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

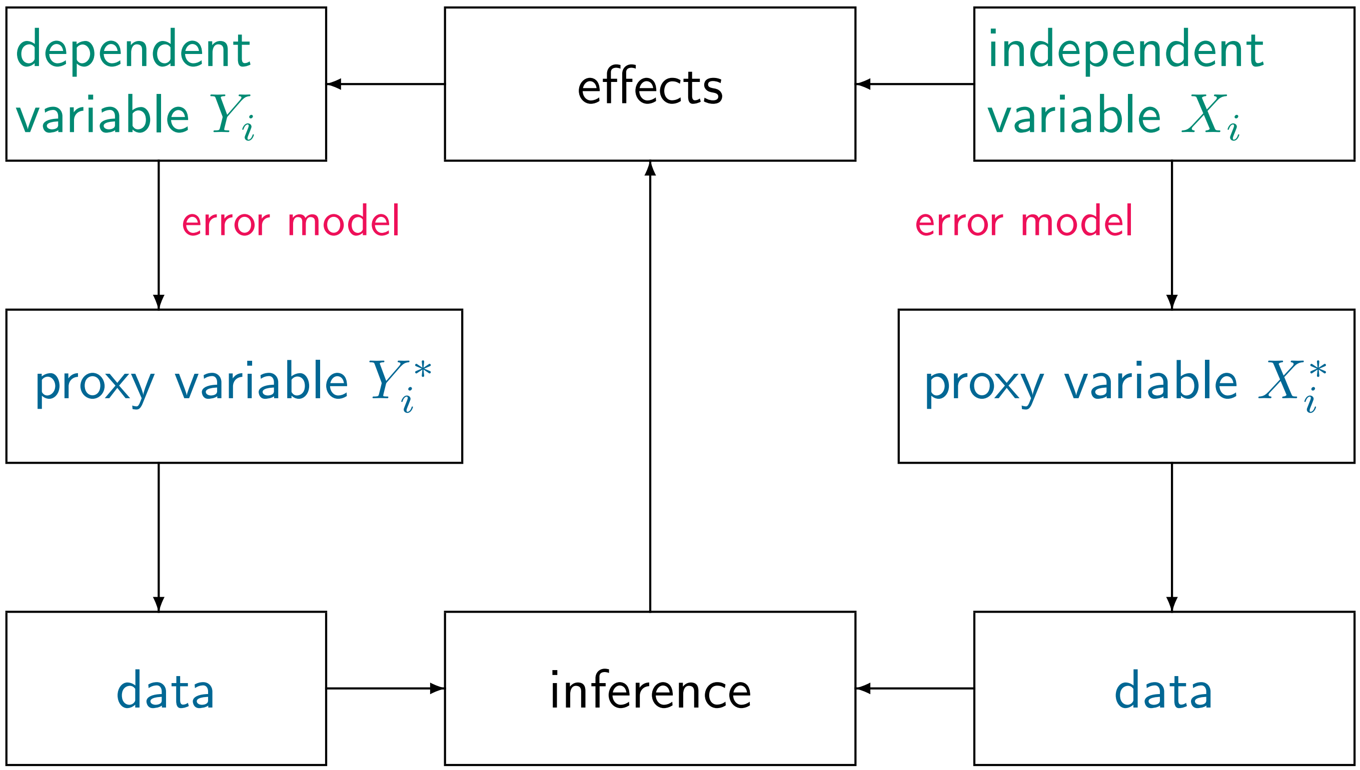
Second Talk

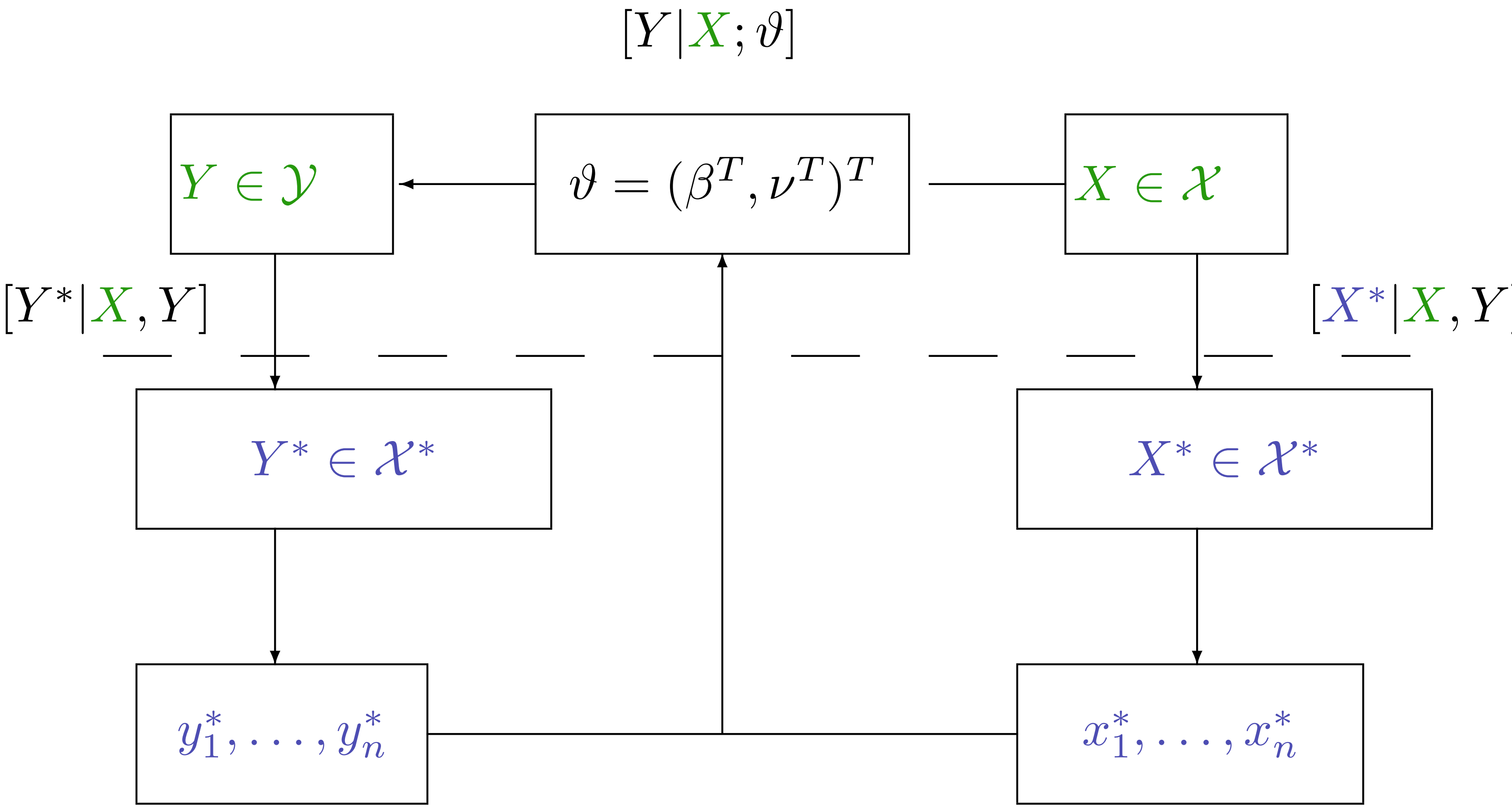
Thomas Augustin
Department of Statistics,
Ludwig-Maximilians Universität Munich (LMU)

