LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Julia Plass, Marco Cattaneo, Thomas Augustin,
Georg Schollmeyer, Christian Heumann

Towards a reliable categorical regression analysis for non-randomly coarsened observations: An analysis with German labour market data

Technical Report Number 206, 2017
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Towards a reliable categorical regression analysis for non-randomly coarsened observations: An analysis with German labour market data

Julia Plass
Department of Statistics, LMU Munich
julia.plass@stat.uni-muenchen.de

Marco Cattaneo
School of Mathematics and Physical Sciences
University of Hull
m.cattaneo@hull.ac.uk

Thomas Augustin
Department of Statistics, LMU Munich
augustin@stat.uni-muenchen.de

Georg Schollmeyer
Department of Statistics, LMU Munich
georg.schollmeyer@stat.uni-muenchen.de

Christian Heumann
Department of Statistics, LMU Munich
christian.heumann@stat.uni-muenchen.de

Abstract

In most surveys, one is confronted with missing or, more generally, coarse data. Many methods dealing with these data make strong, untestable assumptions, e.g. coarsening at random. But due to the potentially resulting severe bias, interest increases in approaches that only include tenable knowledge about the coarsening process, leading to imprecise, but credible results. We elaborate such cautious methods for regression analysis with a coarse categorical dependent variable and precisely observed categorical covariates. Our cautious results from the German panel study "Labour market and social security" illustrate that traditional methods may even pretend specific signs of the regression estimates.

Keywords: coarse data, (cumulative) logit model, missing data, partial identification, PASS data, (profile) likelihood

1 Introduction: How to respect the (lack of) knowledge about incompleteness

In almost all surveys the problem of item-nonresponse occurs [e.g. 19, 40]. One of the principal challenges in the statistical analysis of missing data is the impossibility to test the associated missingness mechanism without adding strong assumptions [e.g. 20]. Despite the awareness of this problem, frequently untestable assumptions on the missingness process are still included in situations where the validity of these assumptions might actually be doubtful. Examples are the missing at random assumption [introduced by 34] or approaches relying on a specific pattern-mixture or selection model [e.g. developed by 12]. In this way, point-identifiability, i.e. uniqueness of parameters, is forced, which is an important prerequisite for the applicability of traditional statistical methods, as for instance the EM algorithm or imputation techniques [e.g. 22].

Especially due to the substantial bias induced by wrongly imposing such point-identifying assumptions, a proper reflection of the available information about the underlying missingness assumption is indispensable [e.g. 23]. To this end, one departs from insisting on point-identifying assumptions by turning to strategies that only include the achievable knowledge, typically ending up in set-valued estimators. In this way, approaches based on the methodology of partial identification start with no missingness assumptions at all, but then add successively assumptions compatible with the obtainable knowledge [e.g. 23]. A practical example is given in [1], where the worst-case bounds for the HIV rate resulting from an approach without any assumptions about the missingness are then refined by exploiting the longitudinal nature of the data. Similarly, sensitivity analyses for selection models take several different models of missing data processes into account [e.g. 14, 21, 45]. Recently, **(author?)** [24] gave a new impetus to this topic by stressing the advantage of reliable, so to say interval-valued point estimates for official statistics with survey nonresponse.

Against this background, we rely on cautious likelihood-based strategies for incomplete data [similarly as in 6, 8, 21, 47], and pursue the goal of determining regression estimators reflecting the available information about the incompleteness in a careful way.[1] Motivated by the two considered examples regarding the income questions from the German panel study "Labour market and social security" [PASS, 42], the focus is set on the logit model for binary response data and the cumulative logit model for ordinal response data. In doing so, we not only restrict to the issue of nonresponse, but also look at the problem of missingness more generally: Apart from fully observed and fully unobserved values, we additionally consider partially observed values where subsets of the full sample space are observed, thus addressing the coarse data problem [e.g. 13]. Consequently, coarse data contain more information than missing data, wherefore we argue in favor of collecting coarse data in case of a preceding nonresponse. Throughout, we restrict to cases of coarse categorical response variables and precisely observed categorical covariates.

