Imprecise Measurement Error Models and Partial Identification - Towards a Unified Approach for Non-Idealized Data

Thomas Augustin

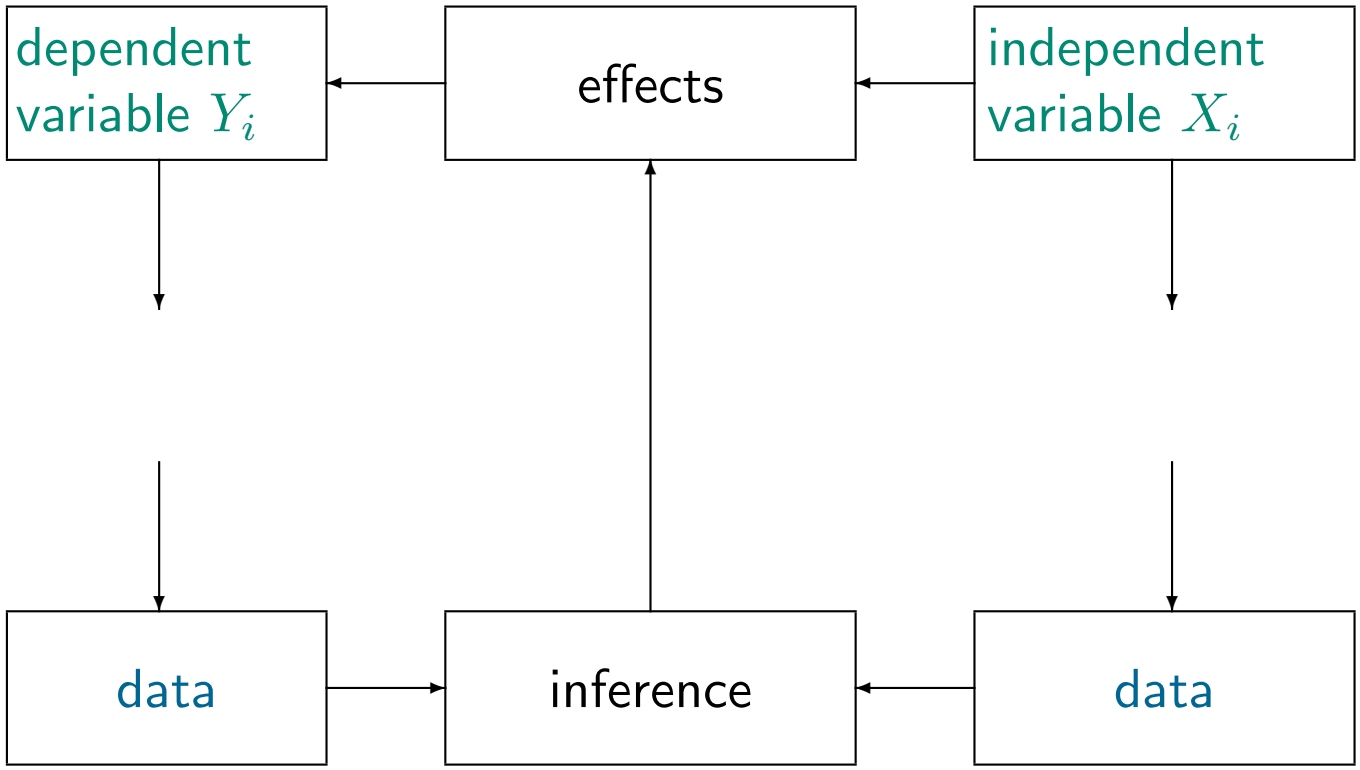
Department of Statistics
Ludwig-Maximilians University Munich

Table of Contents

1. Background and Sketch of the Arguments

1. Background and Sketch of the Arguments

Applied Statistics: Learning from data by sophisticated models
Complex relationships between variables



Often the relationship between variables and data is complex, too:

* Often variables of interest (gold standard) are not ascertainable.

* Only proxy variables (surrogates) are available instead.

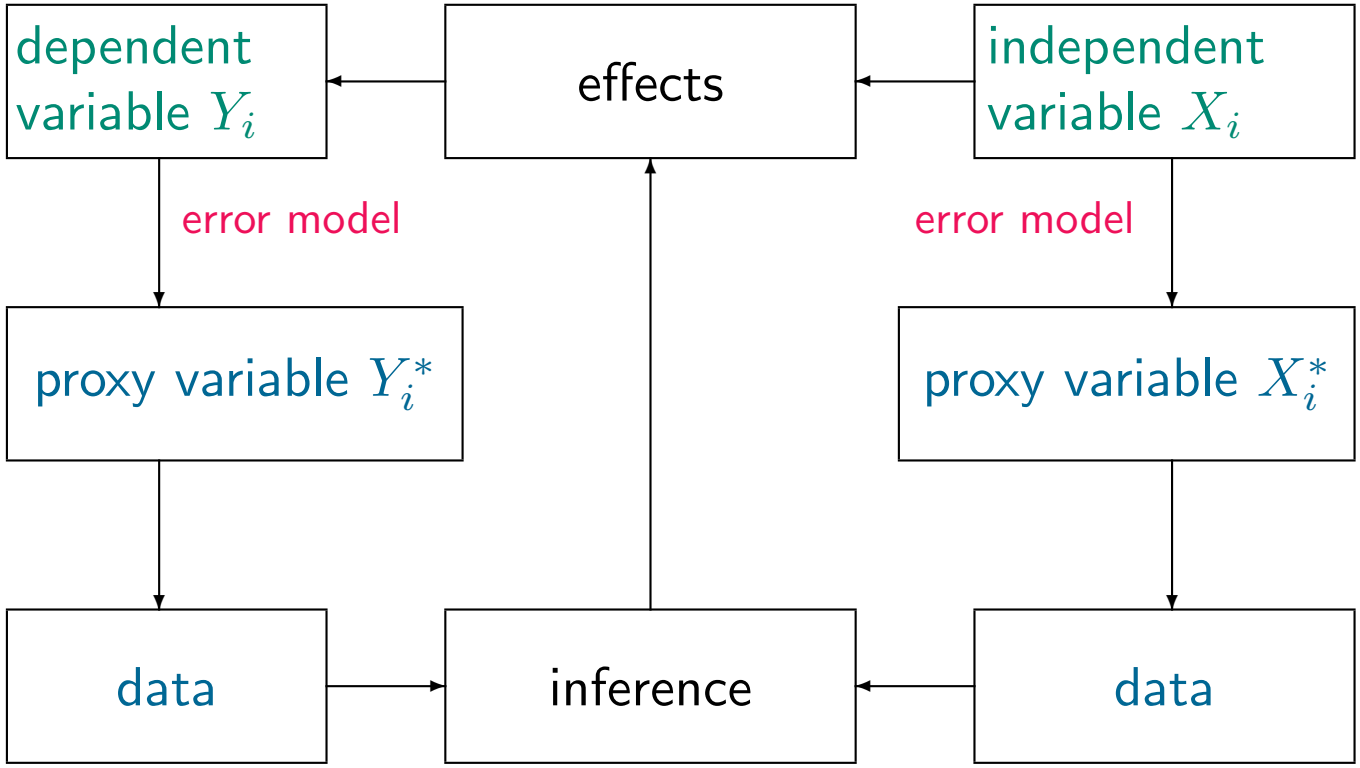
Typical examples: Measurement Error

- Error-prone measurements of true quantities

  * error in technical devices
  * indirect measurement
  * response effects
  * use of aggregated quantities, averaged values, imputation, rough estimates etc.
  * anonymization of data by deliberate contamination

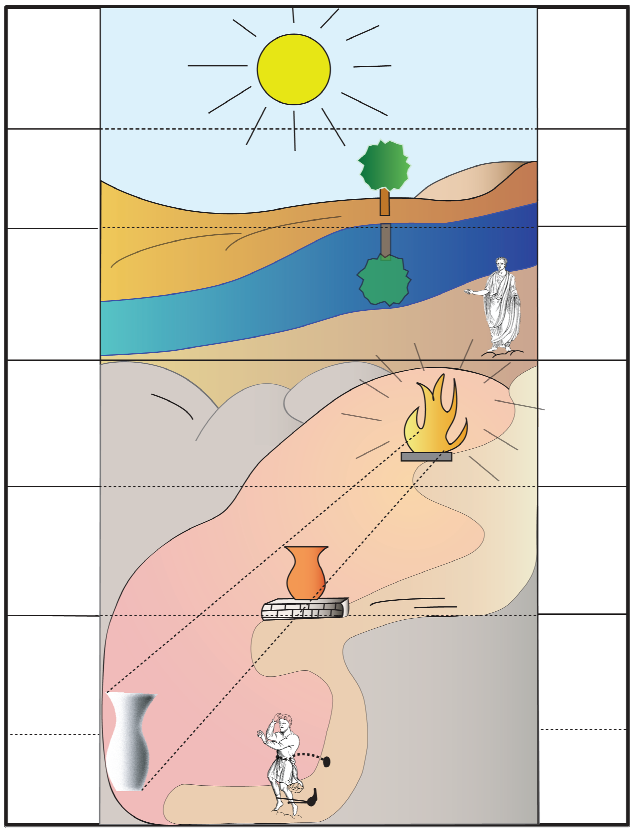
- Operationalization of complex constructs; latent variables

