Survival Analysis Under Measurement Error

Taipei

December 14th, 1999

Thomas Augustin

University of Munich,

thomas@stat.uni-muenchen.de

Contents

1. Introduction

Survival analysis

- $T_i > 0$, $i = 1, \ldots, n$:
  Time until a certain state is left (for the first time)

- Here: The distribution of T_i is assumed to be absolutely continuous.

- Equivalent description by
  * density $f_{T_i}(t)$
  * distribution function $F_{T_i}(t)$
  * survivor function $S_{T_i}(t) = 1 - F_{T_i}(t)$
  * hazard rate, intensity

$$\begin{aligned}\lambda_{T_i}(t) &\stackrel{def}{=} \lim_{h \downarrow 0} \frac{P_{T_i}(t \leq T_i < t + h | \{T_i \geq t\})}{h} \\ &= \frac{f_{T_i}(t)}{S_{T_i}(t)}\end{aligned}$$

  * cumulative hazard rate

$$\Lambda_{T_i}(t) \stackrel{def}{=} \int_0^t \lambda_{T_i}(u)\, du = -\ln(S(t))$$

Duration times $T_i > 0$, $i = 1, \ldots, n$ are assumed to be absolutely continuous.

Covariates X_i, Z_i

non-/semiparametric regression modeling

versus

parametric regression modeling

Usual choice: semiparametric, Cox model (Cox, 1972)

$$\lambda_{T_i}(t|X_i, Z_i) = \lambda_0(t) \cdot \exp\left(\beta_X \cdot X_i + \beta_Z' \cdot Z_i\right)$$

- very flexible! $\lambda_0(t)$ unspecified nuisance function

- Proportional hazard assumption

Some Arguments against an Exclusive Use of the Cox Model

- very sensitive to violations of the proportional hazard assumption.

- Material background often suggests non-proportional hazards.

- Parametric models may allow a more detailed formulation of the knowledge.

- Parametric models are more powerful, if their assumptions are met.

- The baseline hazard rate is not always only a nuisance(-function).

  * Construction of typical time dependency based on material considerations

  * Testing of hypotheses on the time dependency (e.g. does the average risk increase or decrease in time?)

2. Parametric Duration Models

- Renaissance of parametric duration models: distribution of T_i known except some parameters $\theta_1, \ldots, \theta_q$
- $$\theta_1 = \exp\left(\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i\right) .$$

Weibull Model

- The hazard rates have the form

  $$\lambda\,(t|X_i, Z_i) = \alpha{\cdot}\lambda^{\alpha}{\cdot}t^{\alpha-1}{\cdot}\exp\left((\lambda \cdot t)^{\alpha}\right) , \quad \alpha > 0 ,$$

  with

  $$\lambda = \exp\left(\beta_0 + \beta_X \cdot X_i + \beta_Z' \cdot Z_i\right) .$$

- Also proportional hazards among different individuals

- The monotonicity of the hazard rate remains unchanged over time.

  * $\nu > 1$: monotonely increasing ($\nu = 2$ Rayleigh)
  * $\nu < 1$: monotonely decreasing
  * $\nu = 1$: constant (exponential distribution)
  * quite popular in technometrics e.g. as a model for failure of machines and in economic sciences as a model for unemployment spells.

Log-logistic model

- with hazard rates ($\nu > 2$)

$$\lambda\left(t_i | X_i, Z_i\right) = $$

$$= \quad \frac{\nu}{t\left\{1 + t^{-\nu} \exp\left(\nu\left(\beta_0 + \beta_X \cdot X_i + \beta_Z' \cdot Z_i\right)\right)\right\}}$$

- no proportional hazards, instead

- non-monotonic shape with a single maximum depending on the values of the covariates,

- for t going to infinity the hazard ratio for two individuals tends to unity.

- Mainly popular in sociology (for instance for modelling divorce rates).

Accelerated failure time model

Common superstructure upon most parametric duration models:

$$\ln T_i = \beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i + \alpha \cdot \epsilon_i\,, \qquad \alpha > 0,$$

- ϵ_i i.i.d.
  independent of the covariates
  possibly depending on a parameter γ

- All the models mentioned can be produced by appropriate choices of ϵ_i.

- In general $\mathbb{E}\left[\epsilon_i\right] \neq 0$.

Where does the name 'accelerated failure time model' come from?

Typical form (cf. Kalbfleisch& Prentice (1980, p. 34)

$$\lambda_{T_i}(t|X_i, Z_i) = h_0\left(t \cdot \exp\left(\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i\right)\right) \cdot \\ \cdot \exp\left(\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i\right)$$

where the specific shape of $h_0(\cdot)$ is characteristic for the distribution under consideration.