Although analysts might be aware of the consequences of traditional approaches mostly making simplified assumptions such as coarsening at random, they frequently prefer them to cautious approaches for pragmatic reasons. To face this dilemma, we provide an estimation technique that includes all available information about the coarsening in a very natural and flexible way. There are already several methods that try to exploit additional infor-

[1] While in **(author?)** [32] first considerations have already been presented for the special case of a multinomial logit model that included all interactions between the covariates, we here investigate general model specifications.

mation about the incompleteness, as e.g. knowledge about the number of failed-contact attempts in **(author?)** [46] or prior expert beliefs about the differences between responders and nonresponders in **(author?)** [19]. But since these approaches are mostly restricted to either give a precise result or no answer at all, they are incapable to make use of potential available partial knowledge about the missingness that is not sufficient to point-identify the parameters of interest [e.g. 39]. Consequently, the users might conceive the explicit allowance of partially identified parameters as an advantage, since partial knowledge no longer has to be left out of consideration. In our data example we show how partial information about the coarsening such as "respondent with a high income rather tend to give a coarse answer compared to respondents with a low income" can refine the initial results without coarsening assumptions. Furthermore, we give the opportunity to consider "coarsening at random" instead of "exact coarsening at random" models improving the credibility of classical approaches.

The relevance of such a cautious approach, and hence the need of quantifying the underlying uncertainty due to incompleteness, is also apparent from the following latest practical example: Results on the job-seeking refugees in Germany without school-leaving qualification were published by the Federal Employment Agency and provoked a heated debate, mainly reasoned by a different dealing with item-nonresponse. While ignoring the 24.7% nonresponders leads to the result that 34.3% job-seeking refugees are without school-leaving qualification and assumes the refusals to be made randomly, the newspaper "Bild" disseminates an extreme interpretation of the Federal Institute for Vocational Education and Training's (BIBB) conjecture that job-seeking refugees without school-leaving qualification rather tend to disclose their answer and simply counted all nonresponders to this group, hence speaking of 59% in this context [cf., e.g., 15, 4]. A clear communication of the underlying uncertainty would have avoided the discussions and should generally be part of every trustworthy data analysis. As a reaction to the incident several statistical agencies pointed to the importance of reflecting about the reasons why the respondents refused their answers [cf., e.g., 5]. The cautious approach presented in this paper is able to express the underlying uncertainty attributed to nonresponse and could potentially derive weak, but tenable knowledge about the coarsening from the main reasons for nonresponse.

In fact, we not only deal with the uncertainty associated to the incompleteness of the data leading to imprecise results, but also two further kinds of uncertainty: By constructing confidence intervals, we capture the uncertainty arising from the availability of a finite sample only. Studying regression models, we additionally address model uncertainty arising from the parametric assumptions implied by non-saturated regression models. The interaction between the different kinds of uncertainty will be a further aspect of investigation in this paper.

Our paper is structured as follows: In Section 2 we motivate the collection of coarse data, introduce the running example based on the PASS data, explain the way we look at the problem and briefly show the principal idea of the two methods to determine cautious regression estimates that we present and discuss in this paper. Both methods are firstly developed in context of a data example with a binary response variable reducing to the missing data problem in Section 3, where also a way to obtain respective likelihood-based confidence intervals is given. The synergy of the included parametric assumptions on the regression model and the observed data strongly determines the type of results, where three substantially differing cases are elaborated. Afterwards, the applicability of the previous major developments is discussed in the context of coarse data in the strict sense in Section 4. In Section 5 we turn to situations where we benefit from weak auxiliary

information about the coarsening. Section 6 concludes by giving a summary and some remarks on further research.

2 Coarse categorical data

In most surveys, respondents can choose between several predetermined options to answer. Nevertheless, providing answers associated to a specific level of accuracy may be considered as problematic for different reasons: Firstly, respondents might be able to give a more precise answer, but there is no possibility to express it. Secondly, the other way round, respondents potentially may at most be able to decide for a set of categories, but not for the one category they actually belong to, since they are not acquainted enough with the topic of the question. Thirdly, respondents may deliberately refuse their precise answer for reasons of data privacy. While the consequence in the first situation is (only) loss of information, in the second and third situation non-ignorable nonresponse or measurement errors occur in a classical questionnaire design. All these problems could be attenuated by asking in different ways allowing the respondent to report in the required level of accuracy. An example for such an explicit collection of coarse categorical data is given in the following section by introducing the setting of the running example.