  * long term quantities: permanent income,
  * importance of a patent
  * extent of motivation, degree of costumer satisfaction
  * severeness of undernutrition


dependent variable Y_i
effects
independent variable X_i
error model
error model
proxy variable Y_i^*
proxy variable X_i^*
data
inference
data

Cave Allegory

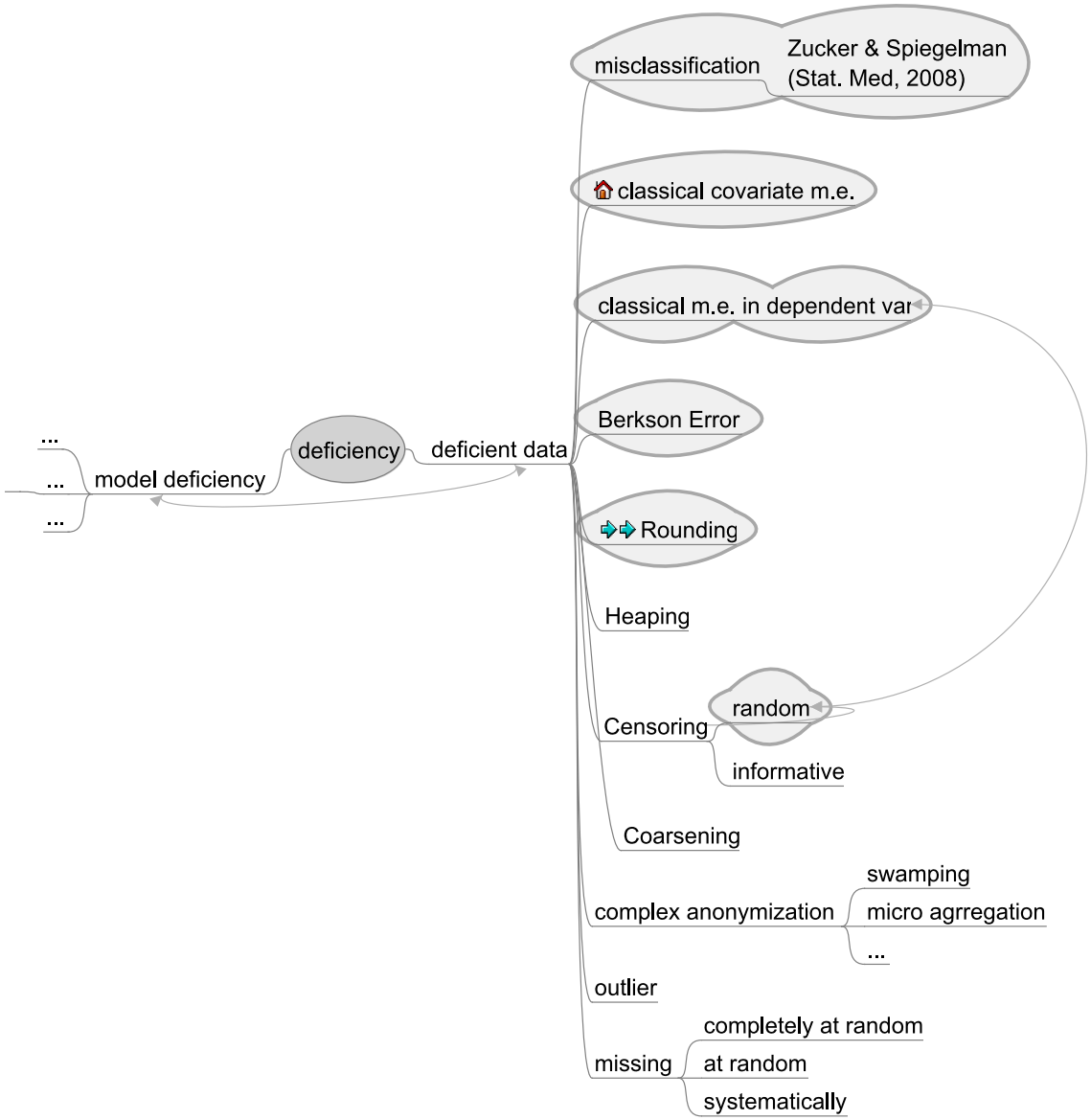
http://commons.wikimedia.org/wiki/File:Allegory_of_the_Cave_blank.png



Notation

We have to distinguish between true (correctly measured) variable and its (possible incorrect) measurement, i.e. between the gold standard and the corresponding surrogate.

* - Notation (here)
X, Z : (unobservable) variable, gold standard
X^*, Z^*: corresponding possibly incorrect measurements analogously: Y, Y^* and T, T^*


misclassification
Zucker & Spiegelman (Stat. Med, 2008)
classical covariate m.e.
classical m.e. in dependent var
Berkson Error
Rounding
Heaping
random
Censoring
informative
Coarsening
swamping
complex anonymization
micro agrregation
...
outlier
completely at random
missing
at random
systematically
deficiency
deficient data
...
...
model deficiency
...

Partial Identification and Beyond

Manski's Law of Decreasing Credibility

Reliability !? Credibility ?

> "The credibility of inference decreases with the strength of the assumptions maintained." (Manski (2003, p. 1))

Identifying Assumptions Very strong assumptions needed to ensure identifiability = precise solution

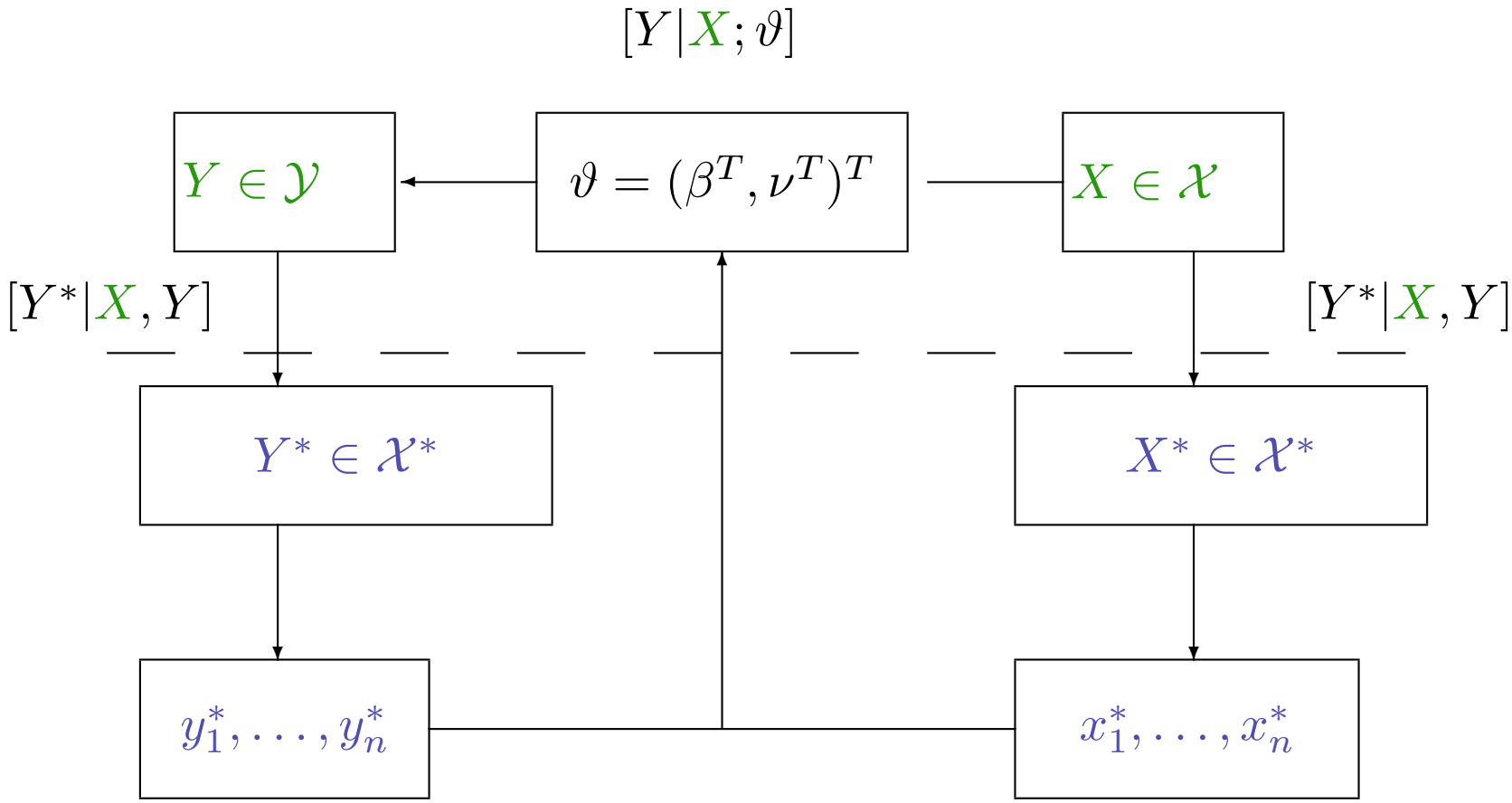
- Measurement error model completely known
  - type of error, in particular assumptions on (conditional) independence
  - type of error distribution
  - moments of error distribution
- validation studies often not available