3. Measurement Error

The available variables are only inexact

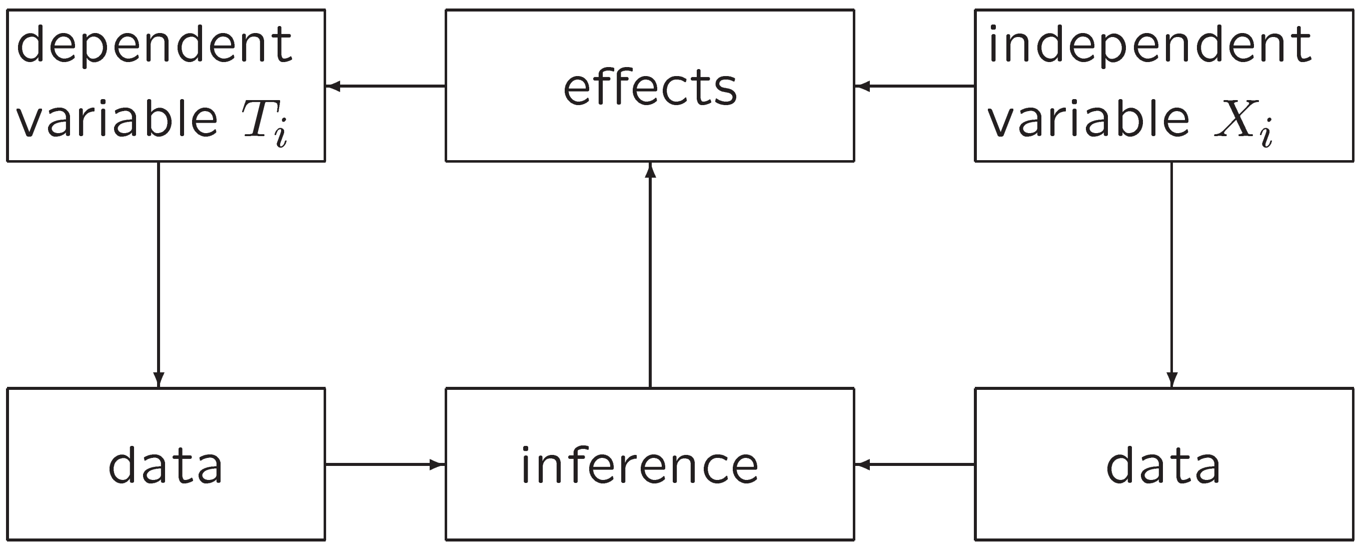
- measurements of a true quantity (physics)
- operationalisations of theoretic constructs (sociology/psychology)
- answers to questions on facts, behavioral patterns or attitudes (methodology of questionnaires)

This has to be taken into account. Ignoring measurement error may lead to severely biased parameter estimates.

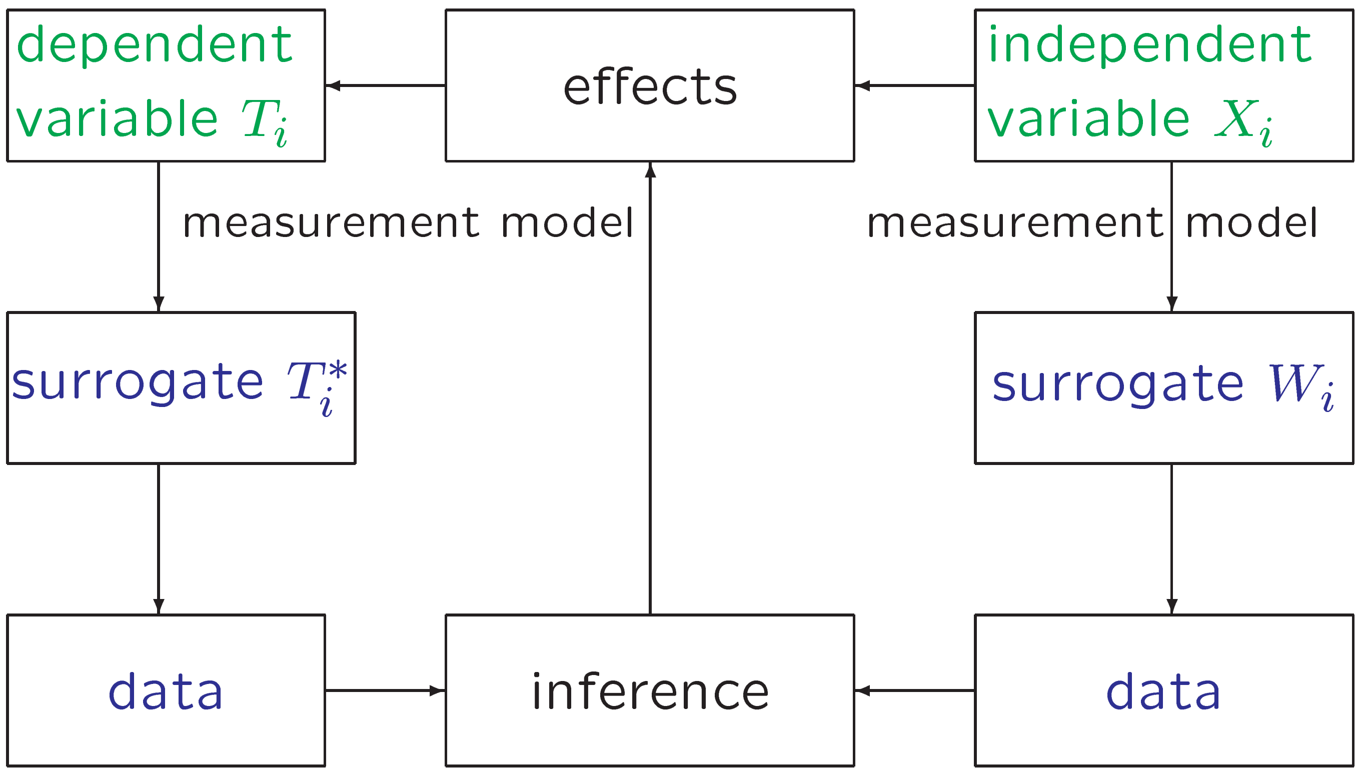
Basic references:

- Schneeweiß & Mittag (1986),
- Fuller (1987),
- Carroll, Ruppert & Stefanski (1995),
- Cheng & van Ness (1999).

Usual regression:



Regression with measurement error:



Covariate measurement error. Detailed assumptions

[A1] Only one covariate, X_i say, is measured with non-negligible error.

[A2] T_i is measured without error.