2.1 The running data example

Since the income question is known to be highly affected by nonresponse [e.g. 41], the German Panel study "Labour market and social security" [PASS study[2], 42] intends to mitigate this problem by using the following questioning technique illustrated in Figure 1: Respondents refusing to disclose their precise income (in the following called nonresponders) are asked to answer additional questions starting from providing rather large income classes (e.g. < 1000 € or not) that are successively narrowed (e.g. < 500 €).[3] In this way, answers with different levels of coarseness are received by simultaneously ensuring the individual degree of data privacy demanded by the respective respondent. This strategic questioning technique to increase response rates is sometimes referred to as non-response follow-up [e.g. 30, where this is distinguished from "follow-up attempts", i.e. repeated efforts to contact respondents]. Depending on the research question, various ways to integrate the answers from the respondents reporting their precise, non-categorical income are conceivable, where we first point to some general options before we mention how we proceed here: To include all answers in the most precise level inferable from the data, a mixture model [e.g. 25] may be used differentiating between nonresponders and responders. In some situations, as e.g. in the context of poverty measurement, an answer on a certain ordinal level might be sufficient, hence the precise answers could be classified to the most precise income categories reported by the nonresponders, allowing a joint analysis. An alternative might be a joint likelihood approach accounting for responders, nonresponders and different groups of partial responders by distinct likelihood contributions [cf. 10, who use an imputation based technique and illustrate their results by the PASS data as well]. Restricting to the answers of the nonresponders, here we consider the most precise collectable categorical income as the true income category, ignoring that a (quasi-) continuous variable is underlying. In a second step a mixture model or a comparative analysis

[2] Here we rely on the data from wave 1 to 4

[3] For ease of presentation, we here restrict to the granularity of categories given in Figure 1. In fact, the PASS data partly provide even finer categories.

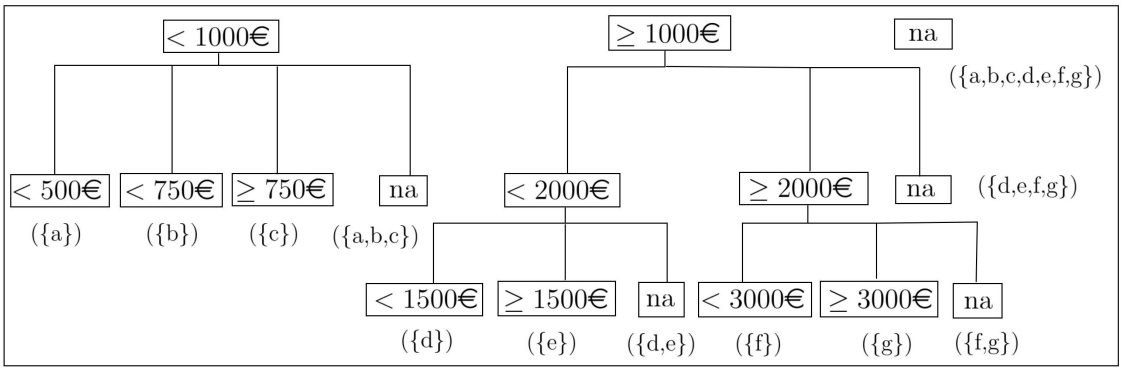

Figure 1: In the PASS study for nonresponders the income questions are individually adjusted, providing for instance categories abbreviated by "$<$ 500 €", "$<$ 750 €" (actually meaning $<$ 750 € and $\geq$ 500 €) and "$\geq$ 750 €" ($\geq$ 750 € and $<$ 1000 €) to original nonresponders who already reported to be in class $<$ 1000 € in an earlier question. The notation in brackets refers to **Example 2**, introduced later on, where the cardinality of the sets gives some indication about the level of accuracy.

to the responders could follow.
Our main goal will be the investigation of some covariates' impact on a true categorical response variable partly observed in a coarse way. In the example, the true categorical income is used as a response variable distinguishing the following two settings, referred to as "Example 1" and "Example 2" later on:

Example 1: Binary response variable
Here we restrict the available income data to the answers obtained from the first question. Thus, categories "$<$ 1000 €", "$\geq$ 1000 €" and "no answer" (i.e. coarse answer "either $<$ 1000 € or $\geq$ 1000 €") are observed, reducing the coarsening probem to the missing data problem. When we consider **Example 1**, the categories are abbreviated by "$<$", "$\geq$" and "na" in the following.

Example 2: Ordinal response variable
Here we account for the whole ordinal structure inherent in the data, and the observed income variable includes different levels of coarseness. In the context of **Example 2**, the abbreviations given in brackets in Figure 1, i.e. categories "{a}" to "{a,b,c,d,e,f,g}", are utilized, where the latter one is interpreted as "either a or b or ... or g".

In this way, we constructed one data situation with a binary and one with an ordinal true response variable (with values "$<$ 1000 €" and "$\geq$ 1000 €" and values "{a}" to "{g}", respectively) in order to exemplify the results obtained by the two considered models. In Section 3, we use **Example 1** to illustrate the respective proposals, while in Section 4 the applicability of the previous ideas for coarse data, not reducing to the missing data case, is studied by referring to **Example 2**.
We use the highest school leaving certificate (first covariate) and age (second covariate) as covariates. Both variables are dichotomized, thus showing values "Abitur no (0)" and "Abitur yes (1)"[4] as well as "$<$ 40 (0)" and "$\geq$ 40 (1)", respectively. Since the categorical income questions are only directed to respondents refusing to disclose their precise income,

[4]The "Abitur" is the general qualification for university entrance in Germany.