- A Brief Look at the First Talk
- The Technical Argument Condensed
- Some Results on Direct Correction In the Poisson Model

3. Overcoming the Dogma of Ideal Precision in Deficiency Models

3.1 Credal Deficiency Model as Imprecise Measurement Error Models
3.2 Credal Consistency of Set-Valued Estimators
3.3 Minimal and complete Sets of Unbiased Estimating Functions
3.4 Some Examples


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



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

The triple whammy effect of measurement error
Carroll, Ruppert, Stefanski, Crainiceanu (2006, Chap.H.)

– bias
– masking of features
– loss of power

• **classical error**: "attenuation "

Results

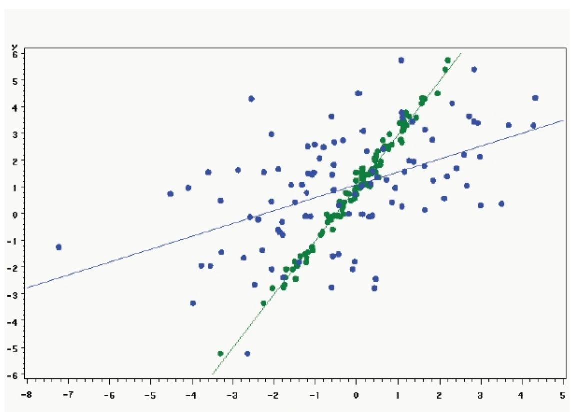


Figure 1: Effect of additive measurement error on linear regression

Terminology

continuous variables

$\downarrow$

measurement error

$\downarrow$

error U_i

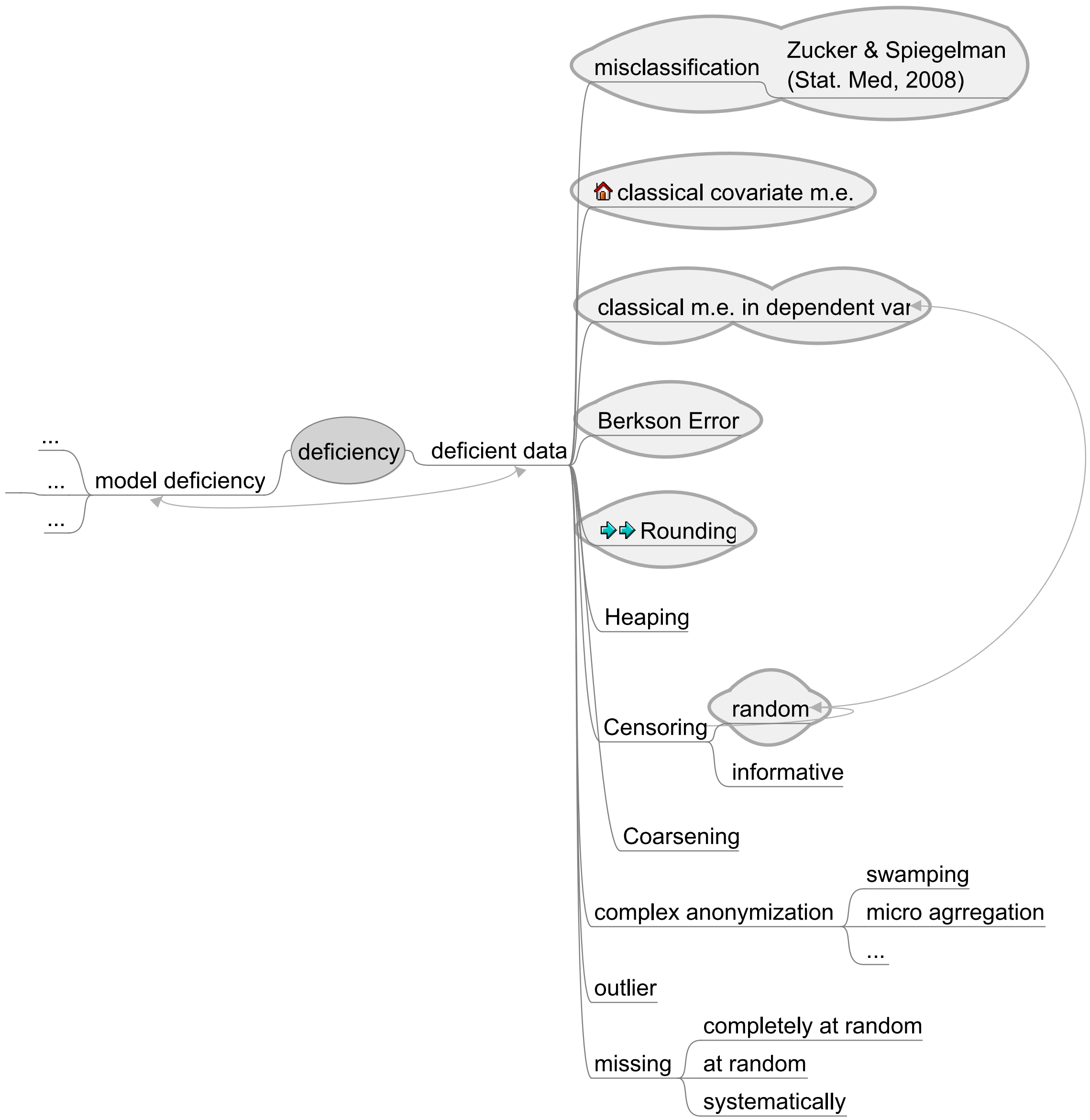
$\swarrow$	$\downarrow$	$\searrow$	$\searrow\ \swarrow$
classical meas. error	Berkson error	Rounding Heaping	Anonymization Techniques
$X_i \perp U_i$	$X_i^* \perp U_i$		

