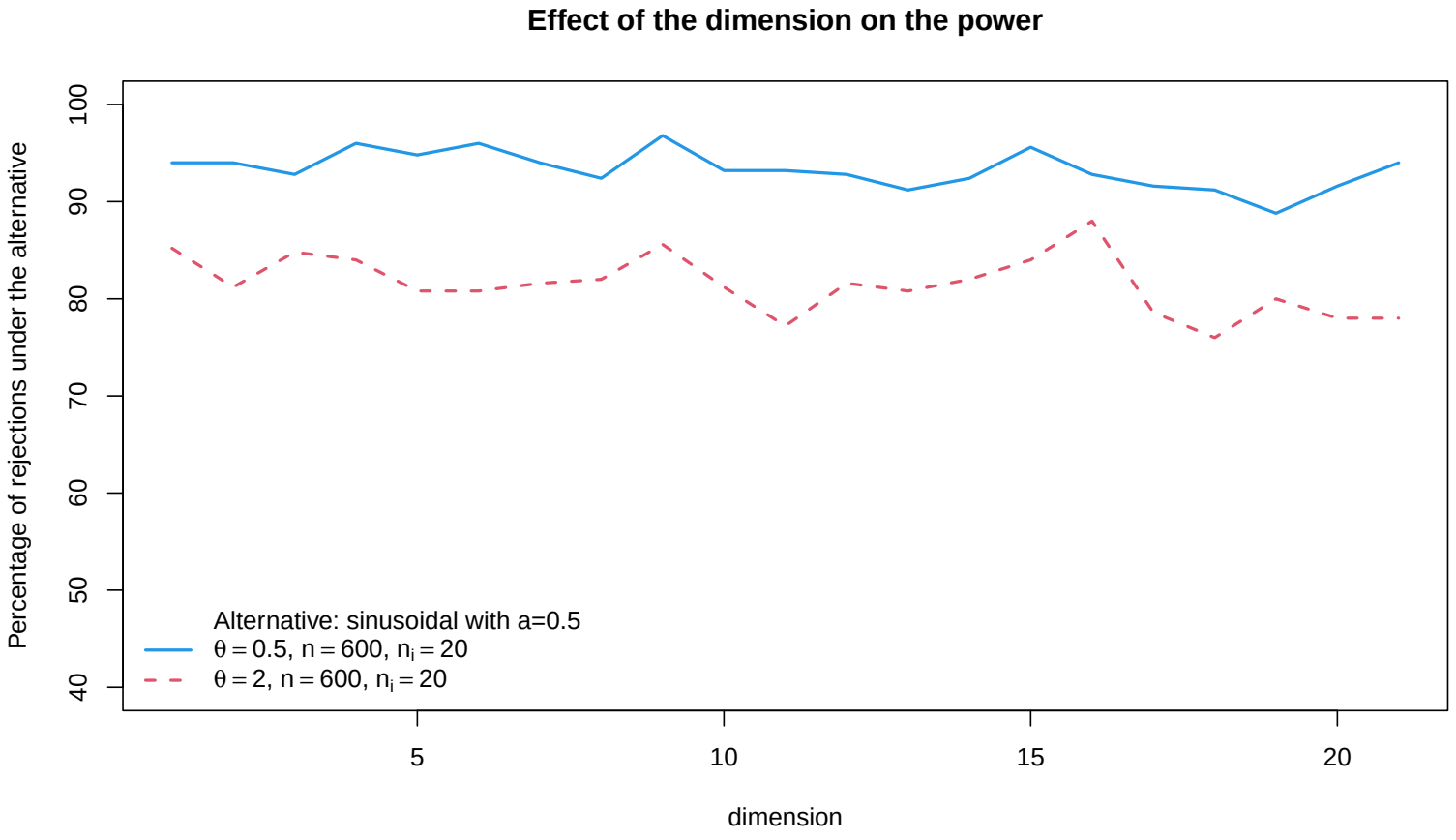


Figure: Effect of the dimension on the power of our test. The curves are the percentages of rejections under the alternative for two different settings.