[A3] Additive measurement error:

$$W_i = X_i + U_i.$$

[A4] U_i is independent of T_i, X_i, Z_i and U_j, $j \neq i$.

[A5] Normal measurement error: $U_i \sim \mathcal{N}(0, \sigma_U^2)$ with σ_U^2 known.

[A6] Structural model: X_i is stochastic.

[A7] The conditional distribution of X_i given W_i and Z_i does not depend on Z_i.

These assumptions imply that the measurement error is 'nondifferential'. T_i and W_i are conditionally independent given X_i and Z_i.

Additionally, an appropriate class of parametric distributions for X has to be chosen:

$$X_i \sim ??$$

Modeling the distribution of X_i, $i = 1, \ldots, n$

Often the empirical distribution of W has several modes.

- Assume the population to be divided into m groups.
- Conditional on being in group j: $X_i \sim \mathcal{N}(\mu_j, \sigma_j^2)$, density $\varphi(\cdot \,\|\mu_j, \sigma_j^2)$
- κ_j: probability to belong to group j
- overall distribution: mixed normal

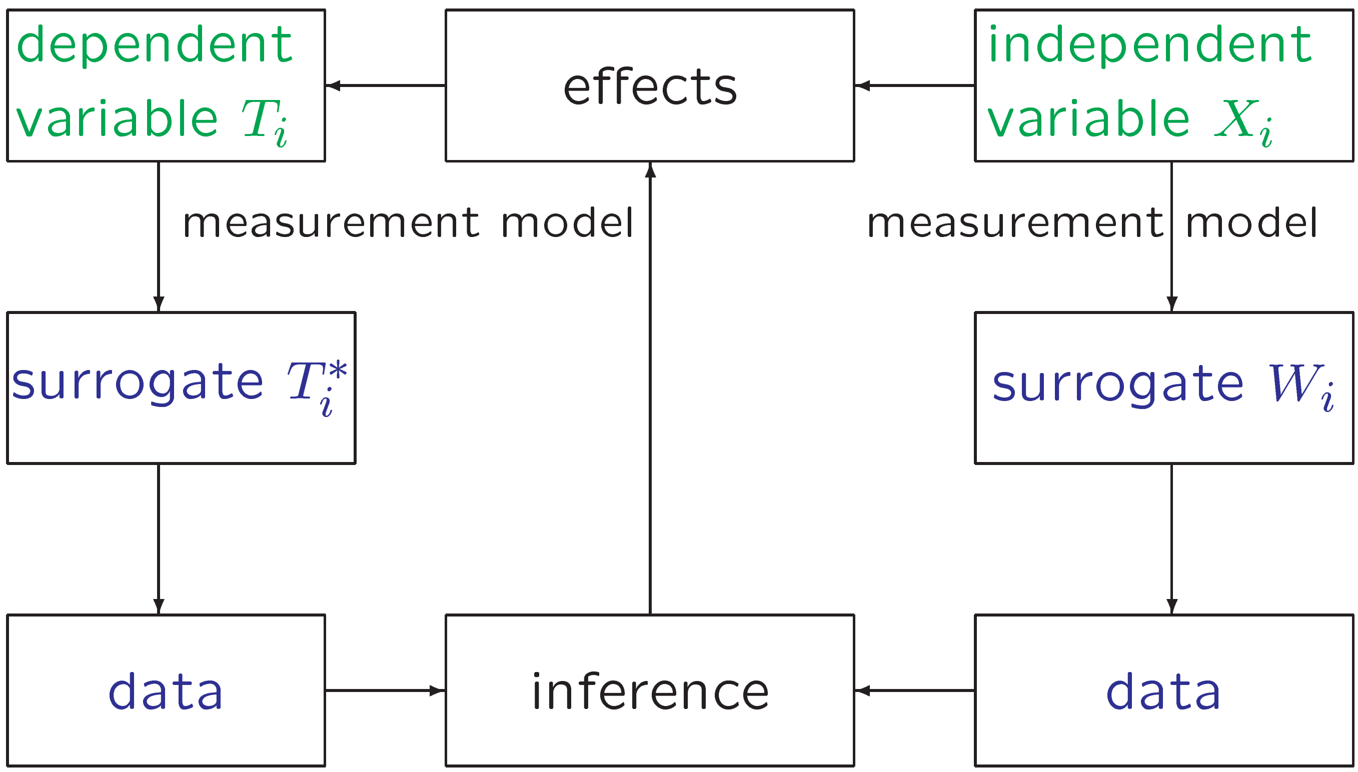
$$X_i \sim \mathcal{MIXN}\Big(m;\, \kappa_1, \ldots, \kappa_m;\; \mu_1, \ldots, \mu_m;\, \sigma_1^2, \ldots, \sigma_m^2\Big)$$

- overall density: $f_{X_i}(\cdot) = \sum_{j=1}^{m} \kappa_j \cdot \varphi(\cdot \,\|\mu_j, \sigma_j^2)$
- For every function $g(\cdot)$

$$\mathbb{E}[g(X_i)] = \sum_{j=1}^{m} \kappa_j \int_0^{\infty} g(x) \cdot \varphi\left(x \,\|\mu_j, \sigma_j^2\right)\, dx$$

- highly flexible, but nevertheless mathematically tractable

Regression with measurement error:



4. Quasi-Likelihood Based Correction for Covariate Measurement Error

4.1 Basic principles

- Problem: estimation of $\theta = (\alpha, \gamma^T, \beta_0, \beta_X, \beta_Z^T)^T$
  - * 'Ideal model': in terms of X_i, Z_i
  - * Data: realizations of W_i, Z_i
- impossible to calculate the 'data-based (conditional) likelihood'

$$\text{Lik}(\theta \| T_i, W_i, Z_i)$$

from the 'ideal (conditional) likelihood'

$$\text{Lik}(\theta \| T_i, X_i, Z_i)\,.$$

- Therefore: Search for another general estimation procedure.
- Idea: Use the quasi-likelihood approach
  - * In general: Wedderburn (1974), McCullagh (1983), McCullagh, (1991), Heyde (1997)
  - * For error in variables models outside survival analysis: Armstrong (1985),

Caroll, Ruppert, & Stefanski (1995; Section 7.8 and Appendix A.4), Thamerus (1997, 1998).