discrete variables

$\downarrow$

misclassification

$\downarrow$



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

2. Measurement Error Correction based on Precise Error Models

2.1 Measurement Error Modelling

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

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

2.4 Corrected Score Functions for Berkson Models

(2.5) (Unconditionally Corrected Score Functions and Rounding) (Felderer)

The Technical Argument Condensed

On the construction of unbiased estimating equations:

- ϑ_0 true parameter value
- Ideal estimating function: $\psi^{X,Y}(\mathbf{X}, \mathbf{Y}, \vartheta)$
- Naive estimating function: $\psi^{sic!\ X,Y}(\mathbf{X}^*, \mathbf{Y}^*, \vartheta)$
- Find $\psi^{X^*,Y^*}(\mathbf{X}^*, \mathbf{Y}^*, \vartheta)$ such that

$$\mathbb{E}_{\vartheta_0}\left(\psi^{X^*,Y^*}(\mathbf{X}^*, \mathbf{Y}^*, \vartheta)\right) \stackrel{!}{=} 0 \qquad (*)$$

- Idea: use the ideal score function as a building block!
- Try $\psi^{X^*,Y^*}(\mathbf{X}^*, \mathbf{Y}^*, \vartheta) = f(\psi^{X,Y}(\mathbf{X}^*, \mathbf{Y}^*, \vartheta))$ for some appropriate $f(\cdot)$

- In general, $\psi^{X^*,Y^*}(\cdot)$ can not be determined directly.

- Note that, since $\mathbb{E}_{\vartheta_0}\left(\psi^{X,Y})(X,Y,\vartheta)\right) = 0$, $(*)$ is equivalent to

$$\mathbb{E}_{\vartheta_0}\left(\psi^{X^*,Y^*}(\mathbf{X}^*,\mathbf{Y}^*,\vartheta)\right) = \mathbb{E}_{\vartheta_0}\left(\psi^{X,Y}(\mathbf{X},\mathbf{Y}^*,\vartheta)\right)$$

- Look at the expected difference between $\psi^{X^*,Y^*}(\cdot)$ and $\psi^{X,Y}(\cdot)$.

- Try to break $\psi^{X,Y}(\mathbf{X},\mathbf{Y},\vartheta)$ into „additive pieces“, and do it piece by piece

- Typically, $\psi(\cdot)$ has the form

$$\psi(X,Y,\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\psi_i(X_i,Y_i,\vartheta),$$

and there are representations such that, for $i = 1, \ldots, n$,

$$\psi_i(X_i, Y_i, \vartheta) = \sum_{j=1}^{s} g_j(X_i, Y_i, \vartheta).$$

- Then try to find $f_1(\cdot), \ldots, f_s(\cdot)$ such that

$$\mathbb{E}_{\vartheta_0}\left(f_j(g_j(X_i^*, \mathbf{Y}_\mathbf{i}^*, \vartheta))\right) = \mathbb{E}_{\vartheta_0}\left(g_j(X_i, \mathbf{Y}_\mathbf{i}, \vartheta)\right) \qquad (**)$$

- (conditionally/locally) corrected score functions (Nakamura (1990, Biometrika), Stefanski (1989, Comm. Stat. Theory Meth.))

- Try to find $f_1(\cdot), \ldots, f_s(\cdot)$ such that

$$\mathbb{E}_{\vartheta_0}\left(f_j(g_j(X_i^*, \mathbf{Y}_\mathbf{i}^*, \vartheta))|X_i, \mathbf{Y}_\mathbf{i}\right) = g_j(X_i, \mathbf{Y}_\mathbf{i}, \vartheta) \quad (**),$$

then the law of iterated expectation leads to (*).

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

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

or

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

- Same techniques as before

  * piece by piece
  * globally or locally

- Under regularity conditions unbiased estimating function by taking the derivative with respect to ϑ.

Some Results on Direct Correction in the Poisson Model

Berkson Error II: A Direct Correction for the Poisson Model under a Linear Error Structure

- Ideal score function:

$$\mathbb{E}\left(X_i Y_i - X_i \exp(X_i \beta)\right) = 0$$

- Naive score function:

$$\mathbb{E}\left(X_i^* Y_i - X_i^* \exp(X_i^* \beta)\right) = 0$$

- Show that there is $a, c \in \mathbb{R}$ such that

$$\mathbb{E}\left(a X^* Y_i + c \cdot \exp(X^* \beta) - X_i^* \exp(X_i^* \beta)\right) = 0$$

$$
\begin{aligned}
\mathbb{E}(aX^*Y_i) &= \mathbb{E}\left(\mathbb{E}(aX_i^*Y_i|X_i)\right) = \\
&= a \cdot \mathbb{E}\left(\mathbb{E}(X_i^*|X_i) \cdot \mathbb{E}(Y_i|X_i)\right)
\end{aligned}
$$

Here an important difference occurs between the Berkson model and a rounding model. In the latter case $\mathbb{E}(X_i^*|X_i) = X_i^*$ by definition, in the former case assume a linear error structure such that $\mathbb{E}(X_i^*|X_i) = \gamma_0 + \gamma_1 X_i$; $\mathbb{E}(X_i^* + U_i|X_i) = X_i + \mathbb{E}(U|X_i)$

Then, for the Berkson model,

$$
\begin{aligned}
\mathbb{E}(aX^*Y_i) &= a \cdot \mathbb{E}\left((\gamma_1 X_i + \gamma_0) \cdot \exp(X_i\beta)\right) = \\
&= a \cdot \mathbb{E}\left(\gamma_1 X_i \exp(X_i\beta) + \gamma_0 \exp(X_i\beta)\right) = \\
&= a \cdot \mathbb{E}\left(\gamma_1 (X_i^* + U_i) \exp\left((X_i^* + U_i)\beta\right) + \gamma_0 \exp\left((X_i^* + U_i)\beta\right)\right) = \\
&= a \cdot \left(\mathbb{E}\left(\gamma_1 X_i^* \exp(X_i^*\beta) \cdot \exp(U_i\beta) + \right.\right. \\
&\quad + \gamma_1 \cdot U_i \cdot \exp(U_i\beta) \cdot \exp(X_i^*\beta) + \\
&\quad \left.\left. + \gamma_0 \exp(X_i^*\beta) \cdot \exp(U_i\beta)\right)\right)
\end{aligned}
$$

Note that here X^* and U are independent.