Evaluating the estimation under the alternative

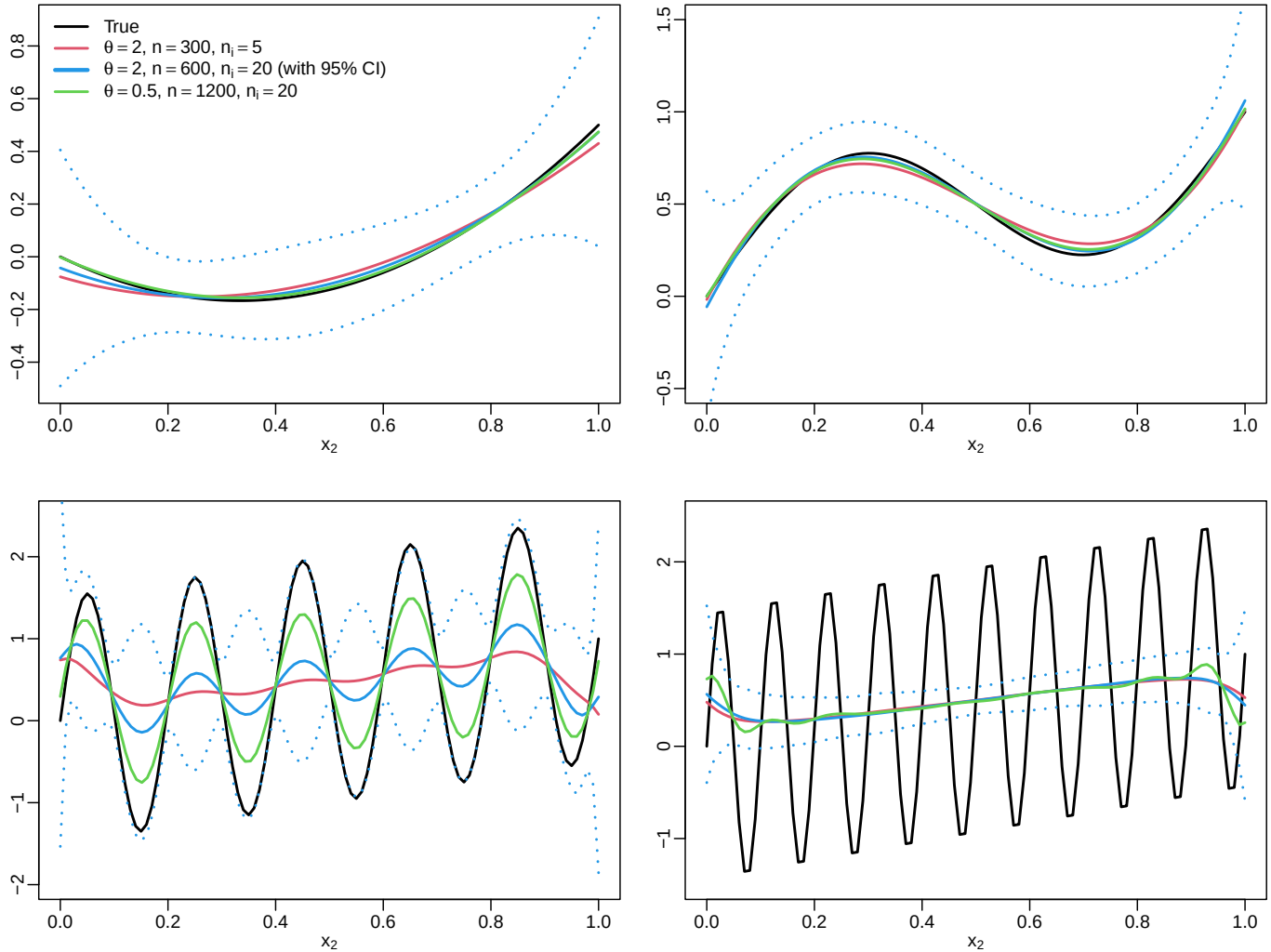


Figure: Average estimates (with 95% confidence bands) using our nonparametric estimator based on a orthogonal representation.