* Estimate the vector θ of unknown parameters from an i.i.d. sample $T_1, \ldots, T_n$ by using
  * a mean function $m_i(\theta)$ and
  * a variance function $v_i(\theta)$

  and by solving the quasi-score equation

$$\sum_{i=1}^{n} \frac{\partial\, m_i(\theta)}{\partial\, \theta} \cdot \frac{(T_i - m_i(\theta))}{v_i(\theta)} = 0\,.$$

Three Steps to Motivate Quasi-Likelihood Estimation

1. The linear model in the classical sense

- $$Y_i = x_i'\beta + \epsilon_i\,, \qquad [\epsilon_i] = \mathcal{N}\left(0, \sigma^2\right)$$

  which means:

  $$[Y_i] = \mathcal{N}\left(m_i(\beta), v\right)$$

  with

  $$\begin{array}{rcccl} m_i(\beta) & \stackrel{def}{=} & \mathbb{E}(Y_i) & = & x_i'\beta \\ v & \stackrel{def}{=} & Var(Y_i) & = & \sigma^2 \end{array}$$

- Maximum-likelihood estimation and least squares estimation coincide.

  $$\sum_{i=1}^{n} \left(Y_i - m_i(\beta)\right)^2 \to \min \qquad (*)$$

  $$\iff \quad \sum_{i=1}^{n} \frac{\partial\, m_i(\beta)}{\partial\, \beta_j} \cdot \left(Y_i - m_i(\beta)\right) = 0\,, \quad \forall j = 1, \ldots, p \qquad (**)$$

- Even if the data are not normally distributed, the estimates are still reasonable. (Gauss-Markov Theorem!).

2. Generalization to heteroscedasticy

- Now: $\mathbb{V}\left(Y_i\right) = v_i$

- Maximum-likelihood estimation is equivalent to weighted least squares estimation.

$$\sum_{i=1}^{n} \frac{1}{v_i}\left(Y_i - m_i\left(\beta\right)\right)^2 \to \text{min} \qquad (*)$$

$$\Longleftrightarrow \quad \sum_{i=1}^{n} \frac{\partial\, m_i\left(\beta\right)}{\partial\, \beta_j} \cdot \frac{\left(Y_i - m_i\left(\beta\right)\right)}{v_i} = 0\,,$$

$$\forall j = 1, \ldots, p \qquad (**)$$

3. Generalized linear models

- $[Y_i]$ belongs to the exponential family with mean function $m_i(\beta) = \mathbb{E}(Y_i) = h\left(x_i'\beta\right)$ and
  variance function $v_i(\beta) = \phi \cdot \mathbb{V}(Y_i)$

- Now v_i depends on β, and (*) and (**) are no longer equivalent.

- Optimal estimation by generalizing (**):

$$\sum_{i=1}^{n} \frac{\partial\, m_i(\beta)}{\partial\, \beta_j} \cdot \frac{(Y_i - m_i(\beta))}{v_i(\beta)} = 0\,, \quad \forall j = 1, \ldots, p\,.$$

- Wedderburn (1974): *Quasi-likelihood estimation, (**)=quasi-score equation*

- Even if the distribution does not belong to the exponential family, the estimates gained form (**) are still reasonable.

$\boxed{\text{mean and variance function based on the observable quantity } W_i}$	data with measurement error

can be related to

$\boxed{\text{the mean and variance function based on the unobservable quantity } X_i}$	'ideal model' without measurement error

and therefore the data-based quasi-likelihood can be obtained.

$$\boxed{\begin{aligned} &\mathbb{E}\left[T_i^r | W_i, Z_i; \theta\right] = \\ &\quad = \mathbb{E}\left(\mathbb{E}\left[T_i | X_i, W_i, Z_i; \theta\right] | W_i, Z_i; \theta\right) = \\ &\quad = \mathbb{E}\Big(\underbrace{\mathbb{E}\left[T_i^r | X_i, Z_i; \theta\right]}_{\text{ideal model}} | \underbrace{W_i, Z_i}_{\text{observable}} ; \theta \Big) \end{aligned}}$$

$$\mathbb{V}\left[T_i|W_i, Z_i;\theta\right] =$$

$$\mathbb{V}\left(\mathbb{E}\left[T_i|X_i, W_i, Z_i;\theta\right]|W_i, Z_i;\theta\right) + \mathbb{E}\left(\mathbb{V}\left[T_i|X_i, W_i, Z_i;\theta\right]|W_i, Z_i;\theta\right) =$$

$$= \mathbb{V}\left(\underbrace{\mathbb{E}\left[T_i|X_i, Z_i;\theta\right]}_{\text{ideal model}}|\ \underbrace{W_i, Z_i}_{\text{observable}};\theta\right) + \mathbb{E}\left(\underbrace{\mathbb{V}\left[T_i|X_i, Z_i;\theta\right]}_{\text{ideal model}}|\ \underbrace{W_i, Z_i}_{\text{observable}};\theta\right)$$

or

$$\mathbb{V}\left[T_i|W_i, Z_i;\theta\right] = \mathbb{E}\left[T_i^2|W_i, Z_i;\theta\right] - \left(\mathbb{E}\left[T_i|W_i, Z_i;\theta\right]\right)^2 =$$

$$= \mathbb{E}\left(\underbrace{\mathbb{E}\left[T_i^2|X_i, Z_i;\theta\right]}_{\text{ideal model}}|\ \underbrace{W_i, Z_i}_{\text{observable}};\theta\right) - \left(\mathbb{E}\left(\underbrace{\mathbb{E}\left[T_i|X_i, Z_i;\theta\right]}_{\text{ideal model}}|\ \underbrace{W_i, Z_i}_{\text{observable}};\theta\right)\right)^2.$$