Therefore

$$\begin{aligned}\mathbb{E}(aX^*Y_i) &= a\cdot\gamma_1\left(\mathbb{E}(\exp(U_i\beta))\cdot\mathbb{E}\left(X_i^*\exp(X_i^*\beta)\right)+\right.\\ &\quad +\mathbb{E}\left(U_i\exp(U_i\beta)\right)\cdot\mathbb{E}\left(\exp(X_i^*\beta)\right)\left.\right)+\\ &\quad +a\gamma_0\mathbb{E}(\exp(U_i\beta))\cdot\mathbb{E}\left(\exp(X_i^*\beta)\right)\end{aligned}$$

- First condition

$$a\gamma_1\mathbb{E}(U\exp(U\beta))+a\gamma_0\mathbb{E}(\exp(U\beta))+c=0$$

  (Note that γ_1 and γ_0 are fixed, not to be chosen.)

- Second condition

$$a\cdot\gamma_1\mathbb{E}(\exp(U\beta))\mathbb{E}(X_i^*\exp(X_i^*\beta))-\mathbb{E}\left(X_i^*\exp\left(X_i^*\beta\right)\right)\stackrel{!}{=}0$$

- $$\begin{aligned}a &= -\left(\gamma_1\cdot\mathbb{E}(\exp(U\beta))\right)^{-1}\\ c &= -a\gamma_1\mathbb{E}(U\exp(U\beta))-a\gamma_0\mathbb{E}(\exp(U\beta))=\frac{\mathbb{E}(U\exp(U\beta))}{\mathbb{E}(\exp(U\beta))}-\frac{1}{\gamma_1}\cdot\gamma_0\end{aligned}$$

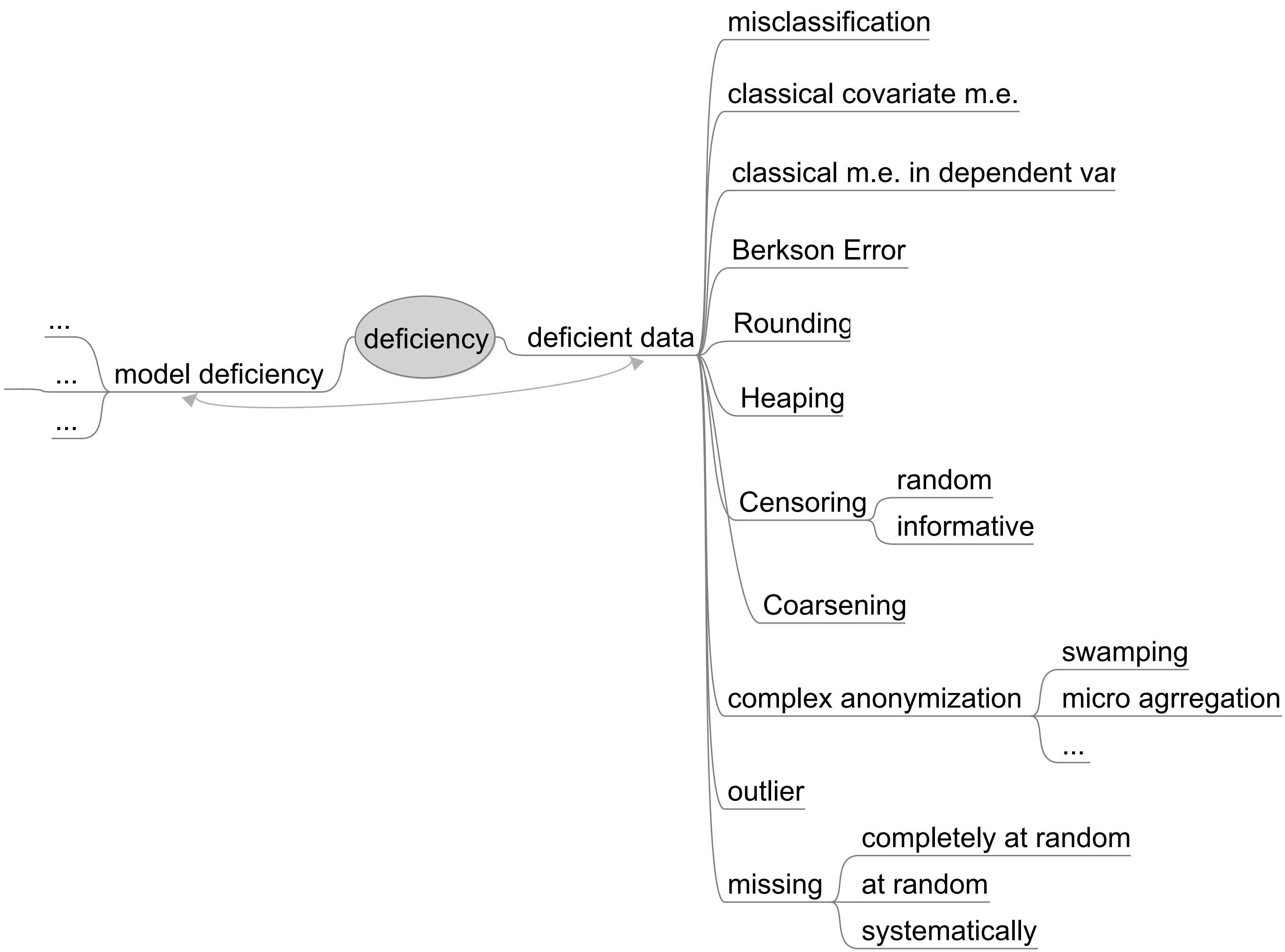
A Direct Correction for Rounding in the Poisson Model