Comparison with the splines approach by Lin et al. (2012)

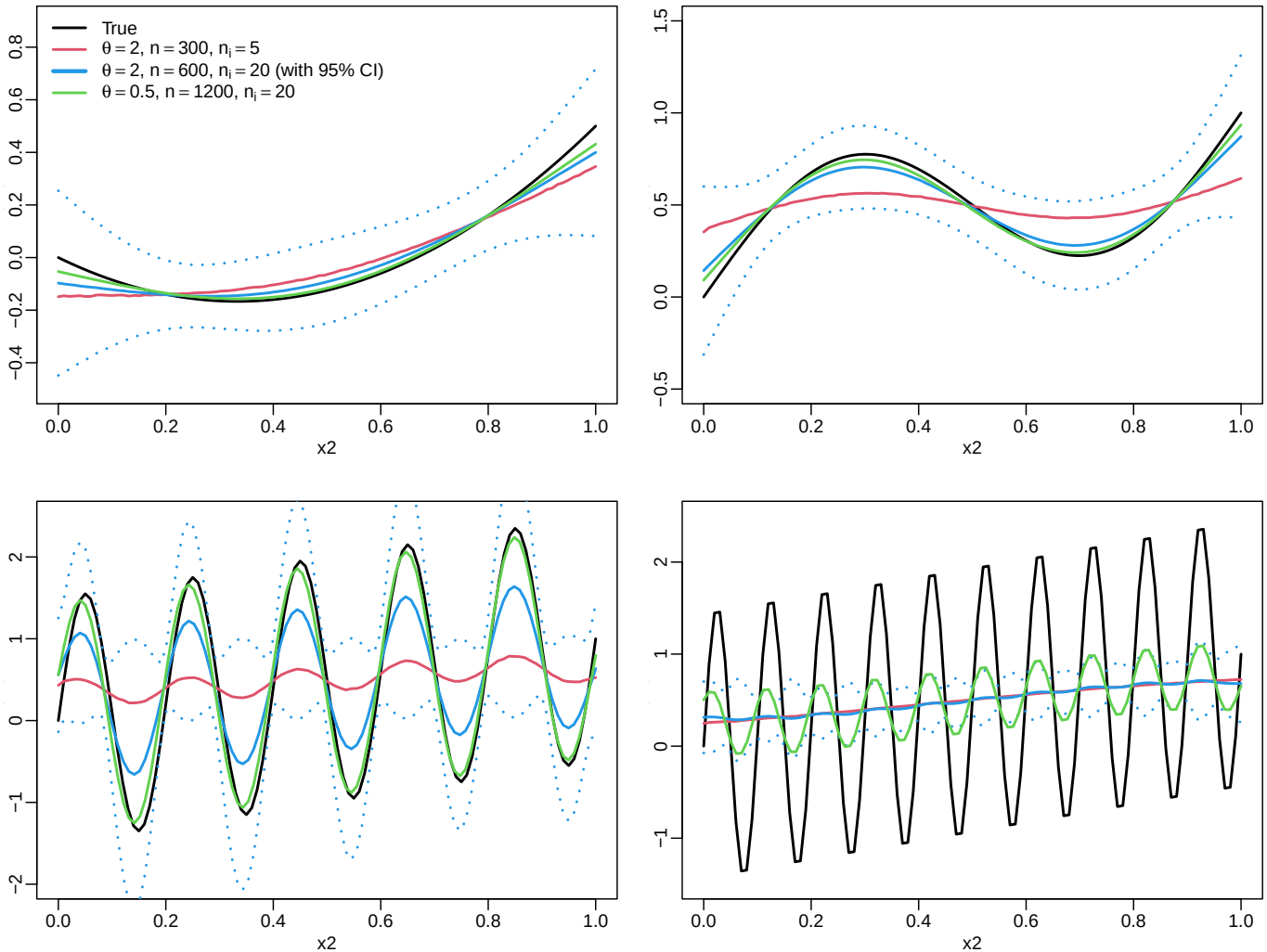


Figure: Average estimates (with 95% confidence bands) using splines.

Data example

Consider data from a randomized trial on chronic granulotomous disease (CGD) :

- $q = 128 =$ number of patients
- For each patient i :
  - $n_i =$ number of records (at least 1)
  - $Y_{ij} =$ gap time (days) between $(j-1)$-st and j-th infection
- Sample size $n = 203$. Censoring percentage is 62% (time interval does not finish with one infection)
- Patients were randomized to either gamma interferon or placebo

The data are shown in Appendix D2 of Fleming and Harrington (1991). Available also in the R-package survival (cgd).