4.2 Correcting for Covariate Measurement Error in Accelerated Failure Time Models

$$\mathbb{E}\left[T_i^r | W_i, Z_i; \theta\right] = \mathbb{E}\Big(\underbrace{\mathbb{E}\left[T_i^r | X_i, Z_i; \theta\right]}_{\text{ideal model}} | \; \underbrace{W_i, Z_i}_{\text{observable}} \; ; \theta \Big)$$

Determine the 'ideal' moments!

Determine the conditional density of X_i given W_i and Z_i and estimate its parameters

Proposition: Let

- $X_i \sim \mathcal{MIXN}\Big(m;\ \kappa_1, \ldots, \kappa_m;\ \mu_1, \ldots, \mu_m;\ \sigma_1^2, \ldots, \sigma_m^2\Big)$
- $U_i \sim \mathcal{N}\left(0,\ \sigma_U^2\right)$
- $X_i \perp U_i$
- $W_i = X_i + U_i$

then

- $W_i \sim \mathcal{MIXN}\Big(m;\ \kappa_1, \ldots, \kappa_m;\ \mu_1, \ldots, \mu_m;\ \sigma_1^2 + \sigma_U^2, \ldots, \sigma_m^2 + \sigma_U^2\Big)$
- $X_i \mid W_i \sim \mathcal{MIXN}\Big(\bar{m};\ \bar{\kappa_1}, \ldots, \bar{\kappa_m};\ \bar{\mu}_1, \ldots, \bar{\mu}_m;\ \bar{\sigma}_1^2, \ldots, \bar{\sigma}_m^2\Big)$

with $(j = 1, \ldots, \bar{m})$

$$\bar{m} = m$$

$$\bar{\kappa}_j = \frac{\kappa_j \cdot \varphi(w_i \,\|\mu_j, \sigma_j^2 + \sigma_U^2)}{\sum_{l=1}^{\bar{m}} \kappa_l \cdot \varphi(w_i \,\|\mu_l, \sigma_l^2 + \sigma_U^2)}$$

$$\bar{\mu}_j = \mu_j + \frac{\sigma_j^2}{(\sigma_j^2 + \sigma_U^2)} \cdot (W_i - \mu_j)$$

$$\bar{\sigma}_j^2 = \sigma_j^2 \cdot \left(1 - \frac{\sigma_j^2}{\sigma_j^2 + \sigma_U^2}\right)$$

Basic ideas

- Use the accelerated failure time model as a superstructure!

$$\Downarrow$$

- General form of the 'ideal' mean and variance functions

Results on
X_i given W_i
$$\Downarrow$$

- General form of the 'data-based' mean and variance functions

$$\Downarrow$$

- General form for measurement error correction in the quasi-score function

$$\Downarrow$$

- Measurement error correction in a concrete model through appropriate specialization (i.e. choice of $c_r(\cdot)$ defined below)

The general form of the 'ideal' moments

$$\ln T_i = \beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i + \alpha \cdot \epsilon_i$$

$$\Downarrow$$

$$\mathbb{E}\left[T_i^r | X_i, Z_i; \theta\right] =$$

$$= \mathbb{E}\left[\exp\left(r \cdot \left(\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i\right)\right) \cdot \exp\left(r\alpha\epsilon_i\right)\right]$$

$$= \underbrace{c_r(\alpha)}_{distribution} \cdot \underbrace{\exp\left(r \cdot (\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i)\right)}_{covariates},$$

with $c_r(\alpha) = M_\epsilon(r \cdot \alpha)$ and $M_\epsilon(\cdot)$ denoting the moment generating function of ϵ_1.

Examples:

- Log-logistic model ($\alpha > r$): $c_r(\alpha) = \frac{r\pi}{\alpha \cdot \sin(r\pi/\alpha)}$
- Weibull model: $c_r = \Gamma\left(\frac{\alpha+r}{\alpha}\right)$
- Generalized Gamma model with parameter γ (Gamma model $\alpha = 1$):

$$c_r(\alpha) = \frac{1}{\gamma^\alpha} \cdot \frac{\Gamma(\gamma + \alpha \cdot r)}{\Gamma(\gamma)}$$

- Generalized F-model with parameters γ_1 and γ_2:

$$c_r(\alpha) = \frac{\Gamma(\gamma_1 + \alpha \cdot r) \cdot \Gamma(\gamma_1 - \alpha \cdot r) \cdot (\gamma_1/\gamma_2)^{r \cdot \alpha}}{(\Gamma(\gamma_1) \cdot \Gamma(\gamma_2))}$$