$\mathbb{E}(X_i^*|X_i) = X_i^*$, and therefore

$$\begin{aligned}
\mathbb{E}(aX_i^*Y_i) &= a\mathbb{E}(X_i^* \cdot \exp(X_i\beta)) = \\
&= a\mathbb{E}\left(\mathbb{E}(X_i^* \cdot \exp(X_i\beta)|X_i^*)\right) = \\
&= a\mathbb{E}\left(\mathbb{E}\left(X_i^* \cdot \exp\left((X_i^* + U_i)\beta|X_i^*\right)\right)\right) = \\
&= a\mathbb{E}\left(X_i^* \exp(X_i^*)\right)\mathbb{E}(\exp(U_i\beta|X_i^*)) \\
a &= \left(\mathbb{E}\left(\exp\left(U_i\beta\right)|\, X_i^*\right)\right)^{-1}
\end{aligned}$$

3. Overcoming the Dogma of Ideal Precision in Deficiency Models

3.1 Credal Deficiency Model as Imprecise Measurement Error Models


...
...
...
model deficiency
deficiency
deficient data
misclassification
classical covariate m.e.
classical m.e. in dependent var
Berkson Error
Rounding
Heaping
Censoring
random
informative
Coarsening
complex anonymization
swamping
micro agrregation
...
outlier
missing
completely at random
at random
systematically

Manski's Law of Decreasing Credibility

Reliability !? Credibility ?

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

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" assumption and let the data speak for themselves!
- Consider the *set* of *all* models that maybe compatible with the data (and then add successively additional assumptions, if desirable)
- The results may be imprecise, but are more reliable for sure
- **The extend of imprecision is related to the data quality!**
- As a welcome by-product: clarification of the implication of certain assumptions
- parallel developments (missing data; transfer to measurement error context!)
  - economics: *partial identification:* e.g., Manski (2003, Springer)
  - biometrics: *systematic sensitiviy 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.); Kunz, Augustin, Küchenhoff (2010, TR)

How to proceed with a set of results ?

- Imprecise probabilities (IP)
  - * Roughly speaking: probabilistic modelling with *sets* of models: *credal sets*
  - * Walley (1991, Chapman & Hall), Weichselberger (2001, Physika), Augustin, Coolen, de Cooman, Troffaes (eds., 2009, Proc ISIPTA'09)
  - * Generalized asymptotics: Fierens, Rego, Fine (2008, JSPI), de Cooman, Miranda (2008, JSPI), Cozman (2010, IJAR)

- Construction of unbiased sets of estimating functions

- Credal consistency

Some promising development in IP

- IP approaches for handeling coarsened or missing data: de Comman, Zaffalon (2004, AI), Utkin, Augustin (2007, IJAR)
- Technical handling by generalized BPAs: Miranda, de Cooman, Couso (2004, JSPI), Augustin (2005, IJGS), Coolen, Augustin (2009, IJAR)
- "Soft independence" with given marginals: e.g., Held H., Augustin, Kriegler (2008, IJAR)
- Asymptotics for IP: Fierens, Rego, Fine (2008, JSPI), de Cooman, Miranda (2008, JSPI), Cozman (2010, IJAR)
- Strong relationship to robust statistics: Augustin, Hable (2010, Struct. Safety)
- And to robust Bayesian analysis: e.g., Walter, Augustin (2009a, JSTP; 2009b, Fests. Fahrmeir)

Credal Estimation

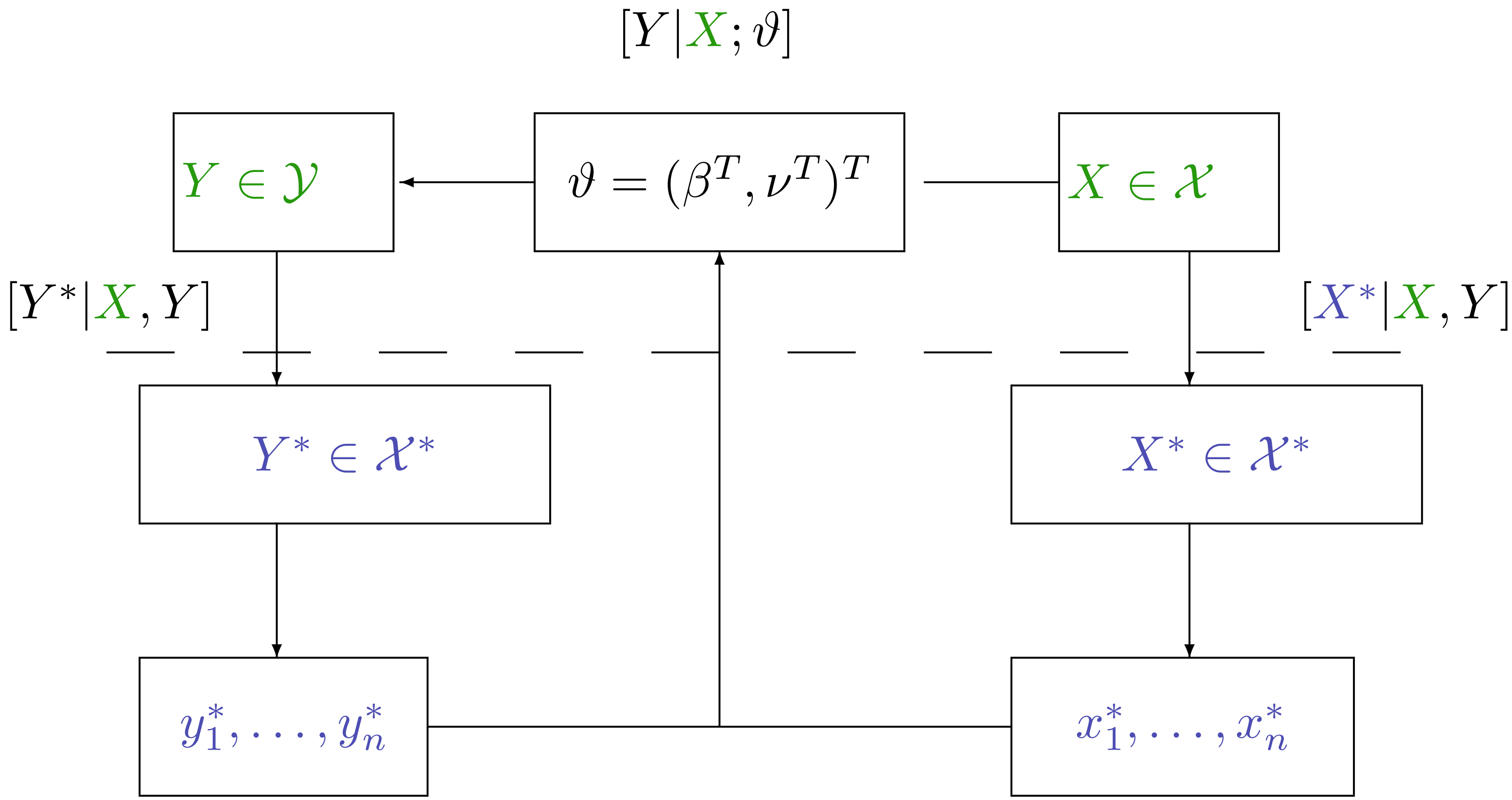
- Natural idea: sets of traditional models $\longrightarrow$ sets of traditional estimators