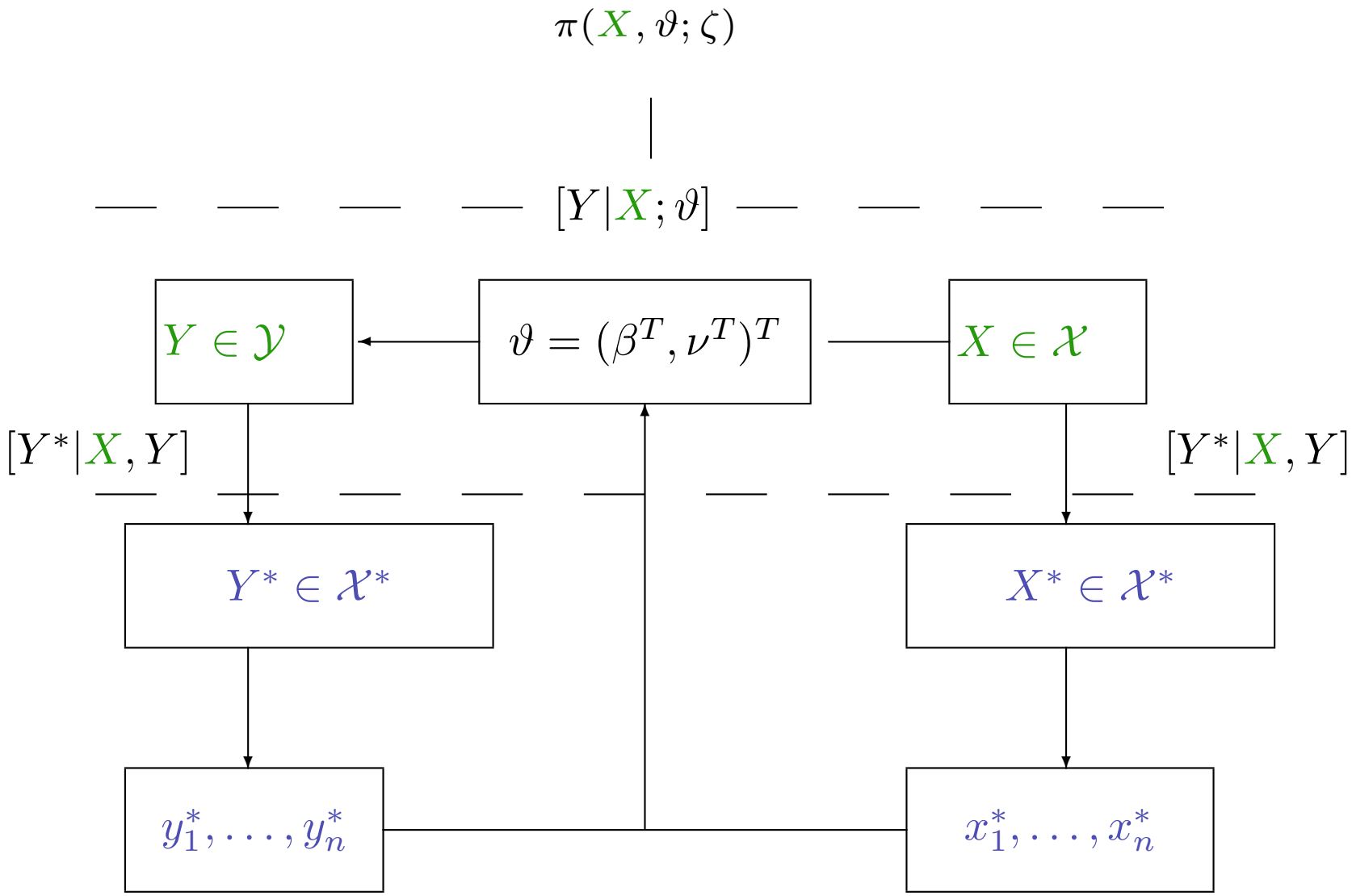
Reliable Inference Instead of Overprecision!!

- Make more „realistic“ assumptions and let the data speak for themselves!
- Consider the *set* of *all* models that are compatible with the data (and then add successively additional assumptions, if desirable)
- The results may be imprecise, but are more reliable for sure
- **The extent of imprecision is related to the data quality!**
- As a welcome by-product: clarification of the implication of certain assumptions
- parallel development (missing data; transfer to measurement error context!)
  * econometrics: *partial identification:* e.g., Manski (2003, Springer)
  * biometrics: *systematic sensitivity analysis:* e.g., Vansteelandt, Goetghebeur, Kenword, Molenberghs (2006, Stat. Sinica)
- current developments, e.g.,
  * Cheng, Small (2006, JRSSB)
  * Henmi, Copas, Eguchi (2007, Biometrics)
  * Stoye (2009, Econometrica)
- Kleyer (2009, MSc., LMU); Kunz, Augustin, Küchenhoff (2010, TR)

Credal Estimation

- Natural idea: sets of traditional models $\longrightarrow$ sets of traditional estimators
- Construct estimators $\widehat{\Theta} \subseteq \mathbb{R}^p$, set-valued, typically interval-valued, point estimators appropriately reflecting the ambiguity (non-stochastic uncertainty, ignorance) in the credal set $\mathcal{P}$.
- $\widehat{\Theta}$ small if, and only if (!), $\mathcal{P}$ "small"
  * Usual point estimator as the border case of precise probabilistic information
  * Connection to Manski's (2003) *identification regions* and Vansteelandt, Goetghebeur, Kenward & Molenberghs (Stat Sinica, 2006) *ignorance regions*.
- Construction of unbiased sets of estimating functions
- Credal consistency


$[Y|X;\vartheta]$
$Y \in \mathcal{Y}$
$\vartheta = (\beta^T, \nu^T)^T$
$X \in \mathcal{X}$
$[Y^*|X,Y]$
$[Y^*|X,Y]$
$Y^* \in \mathcal{X}^*$
$X^* \in \mathcal{X}^*$
$y_1^*, \ldots, y_n^*$
$x_1^*, \ldots, x_n^*$


$\pi(X, \vartheta; \zeta)$
$[Y|X; \vartheta]$
$Y \in \mathcal{Y}$
$\vartheta = (\beta^T, \nu^T)^T$
$X \in \mathcal{X}$
$[Y^*|X, Y]$
$[Y^*|X, Y]$
$Y^* \in \mathcal{X}^*$
$X^* \in \mathcal{X}^*$
$y_1^*, \ldots, y_n^*$
$x_1^*, \ldots, x_n^*$

Credal Deficiency Models

Different types of deficiency can be expressed

- Measurement error problems
- Misclassification
- If $\mathcal{Y}^* \subsetneqq \mathcal{P}(\mathcal{Y}) \times \{0,1\}$: coarsening, rounding, censoring, missing data
- Outliers

Credal set: convex set of traditional probability distributions

$$\begin{aligned} [Y|X,\vartheta] &\in \mathcal{P}_{Y|X,\vartheta} \\ [Y^*|X,Y] &\in \mathcal{P}_{Y^*|X,Y} \\ [X^*|X,Y] &\in \mathcal{P}_{X^*|X,Y} \end{aligned}$$

$$\mathcal{P} := \text{conv}\left(\mathcal{P}_{Y|X,\vartheta} \otimes \mathcal{P}_{Y^*|X,Y} \otimes \mathcal{P}_{X^*|X,Y}\right)$$

2. Measurement Error Correction based on Precise Error Models

2.1. Measurement Error Modelling

“Measure what is measurable,
and make measurable what is not so.“
(Galilei Galileo)

Typical examples

- Error-prone measurements of true quantities
  * error in technical devices
  * indirect measurement
  * response effects
  * use of aggregated quantities, averaged values, imputation, rough estimates etc.
  * anonymization of data by deliberate contamination

- Operationalization of complex constructs; latent variables
  * long term quantities: permanent income,
  * importance of a patent
  * extent of motivation, degree of costumer satisfaction
  * severeness of undernutrition

Notation again

We have to distinguish between true (correctly measured) variable and its (possible incorrect) measurement, i.e. between the gold standard and the corresponding surrogate.