Table 1: Contingency table for the data of **Example 1** (Binary response variable).

		Observed income class		
Abitur	age	$<$	$\geq$	na
no (0)	<40 (0)	97	63	102
	≥ 40 (1)	69	115	131
yes (1)	<40 (0)	33	50	41
	≥ 40 (1)	38	79	59

		Observed income class											
Abitur	age	{a}	{b}	{c}	{a,b,c}	{d}	{e}	{d,e}	{f}	{g}	{d–g}	{f,g}	{a–g}
no (0)	<40 (0)	50	17	18	12	22	11	*	9	*	9	*	102
	≥ 40 (1)	24	18	21	6	23	18	6	16	9	33	10	131
yes (1)	<40 (0)	21	*	*	*	10	7	5	7	8	9	4	41
	≥ 40 (1)	20	9	*	*	*	9	*	14	20	17	10	59

a group expected to be small in a study concerning the labour market, the number of individuals included in our analysis is comparably small. The contingency tables in Table 1 and Table **??** summarize the considered unweighted data including information of 877 individuals. To comply with our data access contract and the non-disclosure regulations of the Federal Employment Agency [cf. 3], we have to prohibit any back-calculations and delete all frequencies that are ≤ 3, here marking them by "*". In each line of Table **??** the sums of the frequencies referring to the categories $\{a\},\ \{b\}, \{c\}$ and $\{a, b, c\}$ (group 1) as well as to $\{d\},\ \{e\},\ \{d, e\},\ \{f\},\ \{g\},\ \{d-g\}$ and $\{f, g\}$ (group 2) can be inferred from Table 1. For that reason, we additionally hide the next smallest entry in each group showing deleted entries; to increase possibilities of potential replacements, one further entry is marked by "*", whenever the sum of the frequencies in the deleted entries is smaller than seven. All frequencies are > 0 (cf. assumptions in Section 2.2).

2.2 The general view of the problem

To frame the problem of coarse data technically, we distinguish between an observed and a latent world.

Let $(x_{11}, \ldots,\ x_{1p},\ y_1),\ \ldots, (x_{n1}, \ldots,\ x_{np},\ y_n)$ be a sample of n independent realizations of categorical random variables $(X_1, \ldots,\ X_p,\ Y)$. Unfavorably, some values y_i are not known precisely, hence the random variable Y refers to the latent world. Instead, we only observe a sample $(x_{11}, \ldots,\ x_{1p},\ \mathfrak{y}_1), \ldots, (x_{n1}, \ldots,\ x_{np},\ \mathfrak{y}_n)$ of n independent realizations of $(X_1, \ldots,\ X_p,\ \mathcal{Y})$, where the random set $\mathcal{Y}$ [e.g. 28] belongs to the observed world. We lay a special focus on the variable Y with sample space Ω_Y and the random set $\mathcal{Y}$ with sample space $\Omega_{\mathcal{Y}} \subset \mathcal{P}(\Omega_Y)$, where we assume the empty set to be generally excluded, but all precise categories $\{y\}$ to be included. Since we aim for a regression analysis here, we are interested in the estimation of the probabilities $\pi_{\mathbf{x}y} = P(Y = y|\mathbf{X} = \mathbf{x})$, $y \in \Omega_Y$, given the – assumed to be – precise values $\mathbf{x} = (x_1, \ldots, x_p)^T \in \Omega_X$ of categorical covariates $X_1, \ldots, X_p$. The associated dependence on the covariates is described by an appropriate response function, $\pi_{\mathbf{x}y} = h(\eta_{\mathbf{x}y})$, with linear predictor $\eta_{\mathbf{x}y} = \beta_{0y} + d(\mathbf{x})^T \boldsymbol{\beta}_y$, where d fills the role of transferring the covariates into appropriate dummy-coded ones [cf., e.g. 11, p. 31]. Our main goal will be a cautious estimation of the regression coefficients β_{0y} and