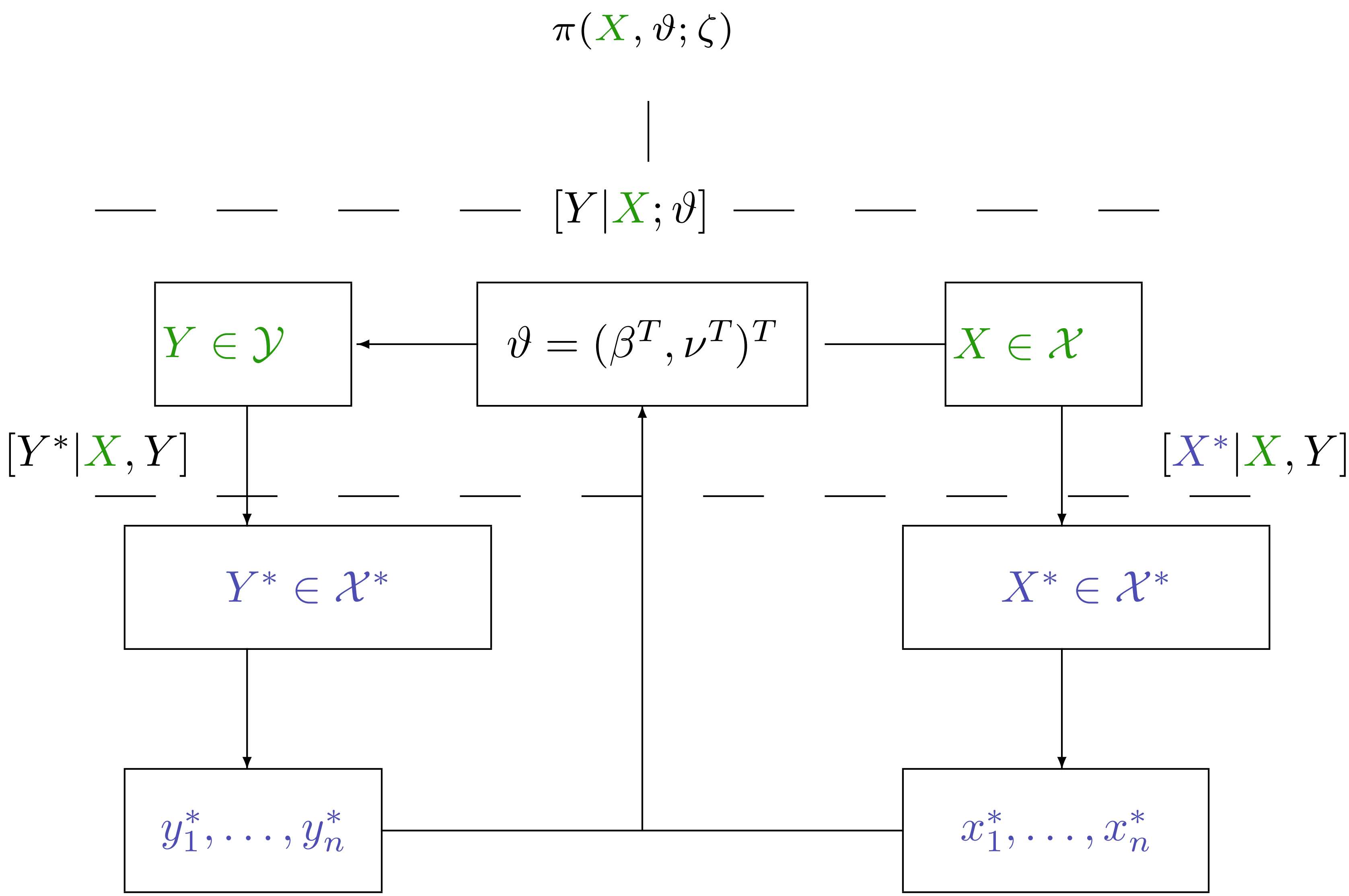
- Construct estimators $\widehat{\Theta} \subseteq \mathbb{R}^p$, set appropriately reflecting the ambiguity (non-stochastic uncertainty, ignorance) in the credal set $\mathcal{P}$.

- $\widehat{\Theta}$ 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


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


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

Credal Deficiency Models

Different types of deficiency can be expressed

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

credal set: convex set of traditional probability distributions

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

3.2 Credal Consistency

- $\left(\widehat{\Theta}^{(n)}\right)_{n\in\mathbb{N}} \subseteq \mathbb{R}^p$ is called *credally consistent* (with respect to the credal set $\mathcal{P}_\vartheta$) if $\forall \vartheta \in \Theta$:

$$\forall p \in \mathcal{P}_\vartheta \; \exists \; \left(\hat{\vartheta}_p^{(n)}\right)_{n\in\mathbb{N}} \in \left(\widehat{\Theta}^{(n)}\right)_{n\in\mathbb{N}} : \operatorname*{plim}_{n\to\infty} \hat{\vartheta}_p^{(n)} = \vartheta.$$

- A credally consistent estimator $\widehat{\Theta}^{(n)}$ is called *minimally* credally consistent if there is no credally consistent estimator $\widehat{\widehat{\Theta}}^{(n)} \subset \widehat{\Theta}^{(n)}$.