Data example

Covariates :

- X_{ij1} = treatment (binary)
- X_{ij2} = pattern of inheritance (binary)
- X_{ij3} = use of corticosteroids (binary)
- X_{ij4} = use of prophylactic antibiotics (binary)
- X_{ij5} = gender (binary)
- $X_{ij6}, X_{ij7}, X_{ij8}$ = hospital category (four categories from which three binary covariates are created)
- X_{ij9} = age (continuous)

Vaida and Xu (2000) analysed these data using a shared frailty model

$$\lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left(\sum_{k=1}^{9} \beta_k X_{ijk} + b_i\right),$$

with $b_i \sim N(0, \theta)$.

Data example: model and testing problem

Why shared-frailty model?

- The risk of recurrent infection remains constant regardless of the number of previous infections.
- Times between infections for a patient may be correlated.

Our goal: To test whether age has indeed a linear effect (assumed by Vaida and Xu, 2000):

$$H_0: \quad \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left(\sum_{k=1}^{9} \beta_k X_{ijk} + b_i\right)$$

$$H_1: \quad \lambda(t|X_{ij}, b_i) = \lambda_0(t) \exp\left(\sum_{k=1}^{8} \beta_k X_{ijk} + m_9(X_{ij9}) + b_i\right).$$

Data example: estimation and testing results

Estimated coefficients under the null hypothesis :

	trtmt	inherit	cortico	prophy	gender	hosp1	hosp2	hosp3	age
$\widehat{\beta}$	1.14	0.82	-1.97	0.95	-0.96	-0.27	-1.09	-0.94	-0.04
sd	0.34	0.37	0.96	0.46	0.51	0.40	0.61	0.59	0.02
lower	0.47	0.09	-3.85	0.04	-1.96	-1.06	-2.30	-2.10	-0.08
upper	1.80	1.56	-0.08	1.86	0.04	0.52	0.11	0.21	-0.01

Estimated variance of the frailty: $\widehat{\theta}_{H_0} = 0.6$

Estimated coefficients under the alternative hypothesis :

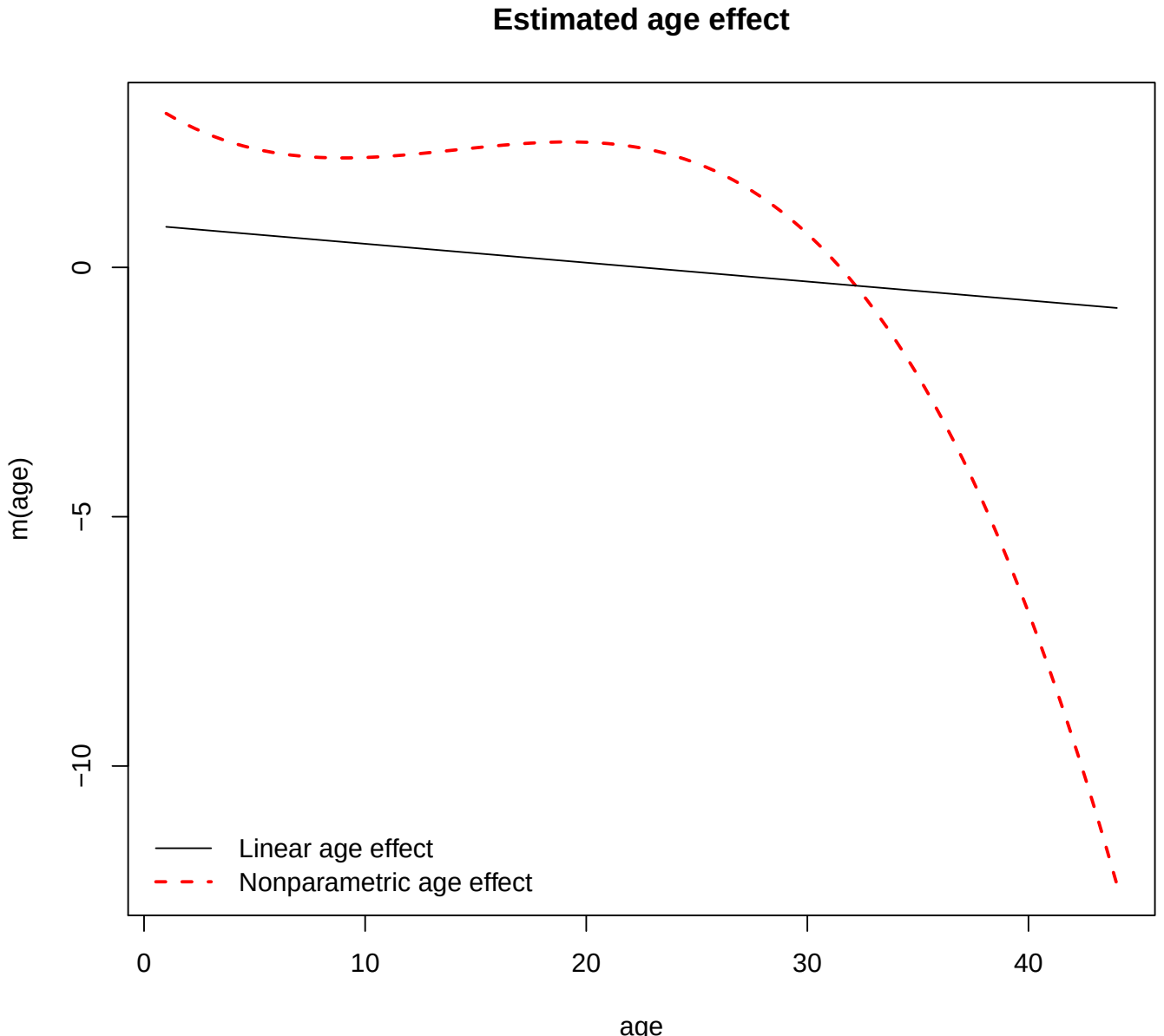
									age		
	trtmt	inherit	cortico	prophy	gender	hosp1	hosp2	hosp3	u_1	u_2	u_3
$\widehat{\beta}$	1.03	1.00	-1.94	1.13	-1.08	-0.28	-1.18	-0.91	-12.6	-12.5	-10.8
sd	0.31	0.37	0.78	0.43	0.49	0.39	0.59	0.57	4.87	6.29	4.92
lower	0.41	0.27	-3.48	0.28	-2.03	-1.04	-2.34	-2.02	-22.1	-24.8	-20.5
upper	1.64	1.73	-0.40	1.98	-0.12	0.48	-0.03	0.21	-3.01	-0.17	-1.16