* - Notation (here)
X, Z : (unobservable) variable, gold standard
X^*, Z^*: corresponding possibly incorrect measurements analogously: Y, Y^* and T, T^*

Sources of measurement error

- Induced by an instrument (laboratory value, blood pressure)
- Induced by the study particants (medical doctors or patients; interviewers and respondents)
- Surrogate variables, e.g. "Job exposure matrix: typical, instead of individual, exposure", "economic wealth of a district instead of individual income"
- Measurement error induced by definition, e.g. "long term mean of daily fat intake", "average income"
- Operationalization of complex constructs ("quality of life", "consumer satisfaction")

Note:

- 'Measurement error' and 'misclassification' are not just a matter of sloppiness.

- Latent variables are eo ipso not exactly measurable.

- "Almost all economic variables are measured with error. [...] Unfortunately, the statistical consequences of errors in explanatory variables are severe."
  (Davidson and Mackinnon (1993),
  Estimation and Inference in Econometrics.)

- In nonlinear models, the later statement may apply (!?) to the dependent variable, too. (Dependence on the DGP: Torelli & Trivellato (1993, J. Econometrics))

The triple whammy effect of measurement error

Carroll, Ruppert, Stefanski, Crainiceanu (2006, Chap.H.)

– bias
– masking of features
– loss of power

- **classical error**: "attenuation "

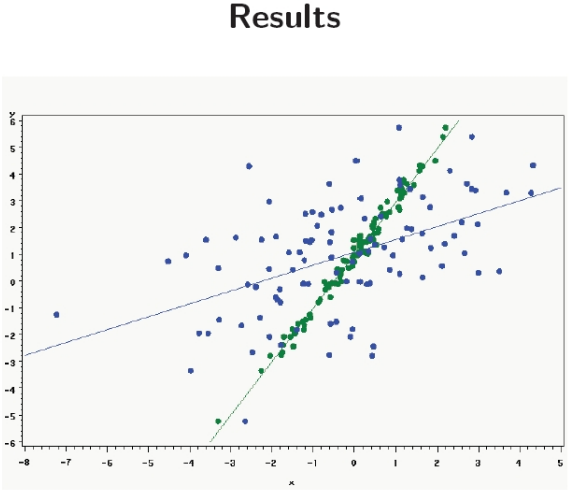
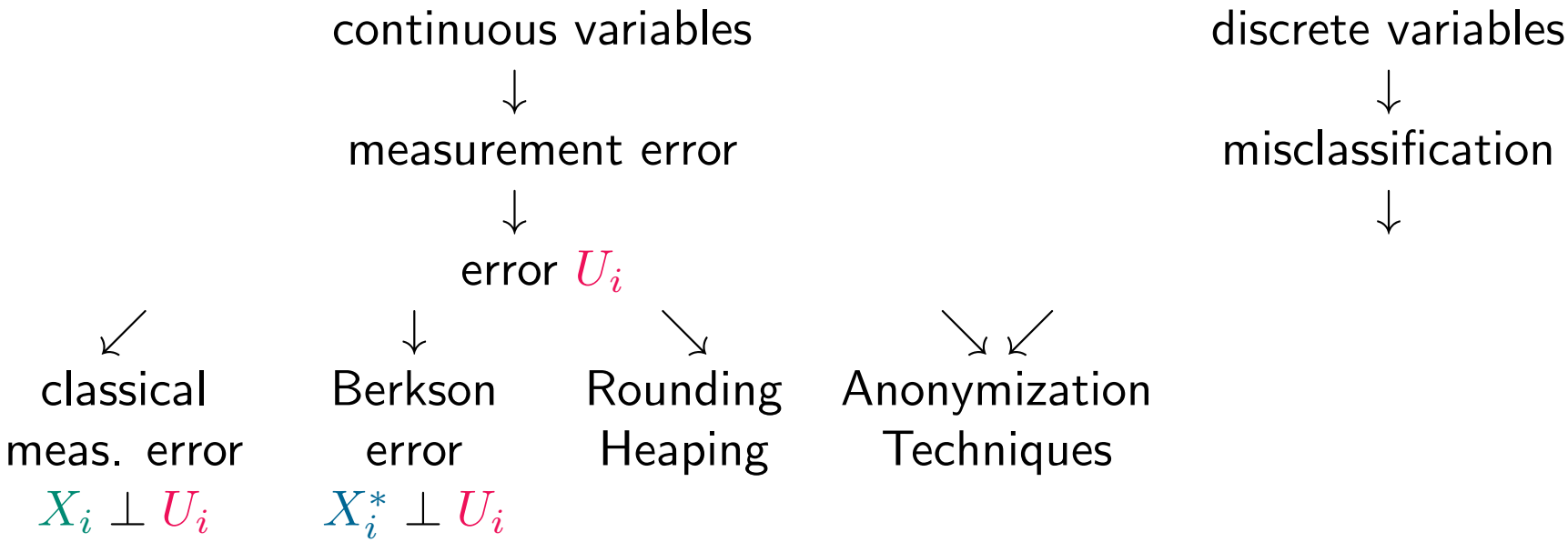


Figure 1: Effect of additive measurement error on linear regression

Terminology



Typical Examples: Berkson Error

$$X_i^* = X_i + U_i \text{ but } X_i^* \perp U_i$$

- Experimental design: X^* target value, X truly absorbed value
- Aggregated data:
  * aggregation for confidentiality, e.g., average income on "block level"
  * X^* mean exposure, X true individual exposure, JEM
- Omitted variable bias
- Remainder term of regression calibration

Aggregated data

- $X_1, \ldots, X_n$ i.i.d. $\sim \left(\mu, \sigma^2\right)$
- $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \operatorname*{plim}_{n \to \infty} \overline{X} = \mu$
- $X_i \approx \left(\overline{X}, \sigma^2\right)$

  $\Longleftrightarrow X_i = \overline{X} + \delta \quad$ with $\delta \sim \left(0; \sigma^2 \left(1 - \frac{1}{n}\right)\right) \quad \delta \perp \overline{X}$

References are missing!!

Omitted variable ???

- Linear predictor with $p = 2$

$$\eta_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i, \quad Z_i \perp X_i$$

- Omitting Z_i means to work with

$$\eta^* = \beta_0^* + \beta_1^* X_i^*$$

where

$$X_i^* = X_i + \underbrace{\frac{\beta_2}{\beta_1}}_{\delta_i} Z_i.$$

- WHAT HAPPENS FOR $p > 2$, HOW IS Z_i SHARED BETWEEN X_{i1}, X_{i2}?

Fehlklassifikation

Y^* beobachtete, Y wahre Klassenzugehörigkeit

- $Y^* = 1$ Test positiv, $Y = 1$ krank,
- Betrugsverdachtsfall, zu Schulungsmaßnahme angemeldet, geschätzte Entwicklung des Auftragseingang

	$Y = 1$	$Y = 0$
$Y^* = 1$	$\mathbb{P}(Y^* = 1 \mid Y = 1)$ sensitivity ☺	$\mathbb{P}(Y^* = 1 \mid Y = 0)$ false positive ☹
$Y^* = 0$	$\mathbb{P}(Y^* = 0 \mid Y = 1)$ false negative ☹	$\mathbb{P}(Y^* = 0 \mid Y = 0)$ specificity ☺

Fehlklassifikationsbias

$$p^* = p \cdot sens + (1 - p) \cdot (1 - spec) = p\,(sens + spec - 1) + (1 - spec)$$