3.3 Construction of Minimal Credally Consistent Estimators

- Transfer the framework of unbiased estimating functions
- A set Ψ of estimating functions is called
  * *unbiased* (with respect to the credal set $\mathcal{P}_\vartheta$) if for all ϑ:

$$\forall \psi \in \Psi \; \exists p_{\psi,\vartheta} \in \mathcal{P}_\vartheta : \quad \mathbb{E}_{p_{\psi,\vartheta}}(\Psi) = 0$$

  * *complete* (with respect to the credal set $\mathcal{P}_\vartheta$) if for all ϑ:

$$p \in \mathcal{P}_\vartheta \; \exists \psi_{p,\vartheta} \in \Psi : \quad \mathbb{E}_p(\psi_{p,\vartheta}) = 0.$$

- A complete and unbiased set ψ of estimating functions is called *minimal* if there is no complete and unbiased set of estimating functions $\tilde{\Psi} \subset \Psi$.

Construction of Minimal Consistent Estimators

Define for some set Ψ of estimating functions

$$\widehat{\Theta}_\Psi = \left\{ \hat{\vartheta} \middle| \, \hat{\vartheta} \text{ is root of } \psi,\, \psi \in \Psi \right\}.$$

Under the usual regularity conditions (in particular unit root for every ψ)

- Ψ unbiased and complete $\Rightarrow \widehat{\Theta}_\Psi$ credally consistent
- Ψ minimal $\Rightarrow \widehat{\Theta}_\Psi$ minimally credally consistent

3.4 Examples

- *Imprecise sampling model*: neighborhood model $\mathcal{P}_{Y|X,\vartheta}$ around some ideal central distribution $p_{Y|X,\vartheta}$
  Let ψ be an unbiased estimation function for $p_{Y|X,\vartheta}$. Then (if well defined)
  $$\Psi = \left\{\psi^* \middle|\, \psi^* = \psi - \mathbb{E}_p(\psi),\, p \in \mathcal{P}_{Y|X,\vartheta}\right\}$$
  is unbiased and complete.

- *Imprecise measurement error model*, e.g. $\mathcal{P}_{X^*|X,Y}$:
  $\Psi = \left\{\psi \middle|\, \psi \text{ is corrected score function for some } p \in \mathcal{P}_{X^*|X,Y}\right\}$ is unbiased and complete.

- Construction of confidence regeions:
  * union of traditional confidence regions
  * can often be improved (Vansteelandt, Goetghebeur, Kenward & Molenberghs (Stat Sinica, 2006), Stoye (2009, Econometrica)).

Appendix A: Some Cautious Rounding Models

$[X_i | X^*] = [X_i^* + U_i | X_i^*]$ with $[U_i | X_i^*] \in \mathcal{M}_{X^*}$: any distribution on the rounding interval $\mathbb{I}$

$\mathbb{I}(X_i^*) = [\underline{\mathbb{I}}(X_i^*), \overline{\mathbb{I}}(X_i^*)]$

E.g. $\overline{\mathbb{I}}(X_i^*) - \underline{\mathbb{I}}(X_i^*) = 1$

Then $\underline{\overline{\mathbb{E}}}(U_i | X_i^*) = [0, 1]$

$$\underline{\overline{\mathbb{E}}}(\exp(\beta U_i) | X_i^*] = [\inf_{p \in \mathcal{M}_{X^*}} \mathbb{E}_p \exp(\beta U_i | X_i^*), \cancel{\sup_{p \in \mathcal{M}_X} \mathbb{E}_p} \sup_{p \in \mathcal{M}_{X^*}} \mathbb{E}_p \exp(\beta U_i | X_i^*)] =$$

$$= [1, \exp(\beta)]$$

Different view at the problem:

Fix a representing value X_i^* of the rounding interval, e.g. the lower interval limit

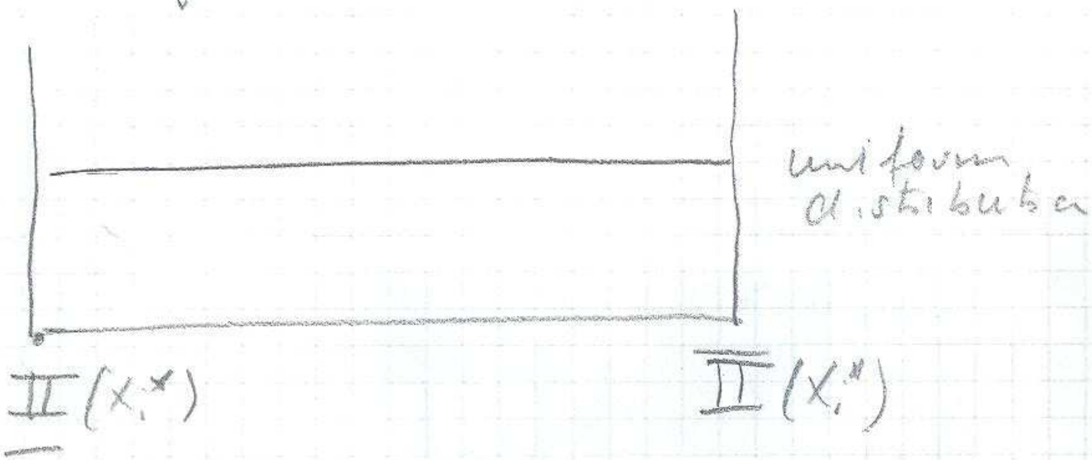
- How do we have to correct X_i^* to get an unbiased estimating equation
- Now no (or weak) knowledge on the rounding process