Test statistic : $LR = 7.168$

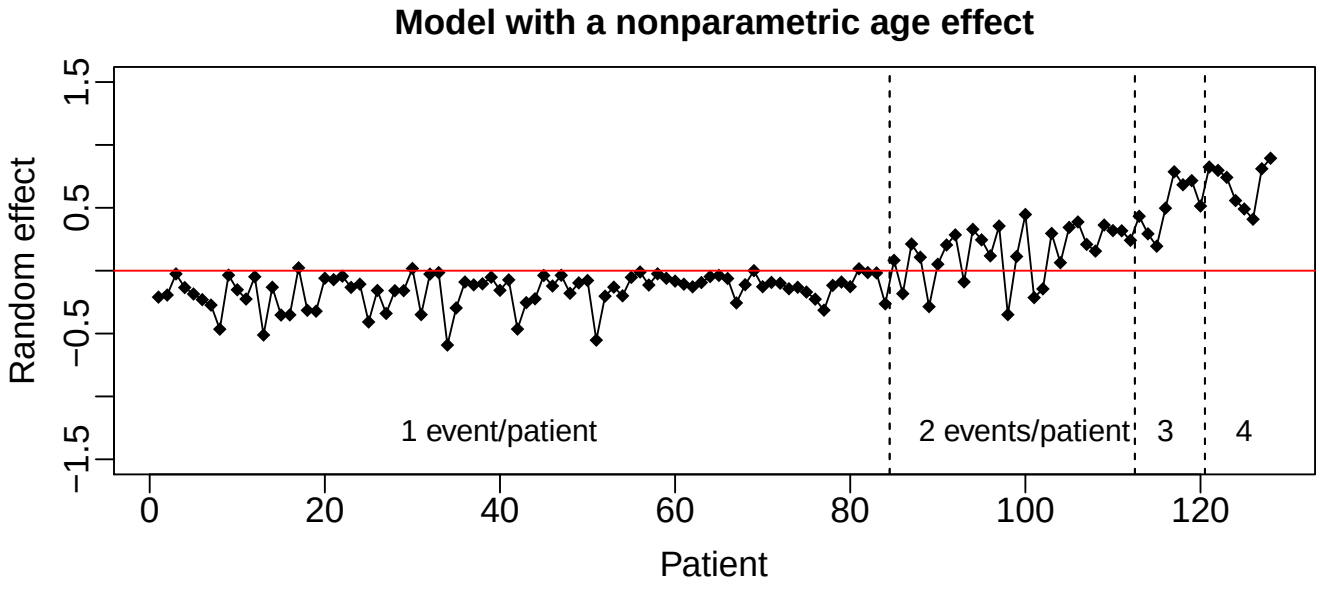
P-value $= 0.10$ based on 500 bootstrap samples