$\boldsymbol{\beta}_y$ that only includes the available information about the coarsening process.
By means of the law of total probability that includes coarsening parameters $q_{\mathfrak{y}|\mathbf{x}y} = P(\mathcal{Y} = \mathfrak{y}|\mathbf{X} = \mathbf{x}, Y = y)$ with $\mathbf{x} \in \Omega_X$, $y \in \Omega_Y$ and $\mathfrak{y} \in \Omega_{\mathcal{Y}}$ (cf. Section 3.1.1), we formalize the connection between both worlds, i.e. the latent world with parameters $\pi_{\mathbf{x}y}$, $y \in \Omega_Y$, $\mathbf{x} \in \Omega_{\mathbf{X}}$ and the observed world with parameters $p_{\mathbf{x}\mathfrak{y}} = P(\mathcal{Y} = \mathfrak{y}|\mathbf{X} = \mathbf{x})$, $\mathfrak{y} \in \Omega_{\mathcal{Y}}$, $\mathbf{x} \in \Omega_X$. Apart from requiring error-freeness in the sense that the true value is contained in the coarse value, $\mathfrak{y} \ni y$, and distinct parameters [cf. 34], we mainly refrain from making assumptions about the coarsening, only discussing in Section 5 how frequently available weak knowledge about the coarsening can be included in a powerful way. Considering the contingency table framework, $n_{\mathbf{x}\mathfrak{y}}$ and $n_{\mathbf{x}}$ represent the counts within the respective cells.

2.3 Two ways of approaching the problem

In this paper, we discuss two procedures to determine cautious maximum likelihood estimators for the regression coefficients:

- **Two-step method:** We firstly estimate the bounds of the latent variable distribution $\pi_{\mathbf{x}y} = P(Y = y|\mathbf{X} = \mathbf{x})$, $y \in \Omega_Y$, $\mathbf{x} \in \Omega_{\mathbf{x}}$, from which the cautious regression estimates are determined in a second step.
- **Direct method:** We rely on the (relative) profile log-likelihood for the regression coefficients of interest, where the set of maxima gives the cautious regression estimates.

Being interested in maximum likelihood estimators of the regression coefficients, maximizing the corresponding (profile log-) likelihood, i.e. the direct method, represents the natural procedure, which is always applicable. In specific situations – which we will characterize here – a two-step method will turn out as a useful alternative. Additionally, the way through the estimation of the latent variable distribution shows to be beneficial when we study how the parametric assumption on the regression model affects the estimated coarsening parameters, since we can implicitly control for the compatibility with the observed data. Nevertheless, it is important to point out that there are situations where only the direct method is worthwhile and hence the two methods cannot be regarded as at the same level.
Both ways aim at the cautious maximum likelihood estimators for each component of the vector of regression coefficients. Consequently, we gain an impression about the magnitude of each effect when no assumptions about the coarsening are imposed, but we cannot directly infer which one-dimensional regression estimates are combinable to achieve the maximum of the likelihood.

3 Cautious estimation of regression coefficients

An important contribution of this paper consists of elaborating how the presence of parametric assumptions on the regression model – in the sense that at least one effect or interaction of the saturated model is set equal to zero – can affect the assumptions about the coarsening process. By comparing the results from a two-step method (cf. Section 2.3) for the case with and without any parametric assumptions on the regression model, interesting insights with regard to this point can be gained. For that reason, we firstly devote ourselves to the case of a saturated model that includes all interactions between

the covariates (cf. Section 3.1) and account for the uncertainty induced by parametric assumptions on the regression model only afterwards (cf. Section 3.2).
If a saturated model is chosen, a two-step method appears to be quite natural and hence we will restrict to this way here: Since there is no reduction of the parameter space and the latent variable distribution basically represents the same information as the regression estimators, we can determine the cautious regression estimators (cf. Section 3.1.2; also cf. 32, where the multinomial logit model is used in this context) by simply transforming the bounds of the latent variable distribution obtained in a first step (cf. Section 3.1.1).
Things become substantially different in the presence of parametric assumptions on the regression model, i.e. if a non-saturated model is specified. Now, due to the reduction of the parameter space a transformation as in the saturated model is no longer valid and the direct approach (cf. Section 2.3) is becoming more important. Nevertheless, basing considerations on a two-step method in some cases still may be useful and we formulate a constraint optimization problem that incorporates the bounds of the latent variable distribution (cf. Section 3.1.1).