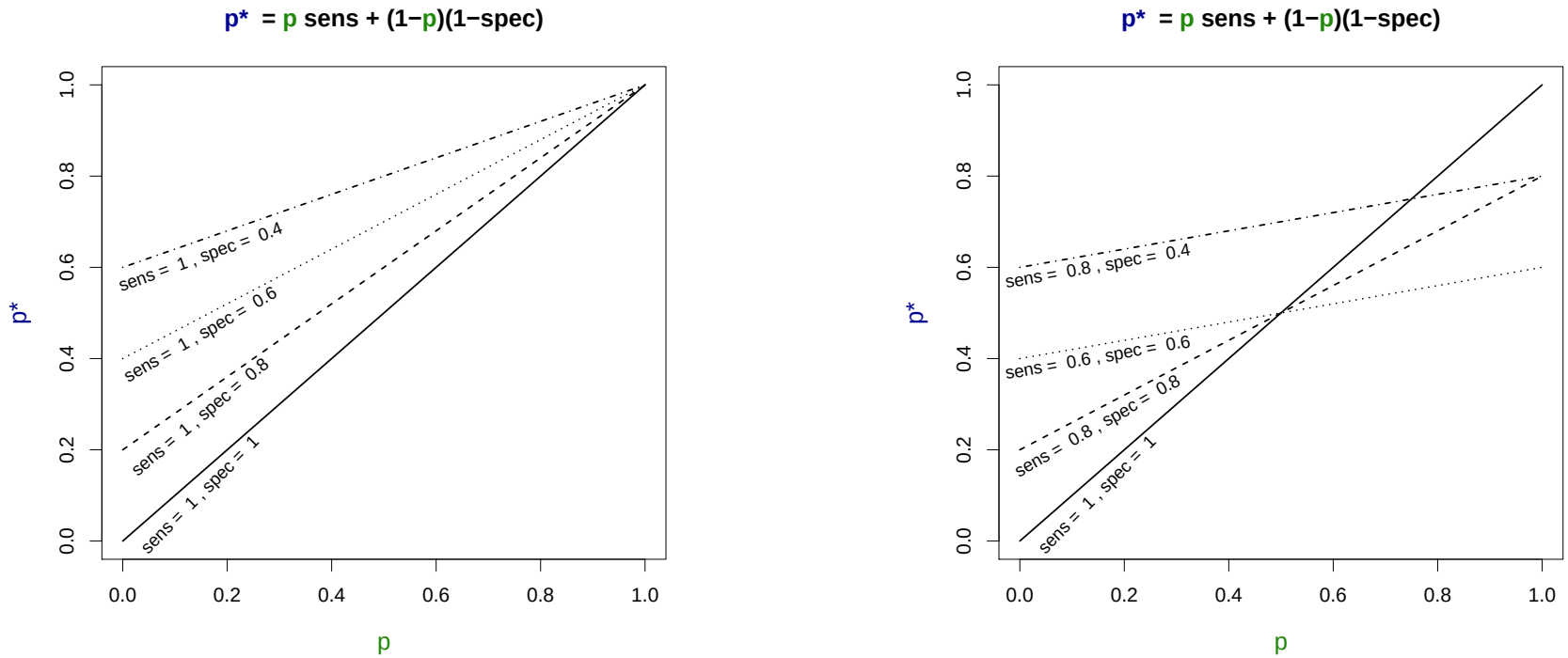


Figure 1: Veranschaulichung Fehlklassifikationsbias: Abweichung von der Winkelhalbierenden

Typical Examples: Rounding and Heaping

- duration data are commonly collected in a retrospective way
- strong memory effects when time spans have to be remembered, e.g., Skinner & Humphreys (1999, Lifetime Data Analysis)
- Holt et al (1991, Biemer et al(eds.)): age at menarche
- heaping in episode / spell-based designs: Torelli & Trivellato (1993, J. Econometrics):
  concentration of values of unemployment duration at multiples of six ("identification problem": heaping versus effect of different levels of compensation)
  strong dependence of the bias on the DGP

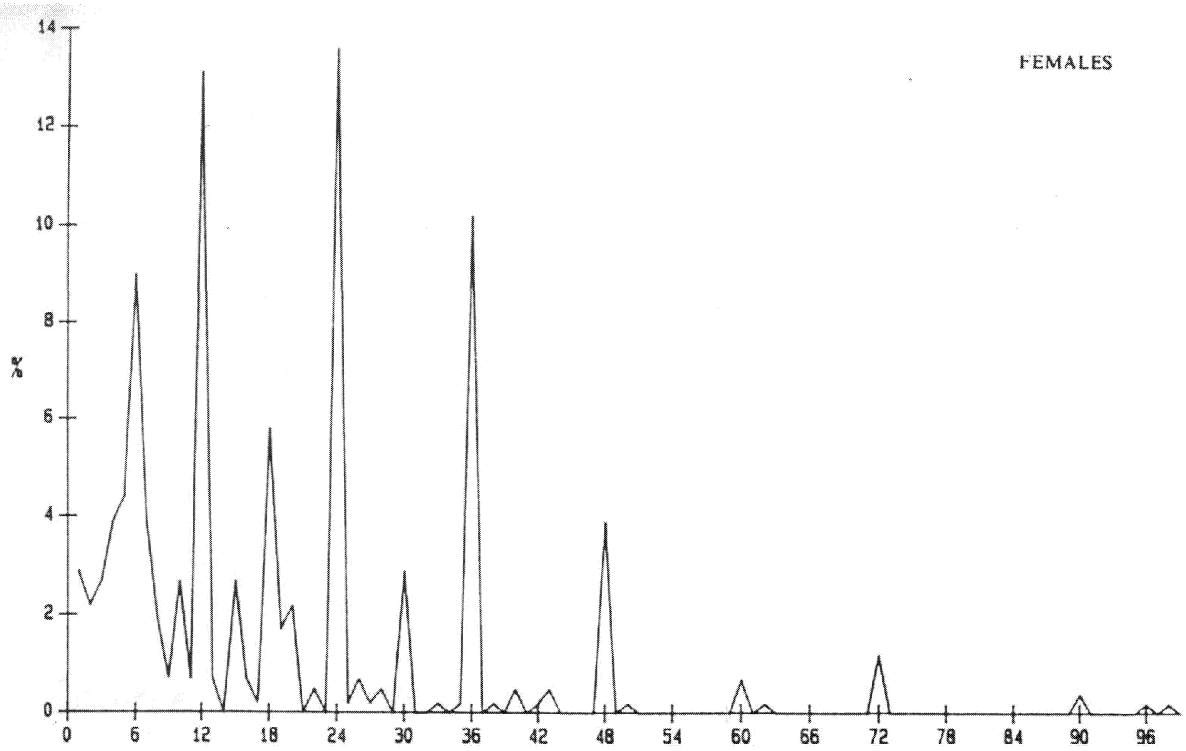


Fig. 2. Percentage distribution of unemployed individuals aged 14–29 years by reported unemployment duration (in months) at initial survey, Italian LFS, matched data for Lombardy, 1986.I–II: males, $N = 267$ and females, $N = 411$.

- heaping in calendar-based designs (German socio-economic panel SOEP)

  * distorted values for entry and leave of state of unemployment
  * bias analysis under simplified assumptions: Augustin & Wolff (2004, Stat.Papers)
  * simulation study (with data constellation based on the SOEP): Wolff & Augustin (2003, ASTA); Jürgens (2007, JRSS A)

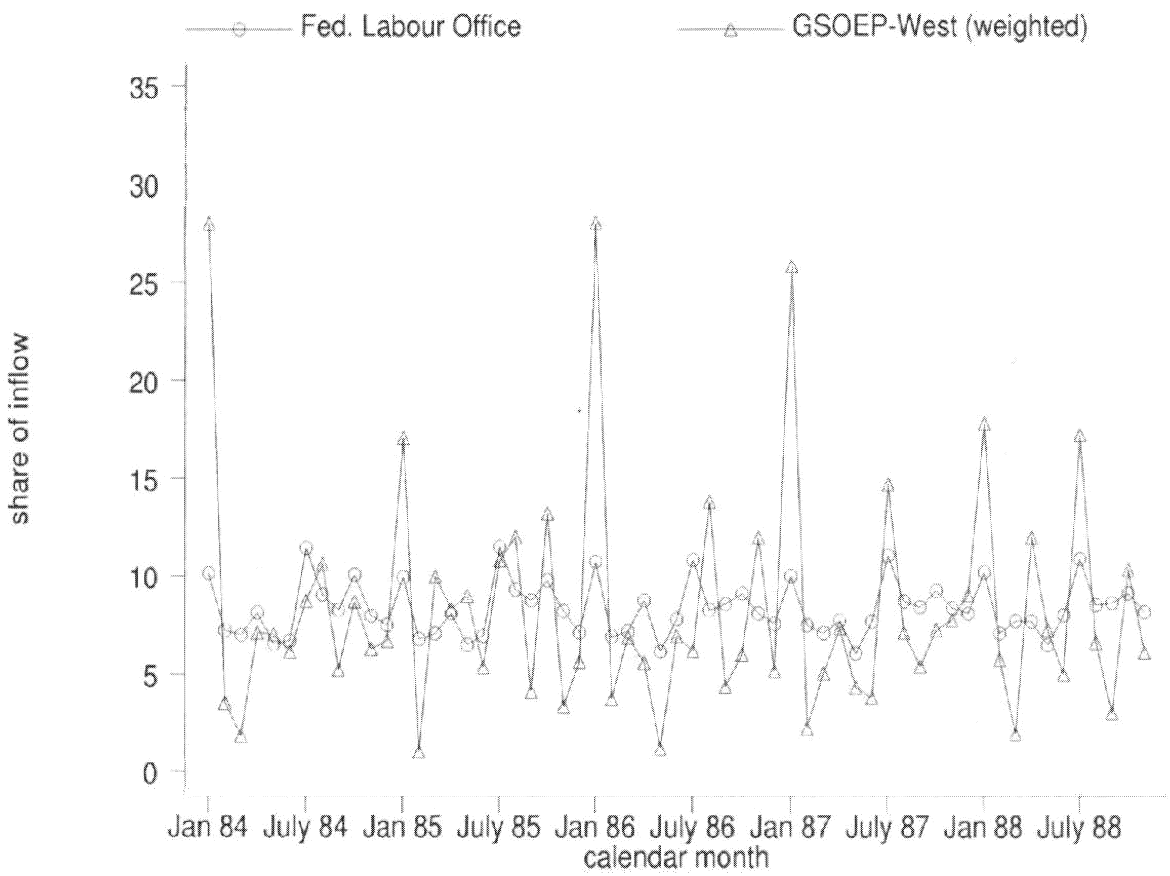


FIGURE 1. Proportion of the annual inflow into registered unemployment in each calendar month – Women, West-Germany.