The general form of the 'data-based' moments

$c_r(\alpha)$ does not depend on X_i. Therefore

$$
\begin{aligned}
&\mathbb{E}\left[T_i^r|W_i, Z_i;\theta\right] = \mathbb{E}\Big(\mathbb{E}\left[T_i^r|X_i, Z_i;\theta\right]|W_i, Z_i;\theta\Big) = \\
&= \mathbb{E}\Big(c_r(\alpha)\cdot \exp\left(r\cdot(\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i)\right)|W_i, Z_i;\theta\Big) = \\
&= c_r(\alpha)\cdot \sum_{j=1}^{\bar{m}} \bar{\kappa}_j \cdot \exp\Big(r\cdot\Big(\beta_0 + \beta_X\cdot\bar{\mu}_{X_i,j} + \\
&\qquad\qquad + \beta_X^2\cdot 0.5\cdot\bar{\sigma}^2_{X_i,j} + \beta_Z^T\cdot Z_i\Big)\Big)
\end{aligned}
$$

with

$$
\bar{\kappa}_j = \frac{\kappa_j\cdot\varphi\left(w_i\,\|\mu_j, \sigma_j^2 + \sigma_U^2\right)}{\sum_{l=1}^{\bar{m}}\kappa_l\cdot\varphi\left(w_i\,\|\mu_l, \sigma_l^2 + \sigma_U^2\right)}
$$

$$
\bar{\mu}_{X_i,j} = \mu_j + \frac{\sigma_j^2}{(\sigma_j^2 + \sigma_U^2)}\cdot(W_i - \mu_j)
$$

$$
\bar{\sigma}^2_{X_i,j} = \sigma_j^2\cdot\left(1 - \frac{\sigma_j^2}{\sigma_j^2 + \sigma_U^2}\right).
$$

Examples:

- Log-logistic model ($\alpha > r$): $c_r(\alpha) = \frac{r\pi}{\alpha \cdot \sin(r\pi/\alpha)}$

- Weibull model: $c_r = \Gamma\left(\frac{\alpha+r}{\alpha}\right)$

- Generalized Gamma model with parameter γ (Gamma model $\alpha = 1$):

$$c_r(\alpha) = \frac{1}{\gamma^\alpha} \cdot \frac{\Gamma(\gamma + \alpha \cdot r)}{\Gamma(\gamma)}$$

- Generalized F-model with parameters γ_1 and γ_2:

$$c_r(\alpha) = \frac{\Gamma(\gamma_1 + \alpha \cdot r) \cdot \Gamma(\gamma_1 - \alpha \cdot r) \cdot (\gamma_1/\gamma_2)^{r \cdot \alpha}}{(\Gamma(\gamma_1) \cdot \Gamma(\gamma_2))}$$

Inserting this into the quasi-score equation yields measurement error corrected quasi-likelihood estimates, which can be shown to have the following

asymptotic properties

- consistency $\Rightarrow$ The measurement error bias is eliminated.
- asymptotic normality
- efficiency among all linear estimating equations

5. Mismeasured Duration Times

- Measurement error in T_i: The time of staying in the state of interest can not be determined exactly (retrospective studies, cf. Skinner & Humphreys, 1999)

- The method developed here is (the first one, which is (?)) able to deal simultaneously with measurement error in covariates as well as in the duration time.

5.1 Additive measurement error

- Instead of T_i the variable $T_i^* = T_i + V_i$ is observed.
  $V_i \perp$ rest, $\mathbb{E}\,(V_i) = m$, $\mathbb{V}\,(V_i) = v$

- Mean and variance functions can be appropriately modified.

- The estimation procedure and the asymptotic results can be applied in a similar way.

5.2 Multiplicative measurement error

- $(T_i^* - T_i)$ proportional to T_i: Instead of T_i one observes T_i^* with $T_i^* = V_i \cdot T_i$.
  $V_i \perp$ rest, $\mathbb{E}(V_i) = m$, $\mathbb{V}(V_i) = v$

- $$\ln T_i^* = \ln T_i + \ln V_i = \beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i + \alpha \cdot \epsilon_i + \ln V_i$$

- This is no accelerated failure time model, but nevertheless

  $$\mathbb{E}\left[(T_i^*)^r | X_i, Z_i; \theta\right] = \bar{c}_r(\alpha) \cdot \exp\left(r \cdot (\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i)\right)$$

  with $\bar{c}_r(\alpha) = M_\epsilon(r \cdot \alpha) \cdot \mathbb{E}_{m,v}\left[V_i^r\right]$.

- Since $\bar{c}_r(\alpha)$ does not depend on X_i, the methods described above can be applied again in a similar way. The asymptotic optimality-properties can be shown to remain valid, too.

6. Censored Duration Times

The main problem

- Instead of T_i one observes

  $$Y_i := \min(T_i, C_i) \quad \text{and} \quad \Delta_i := I(\{T_i < C_i\})$$

  with C_i as the censoring variable

- Idea: Generalize the concept from above to the multivariate observations (Y_i, Δ_i)

- But, even under random censorship, determining the 'ideal' mean and variance functions proves to be difficult:

- For the moment: assume no measurement error

- Standard assumptions on the censoring mechanism:

  * random censorship (T_i and C_i are stochastically independent)

  * non-informative censoring mechanism (independency of C_i the covariates)

- To construct $f_{Y_i,\Delta_i|X_i,Z_i}(t,\delta)$ note:

  * If $\delta = 1$, observation t_i corresponds to the event

$$\{T_i = t_i\} \cap \{C_i > t_i\}$$

  * If $\delta = 0$, observation t_i corresponds to the event

$$\{C_i = t_i\} \cap \{T_i \geq t_i\}$$

- $$f_{Y_i,\Delta_i|X_i,Z_i;\theta}(t_i,\delta_i) = \\ = \left(f_{T_i|X_i,Z_i;\theta}(t_i) \cdot S_{C_i}(t_i)\right)^{\delta_i} \cdot \\ \cdot \left(f_{C_i}(t_i) \cdot S_{T_i|X_i,Z_i;\theta}(t_i)\right)^{1-\delta_i}.$$

$$f_{Y_i,\Delta_i|X_i,Z_i;\theta}\left(t_i,\delta_i\right) =$$
$$= \left(f_{T_i|X_i,Z_i;\theta}\left(t_i\right)\cdot S_{C_i}\left(t_i\right)\right)^{\delta_i} \cdot$$
$$\cdot \left(f_{C_i}\left(t_i\right)\cdot S_{T_i|X_i,Z_i;\theta}\left(t_i\right)\right)^{1-\delta_i}.$$