3.1 The saturated model

3.1.1 Maximum likelihood estimation for the latent variable distribution

In order to estimate the latent variable distribution, we basically split the argumentation by completing three steps [32]: Firstly, we use the random set perspective interpreting all elements in $\Omega_{\mathcal{Y}}$ as categories of their own. Thus, in contrast to the situation in the latent world, knowledge about the "precise" values in the observed world is available, which allows to determine the maximum likelihood estimator (MLE) for the observed variable distribution $p_{\mathbf{x}\mathbf{y}}$, $\mathbf{x} \in \Omega_X, \mathbf{y} \in \Omega_{\mathcal{Y}}$ based on the $n = \sum_{\mathbf{x}\in\Omega_X} n_{\mathbf{x}}$ observations. Since for fixed covariate values $\mathbf{x} \in \Omega_X$, the cell counts $(n_{\mathbf{x}\mathbf{y}})_{\mathbf{y}\in\Omega_{\mathcal{Y}}}$ are multinomially distributed, the MLE for the observed variable distribution is uniquely obtained by the respective conditional relative frequency, i.e. $\hat{p}_{\mathbf{x}\mathbf{y}} = \frac{n_{\mathbf{x}\mathbf{y}}}{n_{\mathbf{x}}}$, $\mathbf{x} \in \Omega_X, \mathbf{y} \in \Omega_{\mathcal{Y}}$, assuming that $n_{\mathbf{x}} > 0$.
Secondly, the information from the observation model relating the latent to the observed world is included. For this purpose, a mapping $\Phi : \gamma \mapsto \vartheta$, with $\gamma = (\pi_{\mathbf{x}y}, q_{\mathbf{y}|\mathbf{x}y})_{\mathbf{x}\in\Omega_X,\mathbf{y}\in\Omega_{\mathcal{Y}},y\in\Omega_Y}$ and $\vartheta = (p_{\mathbf{x}\mathbf{y}})_{\mathbf{x}\in\Omega_X,\mathbf{y}\in\Omega_{\mathcal{Y}}}$, is defined. This mapping describes the transfer between the parametrization in terms of the components of γ and the ones of ϑ by using the theorem of total probability. Consequently, the prescription of the reparametrization is given by

$$p_{\mathbf{x}\mathbf{y}} = \sum_{y\in\mathbf{y}} \pi_{\mathbf{x}y} \cdot q_{\mathbf{y}|\mathbf{x}y} \ , \tag{1}$$

for all $\mathbf{x} \in \Omega_X$, $\mathbf{y} \in \Omega_{\mathcal{Y}}$. Since we already calculated the MLE of ϑ and may express it as a function of the parameter of interest γ, i.e. $\vartheta = \Phi(\gamma)$, by virtue of the invariance of the likelihood we can thirdly determine the MLE of γ as the inverse image of $\hat{\vartheta}$ under the function Φ. Since the mapping Φ is generally not injective, there are several $\hat{\gamma}$, all leading to the same maximum value of the log-likelihood. Thus, we obtain the set-valued estimator

$$\hat{\Gamma} = \{\hat{\gamma} \mid \Phi(\hat{\gamma}) = \hat{\vartheta}\} \tag{2}$$

by replacing the left hand side of (1) by the MLEs $\hat{p}_{\mathbf{x}\mathbf{y}}$ of the observed world, already calculated in the first step, and the right hand side by the empirical analogues of the respective parameters.

Table 2: Estimation of the parameters of the latent world (**Example 1**).

$\hat{\pi}_{\mathbf{x}<}$	$\hat{q}_{na\|\mathbf{x}<}$	$\hat{q}_{na\|\mathbf{x}\geq}$
$\hat{\pi}_{00<} \in [0.37,\ 0.76]$	$\hat{q}_{\mathrm{na}\|00<} \in [0,\ 0.51]$	$\hat{q}_{\mathrm{na}\|00\geq} \in [0,\ 0.62]$
$\hat{\pi}_{01<} \in [0.22,\ 0.63]$	$\hat{q}_{\mathrm{na}\|01<} \in [0,\ 0.66]$	$\hat{q}_{\mathrm{na}\|01\geq} \in [0,\ 0.53]$
$\hat{\pi}_{10<} \in [0.27,\ 0.60]$	$\hat{q}_{\mathrm{na}\|10<} \in [0,\ 0.55]$	$\hat{q}_{\mathrm{na}\|10\geq} \in [0,\ 0.49]$
$\hat{\pi}_{11<} \in [0.22,\ 0.55]$	$\hat{q}_{\mathrm{na}\|11<} \in [0,\ 0.61]$	$\hat{q}_{\mathrm{na}\|11\geq} \in [0,\ 0.43]$