Anonymisation Techniques

- Recent trend in official statistics: public use files (Statistisches Bundesamt, 2005, Statistik und Wissenschaft, Bd. 4) and big economic research institutes (IAB, Nuremberg)
- Error mechanism known!
- Distortion by classical measurement error or misclassification
- Often other techniques, e.g. micro-aggregation (Schmid, Schneeweiß, Küchenhoff (2007) Stat. Neerl.; Schmid, (2007, Diss. LMU))
- Growing importance in biometrics
- Discussion on "privacy preserving data mining"

2.2 Unbiased Estimating Equations and Corrected Score Functions for Classical Measurement Error (in the Cox Model)

Corrected Score Functions (and the Cox Model under Covariate Measurement Error)

- Frame the problem in terms of *unbiased estimating functions (score functions)* for the parameter ϑ

$$s^{\mathbf{X}}(\mathbf{Y}; \mathbf{X}; \vartheta) \quad \text{such that} \quad \mathbb{E}_{\vartheta_0}(s^{\mathbf{X}}(\mathbf{Y}; \mathbf{X}; \vartheta_0)) = 0$$

  at the true parameter value ϑ_0

  * (Huber: (1981, Wiley) M-estimators; Godambe (1991, Oxford UP); Rieder (1994, Springer): Robust Asymptotic Statistics; Wansbeek & Meijer (2000, Elsevier): GMM)

  * Under regularity conditions still consistency and asymptotic normality.

- For the moment classical covariate measurement error only

$$X_i^* = X_i + U_i\,, \quad X_i \perp U_i\,.$$

- Note that typically, even if $\mathbb{E}(X^*) = \mathbb{E}(X)$ then $\mathbb{E}((X^*)^r) \neq \mathbb{E}(X^r), \quad r > 1.$

- Therefore *naive estimation* by simply replacing $\mathbf{X}$ with $\mathbf{X}^*$, leads in general to

$$\left|\mathbb{E}_{\vartheta_0}\left(s^{\mathbf{X}}\left(\mathbf{Y};\mathbf{X}^*;\vartheta_0\right)\right)\right| \geq a > 0\,,$$

resulting in inconsistent estimators. For instance,

$$\mathbb{E}\left(\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \cdot X_i^*)\begin{pmatrix}1\\ X_i{}^*\end{pmatrix}\right) \neq \mathbb{E}\left(\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \cdot X_i)\begin{pmatrix}1\\ X_i\end{pmatrix}\right) = 0$$

- Measurement error correction: Find an estimating function $s^{X^*}(\mathbf{Y}, \mathbf{X}^*, \vartheta)$ in the error prone data with

$$\mathbb{E}_{\vartheta_0} s^{X^*}(\mathbf{Y};\mathbf{X}^*;\vartheta_\mathbf{0}) = \mathbf{0}.$$

Corrected score functions

(Stefanski (1989, Comm. Stat. Theory Meth.), Nakamura (1990, Biometrika))

- Use the ideal score function as a building block, $s^{X^*}(\mathbf{Y}, \mathbf{X}^*, \vartheta)$ such that

$$\mathbb{E}(s^{X^*}(\mathbf{Y}, \mathbf{X}^*, \vartheta)|\mathbf{X}, \tilde{\mathbf{Y}}) = \mathbf{s}^{\mathbf{X}}(\mathbf{Y}, \mathbf{X}, \vartheta) \ ,$$

- Then (via iterated expectation!), indeed:

$$\begin{aligned} \mathbb{E}_{\vartheta_0}\left(s^{X^*}\left(\mathbf{Y}; \mathbf{X}^*; \vartheta_0\right)\right) &= \mathbb{E}_{\vartheta_0}\left(\mathbb{E}_y\left(s^{X^*}\left(\mathbf{Y}; \mathbf{X}^*; \vartheta_0\right) \mid \mathbf{X}, \mathbf{Y}\right)\right) \\ &= \mathbb{E}_{\vartheta_0}\left(s^{X}\left(\mathbf{Y}; \mathbf{X}; \vartheta_0\right)\right) = \mathbf{0} \ . \end{aligned}$$

- Sometimes indirect proceeding: corrected log-likelihood $l^{X^*}(\mathbf{Y}, \mathbf{X}, \vartheta)$ with

$$\mathbb{E}(l^{X^*}(\mathbf{Y}, \mathbf{X}^*, \vartheta)|\mathbf{X}, \mathbf{Y}) = l^{X}(\mathbf{Y}, \mathbf{X}, \vartheta).$$

  Under regularity conditions corrected score function by taking the derivative.

+ Functional method: no (unjustified !?) assumptions on the distribution of X

+ Successful for generalized linear models, polynomial regression, etc. (Survey: Schneeweiß & Augustin, 2006, ASTA, Hübler & Frohn (eds.))

+ Extensions to misclassification (Akazawa, Kinukawa, Nakamura, 1998, J. Jap. Stat. Soc.; Zucker, Spiegelman, 2008, Stat. Med.)

+ Quite general error distribution can often be handled (only existing moment generating function needed)

– Numerical difficulties for small samples

– Handling of transformations (e.g. $\ln X$) complicated or impossible

– Non-existence of corrected score functions for some models

Tabelle 1: $\beta = 1, \nu = 1.2, \lambda = 1, \texttt{size} = 1000$, berechnet aus 100 Schätzern, Varianz des wahren Prädiktors: 1.0, relative Median Fehler (in Prozent)

Fehler-varianz	Zensierung (in %)	rel. Median Fehler für β korrigiert	naiv
0.0	10	0.00	0.00
0.0	40	0.00	0.00
0.0	70	0.00	0.00
0.1	10	0.36	-13.20
0.1	40	1.59	-11.53
0.1	70	1.36	-10.38
0.3	10	0.43	-31.88
0.3	40	1.92	-29.61
0.3	70	1.46	-27.06
0.5	10	-0.76	-44.61
0.5	40	4.75	-41.41
0.5	70	0.96	-38.57
0.7	10	-0.50	-52.42
0.7	40	1.68	-49.89
0.7	70	2.69	-46.89

Example: An Exact Correction for the Breslow Likelihood

(Augustin (2004, Scand. J.Stat.))

- Nakamura (1992, Biometrics): method not applicable to partial likelihood, seemingly approximate estimator
- The Breslow loglikelihood (Breslow (1972, JRSS SerB; 1974, Biometrics)), based on $\lambda_0(t) := \lambda_j$, $\tau_{j-1} \leq t \leq \tau_j$,

$$\ln(\mathcal{L}(\lambda_1, \ldots, \lambda_k, \beta)) = \sum_{j=1}^{k} \Big(\ln \lambda_j + \beta' X_{(j)} - \lambda_j (\tau_j - \tau_{j-1}) \sum_{i \in R(\tau_j)} \exp\left(\beta' X_i\right) \Big)$$

does not have singularities and

$$\sum_{j=1}^{k} \left(\ln \lambda_j + \beta' X^*_{(j)} - \lambda_j (\tau_j - \tau_{j-1}) \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' X^*_i)}{M_{U_i}(\beta)} \right)$$

is a corrected log-likelihood.

Proof / construction principle:

- Start with the naive log-likelihood
- It is sufficient to find constants c_1, c_2 such that

$$\mathbb{E}\left(\sum_{j=1}^{k} \left(\ln \lambda_j + c_1 \cdot \beta' X^*_{(j)} - \lambda_j(\tau_j - \tau_{j-1}) \sum_{i \in \mathcal{R}(\tau_j)} \frac{\exp(\beta' X_i^*)}{c_2} \right) \mid \mathbf{X}, \mathbf{Y} = \right)$$

$$\sum_{j=1}^{k} \left(\ln \lambda_j + \beta' X_{(j)} - \lambda_j(\tau_j - \tau_{j-1}) \sum_{i \in R(\tau_j)} \exp\left(\beta' X_i\right) \right)$$

- For that purpose note that: $\mathbb{E}\left(\beta' X_i^* | \mathbf{X}, \mathbf{Y}\right) = \beta' X_i$; $\mathbb{E}\left(\exp(\beta' X_i^*) | \mathbf{X}, \mathbf{Y}\right) = \exp(\beta' X_i) \cdot M_{U_i}(\beta)$

2.3. Extended Corrected Score Functions – A Unified View at Measurement Error and Censoring

Extended Corrected Score Functions (and Parametric Survival Models)

- Parametric survival models: exponential, log-normal, Weibull, Gamma, log-logistic
- Superstructure: accelerated failure time models

$$\ln T_i = \beta_0 + \beta_X{}' \cdot X_i + \psi \cdot \epsilon\,, \quad \psi > 0,$$

- $\mathbb{E}(T_i) = \exp(\beta_0) \cdot \exp(\beta_X{}' \cdot X_i) \cdot \mathbb{E}(\exp(\psi \cdot \epsilon))$
- Use quasi-likelihood approach with quasi-score function

$$s^X\left(\vec{T}; \mathbf{X}; \beta\right) = \sum_{i=1}^{n} \frac{\partial\ \mathbb{E}\left[T_i | X_i; \beta\right]}{\partial\ \beta} \cdot \frac{\{T_i - \mathbb{E}\left[T_i | X_i; \beta\right]\}}{\mathbb{V}\left[T_i | X_i; \beta, \zeta\right]} = 0$$

- Use the accelerated failure time model as a superstructure!