- $\text{Lik}\left(\theta|t_i,\delta_i;X_i,Z_i\right) \propto \left(f_{T_i}\left(t_i|X_i,Z_i;\theta\right)\right)^{\delta_i} \cdot \left(S_{T_i}\left(t_i|X_i,Z_i;\theta\right)\right)^{1-\delta_i}$

- For the marginal density $f_{Y_i}\left(t_i\right)$ of Y_i:

$$\begin{aligned} f_{Y_i}\left(t_i\right) &= f_{Y_i,\Delta_i|X_i,Z_i;\theta}\left(t_i,1\right) + \\ &\quad + f_{Y_i,\Delta_i|X_i,Z_i;\theta}\left(t_i,0\right) \\ &= f_{T_i|X_i,Z_i}\left(t_i\right)\cdot S_{C_i}\left(t_i\right) + \\ &\quad + f_{C_i}\left(t_i\right)\cdot S_{T_i|X_i,Z_i}\left(t_i\right). \end{aligned}$$

leading to

$$\begin{aligned} \mathbb{E}\left[Y_i^r|X_i,Z_i;\theta\right] &= \int_0^\infty t^r\cdot f_{Y_i}\left(t|X_i,Z_i;\theta\right)\,dt = \\ &= \int_0^\infty t^r\cdot f_{T_i}\left(t|X_i,Z_i;\theta\right)\cdot S_{C_i}(t)\,dt + \\ &\quad + \int_0^\infty t^r\cdot f_{C_i}(t)\cdot S_{T_i}\left(t|X_i,Z_i;\theta\right)\,dt \end{aligned}$$

which depends on the unknown distribution of C_i.

In contrast to the likelihood approach, where the censoring-distribution $f_{C_i}(\cdot)$ enters as a proportionality constant not influencing the inference on the parameters, **mean and variance functions are interrelated with the unknown censoring-distribution in quite a complex manner.**

A first attempt: single censoring

Maximal time η under study fixed, no censoring before η. Then

$$f_{C_i}(t) = \begin{cases} 1 & t = \eta \\ 0 & t \neq \eta \end{cases} \quad S_{C_i}(t) = \begin{cases} 1 & t \leq \eta \\ 0 & t > \eta \,. \end{cases}$$

For the 'ideal' moments:

$$\mathbb{E}\left[Y_i^r | X_i, Z_i; \theta\right] = \\ \int_0^\eta t^r f_{T_i}(t|X_i, Z_i;\vartheta) dt + \eta^r \cdot S_{T_i}(\eta|X_i, Z_i;\vartheta)\,.$$

First term: partial moment $\mathbb{E}^\eta\left[T_i^r|X_i, Z_i;\theta\right]$ of T_i given X_i and Z_i with upper limit η.

$$\mathbb{E}^\eta\left[T_i^r|X_i, Z_i;\right] = \\ = \tilde{c}_r \cdot \exp\left(r \cdot (\beta_0 + \beta_X \cdot X_i + \beta_Z^T \cdot Z_i)\right)$$

with

$$\tilde{c}_r = \mathbb{E}^{\tilde{\eta}}\left[\exp\left(r\alpha\epsilon_i\right)\right]\,.$$

Note that the general form from above is not established here. $\tilde{c}_r$ depends via $\tilde{\eta}$ in a complicated way on the covariates. No general formula for measurement error correction seems to be available.

Further ideas and questions

? Assuming a parametric distribution for C_i
How can one select a (least committal) parametric distribution appropriately?

? Semiparametric procedure: Piecewise exponential model for the distribution of C_i
How does the quasi-likelihood approach work under infinite many nuisance parameters?

? A parametric variant of the Buckley-James approach: Incorporate censoring by considering pseudo variables Y^* with

$$Y_i^* := \begin{cases} T_i & i \text{ uncensored} \\ \mathbb{E}\left[T_i | \{T_i > c_i\}, X_i, Z_i; \theta\right] & i \text{ censored at time } c_i \end{cases}$$

Where has this idea been applied in a parametric setting? What is the experience with it?

Exponential censoring

- Justification: constant risk, 'non-informative prior', $(\nu > 0.)$

- $S_C(t) = \exp(-\nu \cdot t)$, $f_C(t) = \nu \cdot exp(-\nu \cdot t)$.