Throughout the paper, instead of giving the set-valued estimator $\hat{\Gamma}$ in (2) itself, we illustrate it by building its one-dimensional projections. Thus, estimators for the single components of γ are obtained, here represented as

$$\hat{\pi}_{\mathbf{x}y} \in \left[\frac{n_{\mathbf{x}\{y\}}}{n_{\mathbf{x}}},\ \frac{\sum_{\mathfrak{y} \ni y} n_{\mathbf{x}\mathfrak{y}}}{n_{\mathbf{x}}} \right], \quad \hat{q}_{\mathfrak{y}|\mathbf{x}y} \in \left[0,\ \frac{n_{\mathbf{x}\mathfrak{y}}}{n_{\mathbf{x}\{y\}} + n_{\mathbf{x}\mathfrak{y}}} \right], \tag{3}$$

for all $\mathbf{x} \in \Omega_X,\ y \in \Omega_Y$ and all $\mathfrak{y} \in \Omega_{\mathcal{Y}}$ such that $\{y\} \subsetneq \mathfrak{y}$, with $n_{\mathbf{x}} > 0$ and $\frac{0}{0} := 1$. It is important to keep in mind that points in these intervals are constrained by the restrictions in (1). The result in (3) can be shown to correspond to the one obtained from cautious data completion, plugging in all potential precise sample outcomes compatible with the observations [cf. 2, §7.8].

For sake of illustration, we apply this approach to **Example 1**, where four subgroups result from splitting by the different values of the two covariates, hence we consider $\mathbf{x} \in \Omega_X = \{\text{“00”}, \text{“01”}, \text{“10”}, \text{“11”}\}$ interpreted as “age=0, Abitur=0”, “Abitur=0, age=1”, “Abitur=1, age=0” and “Abitur=1, age=1”, respectively. Using Table 1 and referring to the data of the first subgroup, one uniquely obtains

$$\hat{p}_{00<} = \frac{n_{00<}}{n_{00}} = \frac{97}{262}, \quad \hat{p}_{00\geq} = \frac{n_{00\geq}}{n_{00}} = \frac{63}{262} \quad \text{and} \quad \hat{p}_{00na} = \frac{n_{00na}}{n_{00}} = \frac{102}{262}\,,$$

with $n_{00} = n_{00<} + n_{00\geq} + n_{00na}$. There are indeed multiple $\hat{\gamma}$, i.e. estimated combinations of coarsening parameters and latent variable distributions, that are compatible with the restriction in (1) and thus lead to this estimated observed variable distribution. Different scenarios for the estimation of $\pi_{00<}$ are conceivable ranging from attributing all coarse categories “na” to “$\geq$” to including them all in category “$<$”,[5] thus obtaining (cf. (3))

$$\hat{\pi}_{00<} \in [\underline{\hat{\pi}}_{00<},\ \overline{\hat{\pi}}_{00<}] \quad \text{with} \quad \underline{\hat{\pi}}_{00<} = \frac{97}{262} \approx 0.37 \quad \text{and} \quad \overline{\hat{\pi}}_{00<} = \frac{97 + 102}{262} \approx 0.76\,.$$

The resulting estimators (i.e. the one-dimensional projections of $\hat{\Gamma}$) in **Example 1** are shown in Table 2.

[47] presented an approach based on the profile likelihood to describe statistical evidence with missing data without imposing untestable assumptions, hence allowing for an alternative way to achieve the results in (3). Compared to the global log-likelihood $l(\cdot)$ in dependence of all parameters, the profile log-likelihood is a function of the parameter of interest only and arises from the global log-likelihood by considering all other parameters as nuisance parameters [cf., e.g. 31, p. 80]. In our case, a specific parameter $\pi_{\mathbf{x}y}$ or $q_{\mathfrak{y}|\mathbf{x}y}$, $\mathbf{x} \in \Omega_X,\ y \in \Omega_Y,\ \mathfrak{y} \in \Omega_{\mathcal{Y}}$ might be of interest and the profile log-likelihood follows as

$$l(\pi_{\mathbf{x}y}) = \max_{\xi} l(\pi_{\mathbf{x}y}, \xi) \quad \text{or} \quad l(q_{\mathfrak{y}|\mathbf{x}y}) = \max_{\xi} l(q_{\mathfrak{y}|\mathbf{x}y}, \xi) \tag{4}$$