$\Downarrow$ General form of the 'ideal' quasi-score equation based on

$$\mathbb{E}\left[T_i^r | X_i; \beta, \psi, \gamma\right] = \underbrace{c_r(\psi, \gamma)}_{distribution} \cdot \underbrace{\exp\left(r \cdot (\beta_0 + \beta_X' \cdot X_i)\right)}_{covariates}$$

$\Downarrow$ General form of the corrected quasi-score equation

$\Downarrow$ Measurement error correction in a concrete model through appropriate choice of $c_r(\cdot)$

$\Downarrow$ Adopt for deficient dependent variables (measurement error/censoring)

Handling Deficient Dependent Variables

- **Censoring**: Instead of T_i one observes

$$T_i^* = \min(T_i, C_i) \quad \text{and} \quad \Delta_i := I\{T_i \leq C_i\}$$

- **Measurement error**: (with $V_i \perp$ rest, $\mathbb{E}(V_i) = m$, $\mathbb{V}(V_i) = v$)
  * additive ($T_i^* = T_i + V_i$)
  * multiplicative ($T_i^* = T_i \cdot V_i$)
    response effects, memory effects (e.g., Skinner & Humphreys (1999, Lifetime Data Analysis))

Unifying approach: $\Psi^{\mathbf{T}^*,\mathbf{X}^*}(\mathbf{T}^*;\mathbf{X}^*,\vartheta)$ extended corrected score function:

$$\mathbb{E}(\Psi^{\mathbf{T}^*,\mathbf{X}^*}(\mathbf{T}^*;\mathbf{X}^*,\vartheta)|\mathbf{T},\mathbf{X}) = \Psi^{\mathbf{T};\mathbf{X}}(\mathbf{T},\mathbf{X},\vartheta)\,,$$

where $\Psi^{\mathbf{T},\mathbf{X}}(\mathbf{T};\mathbf{X};\vartheta)$ is an ideal unbiased estimating function.

It leads, again by the law of iterated expectation, to unbiased estimating function.

Theorem: Let y $\mathbf{W} \in \{X_i^*, X_i\}$, and $\Psi^{\mathbf{T}}(\mathbf{T_i}, \mathbf{W}, \vartheta)$ be an unbiased estimating function of the very general form (including ML estimation in exponential families or Weibull type models, corrected QL estimation)

$$\psi^{\mathbf{T}}(\mathbf{T_i}, \mathbf{W}, \vartheta) = \sum_{i=1}^{n} \sum_{l=0}^{q} c_l(W_i, \vartheta) \cdot {T_i}^{\alpha_l}.$$

Then, under multiplicative measurement error,

$$\psi^{\mathbf{T}^*}(\mathbf{T}^*, \mathbf{W}, \vartheta) = \sum_{i=1}^{n} \sum_{l=0}^{q} c_l(W_i, \vartheta) \cdot \frac{\mathbf{T_i}^{*\alpha_l}}{\mathbb{E}({V_i}^{\alpha_l})}$$

is an extended corrected score function. Similar results hold for the additive model, if $\alpha_l \in \mathbb{N}$.

Corollary: $q = 1, \alpha = 1, \mathbb{E}(V_i) = 1 \Longrightarrow \widehat{\vartheta}_{ML,naive} \longrightarrow \vartheta$ for exponential families under unbiased measurement error in the *dependent* variable.

Censoring Instead of T_i one observes

$$T_i^* := \min(T_i, C_i) \quad \text{and} \quad \Delta_i := I(\{T_i \leq C_i\})$$

with C_i as the censoring variable.

Solve censoring by inverse probability weighting (e.g. Zhou (1992, Biometrika), Graf et al. (1999, Stat. Med.), Augustin (2002, Habil.), van der Laan & Robins,(2003, Springer), Hothorn et al (2006, Biostatistics), Gerds & Schuhmacher (2006, Biom. J.))

Consider independent random censoring, $W \in \{X, X^*\}$, $C_1, \ldots, C_n$ i.i.d. with $G(t) := P(C_i \geq t|W_i) = P(C_i \geq t) > 0\,, \ \forall t \in \mathbb{R}_+$. Then for "every" $g(\cdot)$

$$\mathbb{E}(\Delta \cdot \frac{g(T_i^*, W, \vartheta)}{G(T_i^*)}|W_i, T_i) = g(T_i, W_i, \vartheta)$$

Theorem: Consider independent random censoring, $W \in \{X, X^*\}$, $C_1, \ldots, C_n$ i.i.d. with $G(t) := P(C_i \geq t | W_i) = P(C_i \geq t) > 0\,,\ \forall t \in \mathbb{R}_+\,,$ and an unbiased estimation function for ϑ of the form

$$\psi^*(\mathbf{T}, \mathbf{W}, \vartheta) = \sum_{i=1}^{n} a_0(W_i, \vartheta) + \sum_{l=1}^{q} a_l(W_i, \vartheta) \cdot T_i{}^{\alpha_l} \qquad [C]$$

Then

$$\psi^{**}(\mathbf{T}^*, \mathbf{W}, \vartheta) = \sum_{i=1}^{n} \frac{\Delta_i}{G(T_i^*)} \cdot \psi_i^*(T_i^*, W_i, \vartheta)$$

and

$$\psi^{***}(\mathbf{T}^*, \mathbf{W}, \vartheta) = \sum_{i=1}^{n} \left(a_0(W_i, \vartheta) + \sum_{l=1}^{q} a_l(W_i, \vartheta) \cdot \frac{\Delta_i}{G(T_i^*)} \cdot T_i^{*\alpha_l} \right)$$

are extended corrected score functions.

Proof: $\Delta_i = 0 \iff T_i^* \neq T_i$. Therefore:

$$\frac{\Delta_i}{G(T_i^*)} \cdot g(T_i^*, W_i, \vartheta) = \frac{\Delta_i}{G(T_i)} \cdot g(T_i, W_i, \vartheta)$$

and

$$\begin{aligned}
& \mathbb{E}\left(\frac{\Delta_i}{G(T_i^*)} \cdot g(T_i^*, W_i, \vartheta) \middle| W_i, T_i \right) \\
= \quad & \mathbb{E}\left(\frac{\Delta_i}{G(T_i)} \cdot g(T_i, W_i, \vartheta) \middle| W_i, T_i \right) \\
= \quad & 1 \cdot \frac{P\left(\Delta_i = 1 | W_i, T_i\right)}{G(T_i)} \cdot g(T_i, W_i, \vartheta) + 0 \\
= \quad & g(T_i, W_i, \vartheta)\,,
\end{aligned}$$

due to $P\left(\Delta_i = 1 | W_i, T_i\right) = P(C_i \geq T_i | W_i, T_i) = G(T_i)$.

- $$P\left(\Delta_i = 1 | W_i, T_i\right) = P(C_i \geq T_i | W_i, T_i) = G(T_i)$$

  Is it possible to allow dependence on W_i indeed?
  Note, W is observable.

- Estimate $G(\cdot)$ in a nonparametric way: Kaplan-Meier-estimator

- In the case of $\psi^{**}(\cdot)$ condition [C] can be replaced by $\psi^*(\mathbf{T}, \mathbf{W}, \vartheta) = \sum_{i=1}^n \psi_i^*(T_i, W_i, \vartheta)$.

- Utilize the generality of the censoring lemma!

Direct extensions:
Utilize the generality of the censoring lemma

- Use soft weighting by truncating large pseudo-observation (Hothorn et al. (2006, Biostat.))
- Consistency under nonparametric estimation of $G(\cdot)$ (with Kukush, Usoltseva):
- Extend sophisticated methods to handle measurement error in the linear model (Cheng & van Ness (1999, Arnold); Wansbeek & Meijer (2000, Elseiver)) and the polynomial model (Cheng & Schneeweiss (1998, JRSS, Ser.B, 2002, TLS)), (Wansbeek & Meijer (2000, Elsevier, Chap. 11)) to the nonparametric AFT

$$\ln(T_i) = \beta' X_i + \epsilon_i \qquad \text{with } \epsilon_i \text{ unspecified.}$$

- M-estimators under censorship
- handling of censored **in**dependent variables!?

- but be careful with standard application of weighting procedures: compare Basu's elephant in nonparametric statistics (Einbeck & Augustin, (2009, Statistica Sinica))

2.4. Corrected Score Functions for Berkson Models

- Up to now, the method of corrected score functions has not been applied to the Berkson model.