· Now less extreme rounding

E.g. $\mathbb{W}_i \mid X_i^* \sim \mathrm{cont}_{\varepsilon_0}\, \mathcal{U}\,[\mathbb{I}(X_i^*)]$

or

density bounds



Appendix B: Generalized Measurement Error Models

B.A. Classical Error Model with generalized Error Distrib

$$X_i^* = X_i + U_i \qquad X_i \perp U_i \quad ⚠$$

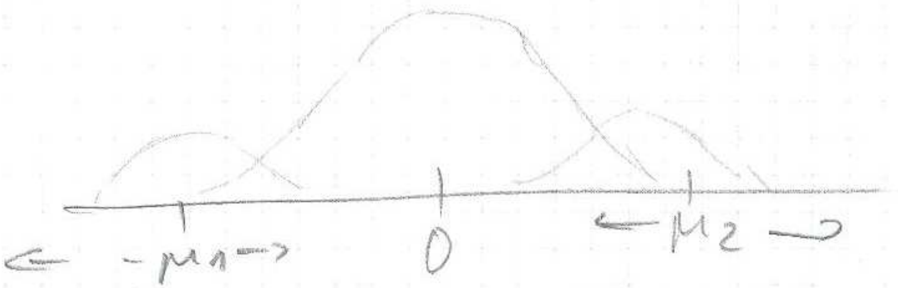
now with U_i belonging to a credal set

E.g. $U_i \sim \varepsilon_1 N(-\mu_1, \sigma_1^2) + (1-\varepsilon_1-\varepsilon_2) N(0, \sigma_0^2) + \varepsilon_2 N(\mu_2, \sigma_2^2)$

with $0 \le \varepsilon_1, \varepsilon_2 \le \varepsilon_0 \left(< \frac{1}{2}\right)$

$$\left.\begin{array}{l} 0 \le \mu_j \le \overline{\mu_j} \\ 0 \le \sigma_j \le \overline{\sigma_j} \end{array}\right\} j = 0,1,2$$

($N(\upsilon, 0)$ interpreted as Dirac measure in υ)



Detour: Anonymization in Official Statistics

Example: Linear Regression (4)

$$\Psi(Y, X, \vartheta) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) \binom{1}{X_i}$$

- Construct a (conditional) corrected score functions for the second equation; the first one can be handled similarly.
- Show that, for all $i = 1, \ldots, n$, there are constants such that

$$\mathbb{E}\big((a_{11} Y_i + a_{12} Y_i X_i^*) - (a_2 \beta_0 X_i^* + b_2) -$$
$$- (a_{31} X_i^* + a_{32} (X_i^*)^2 + b_3 \mid X_i, Y_i\big) =$$
$$= Y_i X_i - \beta_0 X_i - \beta_1 X_i^2$$

• l.h.s first term

$$\mathbb{E}(a_{11} Y_i + a_{12} Y_i X_i^* + b_1 | X_i, Y_i) =$$

$$= \mathbb{E}(a_{11} Y_i + a_{12} Y_i (X_i + U_i) + b_1 | X_i, Y_i) =$$

$$= a_{11} Y_i + a_{12} Y_i X_i + a_{12} Y_i \mathbb{E}(U_i | X_i Y_i) + b_1$$

Comparing to $Y_i X_i$ leads to

$$a_{12} = 1 , \; b_1 = 0$$

and

$$a_{11} - a_{12} \mathbb{E}(U_i | X_i Y_i) = 0 \iff a_{11} = \mathbb{E}(U_i | X_i, Y_i) =$$
$$= \mathbb{E}(U_i)$$

• l.h.s second term

$$\mathbb{E}(a_2 \beta_0 X_i^* + b_2 | X_i Y_i) = \mathbb{E}(a_2 \beta_0 (X_i + U_i) + b_2 | X_i, Y_i) =$$

$$= a_2 \beta_0 X_i + a_2 \beta_0 \mathbb{E}(U_i | X_i Y_i) + b_2$$

comparing to $\beta_0 X_i$ leads to $a_2 = 1$ $\quad b_2 = -\beta_0 \mathbb{E}(U_i | X_i Y_i) = -\beta_0 \mathbb{E}(U_i)$

Third term

$$\mathbb{E}\left(a_{31} X_i^* + a_{32} \beta_1 (X_i^*)^2 + b_3 \mid X_i Y_i \right) =$$

$$= \mathbb{E}\left(a_{31}(X_i + U_i) + a_{32}\beta_1 (X_i^2 + 2X_i U_i + U_i^2) + b_3 \mid X_i Y_i \right) =$$

$$= a_{31} X_i + a_{32}\beta_1 X_i^2 + a_{31}\mathbb{E}(U_i \mid X_i Y_i) + 2a_{32}\beta_1 X_i \mathbb{E}(U_i \mid X_i, Y_i) +$$

$$+ a_{32}\beta_1 \mathbb{E}(U_i^2 \mid X_i Y_i) + b_3$$

Comparing to $\beta_1 X_i^2$ leads to

$$a_{32} = 1$$

$$X_i \left(a_{31} + 2 a_{32} \beta_1 \mathbb{E}(U_i \mid X_i Y_i) \right) = 0$$

$$a_{31} = -2 a_{32} \beta_1 \mathbb{E}(U_i \mid X_i Y_i)$$

$$b_3 = a_{32} \beta_1 \mathbb{E}(U_i^2 \mid X_i Y_i) - a_{31} \mathbb{E}(U_i \mid X_i Y_i)$$

$$= \beta_1 \underbrace{\mathbb{E}(U_i^2 \mid X_i Y_i)}_{\mathbb{E}(U_i^2)} - 2\beta_1 \underbrace{\left(\mathbb{E}(U_i \mid X_i Y_i)\right)^2}_{(\mathbb{E}(U_i))^2}$$

B2 Beyond a Simple Independence Structure (7)

- Note that the whole derivation was done in terms of $\mathbb{E}(U_i | X_i, Y_i)$ and $\mathbb{E}(U_i^2 / X_i Y_i)$
- If $\mathbb{E}(U_i | X_i Y_i) \in [-\underline{u}_1, \overline{u}_1]$ and
  $\mathbb{E}(U_i^2 | X_i Y_i) \in [-\underline{u}_2, \overline{u}_2]$ the argumentation still applies
- Indeed, as long as $\mathbb{E}(U_i | X_i Y_i)$ is of the form $\mathbb{E}(U_i | X_i) = \alpha_0 + \alpha_1 X_i$ similar techniques can be applied