- $$\begin{aligned} f_{Y_i|X_i,Z_i}(t) \;\; = \;\; & f_{T_i|X_i,Z_i}(t) \cdot \exp(-\nu \cdot t) \\ & + S_{T_i|X_i,Z_i}(t) \cdot \nu \cdot \exp(-\nu \cdot t) \end{aligned}$$

- Assume the moment generating function $M_{T_i|X_i,Z_i}(a)$ of T to exist (and to be analytic) for $a \in (-\infty, c_0)$ with $c_0 > 0$. Then the moment generating function $M_{Y_i|X_i,Z_i}(a)$ of Y exists, and

$$\begin{aligned} & M_{Y_i|X_i,Z_i}(a) = \\ = \;\; & \int_0^\infty \exp(a \cdot t) \cdot f_{Y_i|X_i,Z_i}(t)\; dt = \\ = \;\; & \int_0^\infty \exp(a \cdot t) \cdot f_{T_i|X_i,Z_i}(t) \exp(-\nu \cdot t)\; dt \\ & + \int_0^\infty \exp(a \cdot t) \cdot S_{T_i|X_i,Z_i}(t) \cdot \nu \cdot \exp(-\nu \cdot t)\; dt = \\ = \;\; & \int_0^\infty \exp((a - \nu) \cdot t) \cdot f_{T_i|X_i,Z_i}(t)\; dt \\ & + \int_0^\infty \exp((a - \nu) \cdot t) \cdot S_{T_i|X_i,Z_i}(t)\; dt = \ldots = \\ = \;\; & M_{T_i|X_i,Z_i}(a - \nu) + \nu \left(\frac{M_{T_i|X_i,Z_i}(a - \nu)}{a - \nu} - 1 \right) \end{aligned}$$

- Try to correct for the measurement error by calculating

$$M_{Y_i|W_i,Z_i}(a) = \mathbb{E}\left(M_{Y_i|X_i,Z_i}(a)\,|W_i\right)$$

- If this is not feasible,

  * calculate the moments given X_i, Z_i and correct them or

  * relate them by Taylor expansion to the moments of T_i given X_i, Z_i, which can be corrected as described above.

A Semi/Nonparametric Approach

- $f_{C_i}(\cdot)$ piecewise exponential

- Time points

$$0 \overset{def}{=} \tau_0 < \tau_1 < \ldots < \tau_q < \tau_{q+1} \overset{def}{=} \infty$$

- In every interval: constant hazard rate λ_j

- $$\int_0^\infty f_{T_i|X_i,Z_i}(t) \cdot S_{C_i}(t)\, dt =$$
$$= \sum_{j=1}^{q+1} \int_{\tau_{j-1}}^{\tau_j} t^r \cdot f_{T_i|X_i,Z_i}(t) \cdot \exp\left(-\lambda_j \cdot t\right) dt$$

- again bounded range of integration, difficult to correct for measurement error.

- If the τ_j's are chosen data-dependent, the number of nuisance parameter $\lambda_1, \ldots, \lambda_q$ goes to infinity.

Additional Note

(Wallner and T. A.)

- Again and again terms of the form

$$\mathbb{E}\left(S_{T_i|X_i,Z_i}\left(g\left(X_i, Z_i\right)\right| W_i\right)$$

are needed where

$$[X_i|W_i] = \mathcal{N}\left(\tilde{\tilde{\mu}}\,,\,\tilde{\tilde{\sigma}}^2\right)$$

- For instance in the Weibull case

$$\mathbb{E}\left(\exp\left(-a \cdot \exp\left(\tilde{X}_i\right)\right)|W_i\right)\,, \quad a > 0\,,$$

which is equivalent to

$$\mathbb{E}\left(\exp\left(-a \cdot \tilde{Y}_i\right)\right)\,, \quad [Y_i|W_i] = \log\mathcal{N}\left(\tilde{\tilde{\mu}}\,,\,\tilde{\tilde{\sigma}}^2\right)\,.$$

Because of $0 \leq \exp\left(-a \cdot \tilde{Y}_i\right) \leq 1$ one knows

$$0 \leq \mathbb{E}\left(\exp\left(-a \cdot \tilde{Y}_i\right)\right) \leq 1\,.$$

But how to calculate the concrete value?

- $$\mathbb{E}\left(\exp\left(-a \cdot \tilde{Y}_i\right)\right) = ??$$

- Moment generating function of lognormally distributed variables

  Does not exist, is not analytic.

- Δ-Method (Taylor expansion around $\tilde{\tilde{\mu}}$)

  diverges

- Expansion around zero

  diverges

- Idea: Construct an uniformly integrable sequence $(g_n)_{n\in\mathbb{N}}$ by considering "censored (truncated) expansions", i.e. the sequence

$$\left(g_n\left(y\right) \stackrel{def}{=} \sum_{j=1}^{n} (-1)^j \frac{(a\cdot y)^j}{j!} \cdot I_{[0,b_n]}(y) \right)_{n\in\mathbb{N}} .$$

For appropriate choice of the real sequence $(b_n)_{n\in\mathbb{N}}$

$$g_n\left(y\right) \stackrel{n\to\infty}{\to} \exp\left(-a\cdot y\right)$$

and $\{g_n\,, n\in\mathbb{N}\}$ is uniformly integrable.

Therefore

$$\mathbb{E}\left(\exp\left(-a\cdot\tilde{Y}_i\right)\right) = \lim_{n\to\infty}\mathbb{E}\left(g_n\left(Y\right)\right) .$$

7. Concluding Remarks and Some Topics for Further Research

- Easy to handle tool to correct for measurement error in covariates as well as in the duration time.

  > "[. . . via quasi-likelihood] useful inferences are possible even in problems for which a full likelihood-based analysis is either intractable or impossible with the given assumptions."
  >
  > (McCallagh, 1991, p. 265.)

- Accelerated failure time serves as a superstructure to derive general results

- Appropriate modelling of censorship?
- Simulation and application to real data
- Compare the approach with functional approaches and with Bayesian techniques!
- Improve quasi-likelihood estimation by taking into account higher order corrected moments!
- Multivariate extension possible: error in several covariates
- Extension to mixtures of skew distributions for the latent variable ?
- Do mixtures of normals also provide useful models for the error distribution (heavy tails, outliers) ?
- Robustification of quasi-likelihood estimation
- Proper techniques to estimate the number of components of the mixture distribution are needed.