- Recall: Corrrected score function: $s^{X^*}(\mathbf{Y}, \mathbf{X}^*, \vartheta)$ such that

$$\mathbb{E}(s^{X^*}(\mathbf{Y}, \mathbf{X}^*, \vartheta)|\mathbf{X}, \tilde{\mathbf{Y}}) = \mathbf{s}^{\mathbf{X}}(\mathbf{Y}, \mathbf{X}, \vartheta)$$

- Only in the classical error model

$$[X_i^*|X_i; Y_i]$$

  is directly available

- Idea: In the Berkson model

$$[X_i^*|X_i, Y_i]$$

  can be obtained under additional assumptions, e.g., on the marginal distribution of X_i^*.

Let, for the moment, for $i = 1, \ldots, n$,

- $X_i^* \sim N\left(\mu^*, \sigma^{*2}\right)$; w.l.o.g. $\mu^* = 0$
- $\delta_i \sim N\left(0, {\sigma_{\delta_i}}^2\right)$

Berkson model $X_i = X_i^* + \delta_i$; $\qquad X_i^* \perp \delta_i$

Then

$$X_i^* | X_i, Y_i \sim N\left(\tilde{\mu}; \tilde{\sigma}^2\right)$$

with

$$\tilde{\mu} \quad = \quad X_i \left(1 - \frac{\sigma_{\delta_i}^2}{{\sigma_X}^2}\right)$$

and

$$\tilde{\sigma}^2 = \frac{\sigma_{\delta_i}^2 \sigma_{X_i^*}^2}{\sigma_{X_i}^2}$$

Theorem: Corrected loglikelihood under Berkson error:

- Poisson regression:

$$\begin{aligned}
\ell(X, Y, \beta) &= \sum_{i=1}^{n} \beta X_i Y_i - \exp(\beta X_i) \\
\ell_{naive}(X^*, Y, \beta) &= \sum_{i=1}^{n} \beta X_i^* Y_i - \exp(\beta X_i^*) \\
\ell_{corr}(X^*, Y^*, \beta) &= \sum_{i=1}^{n} \frac{\sigma_X^2}{\sigma^{*2}} \cdot \beta X_i^* Y_i - \exp\left(\frac{\sigma_X^2}{\sigma^{*2}} \beta X_i^* - \frac{\sigma_{\delta_i}^2 \sigma_{X_i}^2}{2\sigma^{*2}} \beta^2 \right)
\end{aligned}$$

- Cox model: Breslow log-likelihood, based on failure times τ_j, $j = 1, \ldots, k$,
  * $\mathcal{D}(\tau_j)$ deaths at τ_j; $d_j := |\mathcal{D}(\tau_j)|$, $\mathcal{R}(\tau_j)$ risk set,
  * λ_j piecewise constant part of baseline hazard rate $\lambda_0(t)$

$$\ell(X, Y, \beta) = \sum_{i=1}^{n} d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta X_i - \lambda_j(\tau_{j+1} - \tau_j) \sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta X_i)$$

$$\ell_{naive}(X^*, Y, \beta) = \sum_{i=1}^{n} d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \beta X_i^* - \lambda_j(\tau_{j+1} - \tau_j) \sum_{i \in \mathcal{R}(\tau_j)} \exp(\beta X_i^*)$$

$$\ell_{corr}(X^*, Y, \beta) = \sum_{j=1}^{k} \left(d_j \ln \lambda_j + \sum_{i \in \mathcal{D}(\tau_j)} \frac{\sigma_X^2}{\sigma^{*2}} \cdot \beta X_i^* - \right.$$

$$\left. -\lambda_j(\tau_{j+1} - \tau_j) \sum_{i \in \mathcal{R}(\tau_j)} \exp\left(\frac{\sigma_X^2}{\sigma^{*2}} \beta X_i^* - \frac{\sigma_{\delta_i}^2 \sigma_{X_i}^2}{2\sigma^{*2}} \beta^2 \right) \right),$$

Construction

- Both models have a similar structure
  * linear term: βX_i and $(\beta X_i \cdot Y_i)$, resp.
  * exponential term: $\exp(\beta X_i)$

- For finding a corrected log likelihood

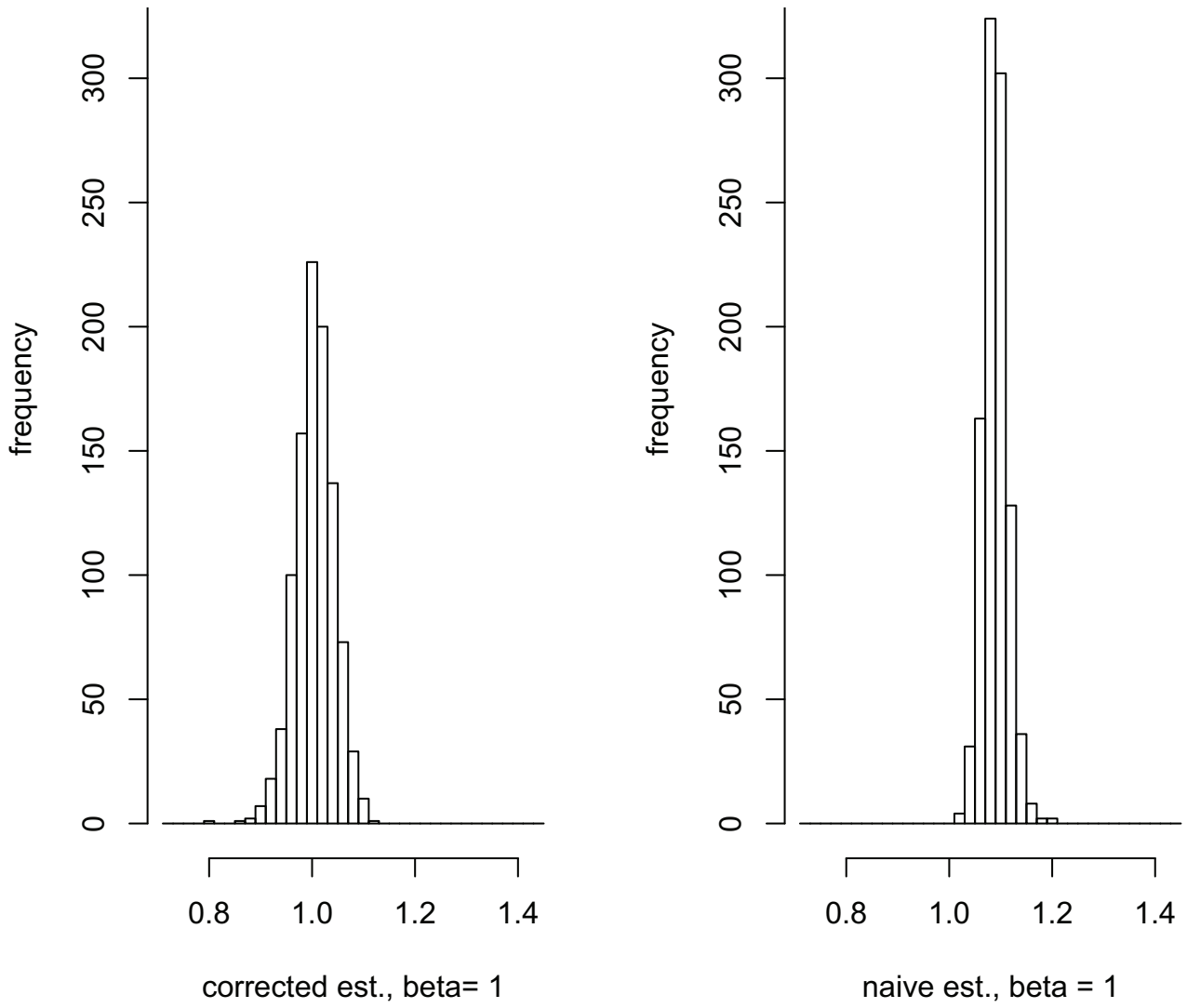
$$\mathbb{E}(\mathrm{l}^{X^*}(\mathbf{Y}, \mathbf{X}^*, \vartheta)|\mathbf{X}, \mathbf{Y}) = \mathrm{l}^{X}(\mathbf{Y}, \mathbf{X}, \vartheta).$$

  it is sufficient that there exist c_1, c_2, c_3, c_4 such that

$$\mathbb{E}\Big((c_1 \beta X_i^* + c_2)|\, X_i, Y_i\Big) \stackrel{!}{=} \beta X_i$$

  and

$$\mathbb{E}\Big((\exp(c_3 \beta X_i^* + c_4))|\, X_i, Y_i\Big) \stackrel{!}{=} \exp(\beta X_i).$$


frequency
0
50
100
150
200
250
300
0.8
1.0
1.2
1.4
corrected est., beta= 1
frequency
0
50
100
150
200
250
300
0.8
1.0
1.2
1.4
naive est., beta = 1

- Unbiasedness of naive score tests ???