$\Rightarrow$ We do no have evidence to reject H_0 at the 5% level. Results are not conclusive at 10% level.

Estimated age effect under the null and the alternative



Data example: estimated random effects



- Random effects of patients with different number of infections are different.
  ⇒ Patients with more infections are different from those with fewer infections, in a way not explained by the covariates in the study.
- The variable "number of previous infections" is "not significant" in the model.
  ⇒ Previous infections do not increase the risk of future infections.

Conclusions

- Development of a goodness-of-fit test for the functional form of the covariate effects in a Cox model with random effects.
- Approach based on the full likelihood.
- Under the alternative we estimate the covariate effects non-parametrically using orthogonal expansions.
- Computations can be performed in R using available packages (e.g. frailtySurv, phmm).
- Simulations show that the proposed bootstrap calibration works well in practice.
- Simulations show that the test is not affected by the misspecification of the frailty distribution, and the dimension of parameters.

Extensions

- Some other appealing models for goodness-of-fit testing:
  - Accelerated failure time model with random effects:

$$\log T_{ij} = m(X_{ij}) + b_i' Z_{ij} + \epsilon_{ij}$$

  - Additive risk model with random effects:

$$\lambda\left(t|X_{ij}, b_i\right) = \lambda_0(t) + m(X_{ij}) + b_i' Z_{ij}$$

  - Proportional odds model with random effects:

$$\frac{Pr\left(Y \leq t|X_{ij}, b_i\right)}{Pr\left(Y > t|X_{ij}, b_i\right)} = \exp\left\{\alpha(t) + m(X_{ij}) + b_i' Z_{ij}\right\}$$

- Explore an extension of the goodness-of-fit test of González-Manteiga et al. (2016) under the formulation of López-de-Ullibarri, Janssen and Cao (2012).

References

1. Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer-Verlag.
2. González-Manteiga, W., Martínez-Miranda, M.D. and Van Keilegom, I. (2016). Goodness-of-fit test in parametric mixed effects models based on the estimation of the error distribution. *Biometrika* **103**, 133–146.
3. Gorfine, M., Zucker, D.M. and Hsu, L. (2006). Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *Biometrika* **93**, 735–741.
4. Gray, R.J. (1994). Spline based tests in survival analysis. *Biometrics* **50**, 640–652.
5. Lin, J., Zhang, D. and Davidian, M. (2008). Smoothing Spline-Based Score Tests for Proportional Hazards Models. *Biometrics* **62**, 803–812.
6. López-de-Ullibarri, I., Janssen, P. and Cao, R. (2012). Continuous covariate frailty models for censored and truncated clustered data. *J. Statist. Plann. Inference* **142**, 1864–1877.
7. Massonnet, G., Burzykowski, T. and Janssen, P. (2006). Resampling Plans for Frailty Models. *Commun. Stat. - Simul. Comput.* **35**, 497–514.
8. Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022.
9. Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Stat. Med.* **19**, 3309–3324.
10. Xu, R. Vaida, F. and Harrington, D.P. (2009). Using profile likelihood for semiparametric model selection with application to proportional hazards mixed models. *Stat. Sin.* **19**, 819–842.
11. Yu, Z. and Lin, X. (2008). Nonparametric Regression Using Local Kernel Estimating Equations for Correlated Failure Time Data. *Biometrika* **95**, 123–137.
12. Yu, Z., Lin, X. and Tu, W. (2012). Semiparametric Frailty Models for Clustered Failure Time Data. *Biometrics* **68**, 429–436.