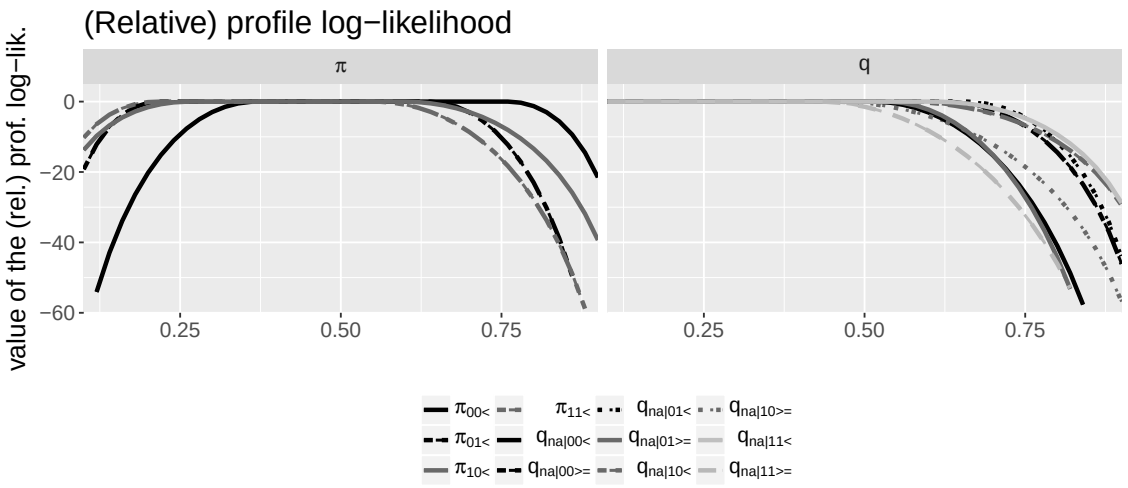

Figure 2: Referring to the the data of **Example 1**, the (relative) profile log-likelihood function for every parameter in γ is depicted.

with nuisance parameters ξ corresponding to γ without $\pi_{\mathbf{x}y}$ and $q_{\mathbf{y}|\mathbf{x}y}$, respectively. Thus, we can graphically represent the profile log-likelihood by varying the values of the parameter of interest on a grid and evaluating the log-likelihood at each fixed value for the parameter of interest and the nuisance parameters maximizing the log-likelihood in this case. Figure 2 shows the (relative) profile log-likelihood for **Example 1**, obtained by shifting the profile log-likelihood by the maximum value of the log-likelihood function along the y-axis. The range of the plateau characterizes the maximum likelihood estimator for the parameter of interest and hence is in accordance with the results in Table 2. The explicit formula for the profile log-likelihood for $\pi_{\mathbf{x}y}$ is given in [6].

3.1.2 Maximum likelihood estimators for the regression coefficients

Whenever a saturated model is used, the reparametrization in terms of the regression coefficients means no reduction of the dimension and the link function $g(\pi_{\mathbf{x}y})$ is bijective. Since it is also continuous, [cf., e.g. 11, p. 304], the bounds of the estimated regression coefficients can be calculated as a direct transformation of the bounds of the latent variable distribution (cf. Section 3.1.1).
To illustrate the procedure, we refer to the data situation of **Example 1**, where the logit model with the response function

$$\pi_{\mathbf{x}<} = P(Y = \text{“} < \text{”} \,|\mathbf{x}) = \frac{\exp(\beta_0 + d(\mathbf{x})^T\boldsymbol{\beta})}{1 + \exp(\beta_0 + d(\mathbf{x})^T\boldsymbol{\beta})} \tag{5}$$

for the category of interest, here “$<$”, and

$$\pi_{\mathbf{x}\geq} = \frac{1}{1 + \exp(\beta_0 + d(\mathbf{x})^T\boldsymbol{\beta})}, \tag{6}$$

for the reference category, here “$\geq$”, is appropriate. Equivalently, the logit model can be described by the link function

$$g(\pi_{\mathbf{x}<}) = \ln\Big(\frac{\pi_{\mathbf{x}<}}{1 - \pi_{\mathbf{x}<}}\Big) = \beta_0 + d(\mathbf{x})^T\boldsymbol{\beta}\,. \tag{7}$$

[5]This is technically related to the Dempster-Shafer Theory [cf. 36].

Table 3: Regression estimates obtained without parametric assumptions (Example 1). interactions: $\hat{\beta}_{12} \in [-2.76,\ 3.64]$ (cautious estimation), $\hat{\beta}_{12} = 0.63$ (traditional)

cautious estimation	$\hat{\beta}_0 \in [-0.53,\ 1.15]$	$\hat{\beta}_1 \in [-2.16,\ 0.92]$	$\hat{\beta}_2 \in [-2.42,\ 1.08]$
traditional procedure	$\hat{\beta}_0 = 0.43$	$\hat{\beta}_1 = -0.85$	$\hat{\beta}_2 = -0.94$

Considering a saturated model, we specify the linear predictor as $\beta_0 + \beta_1 \cdot \text{Abitur} + \beta_2 \cdot \text{age} + \beta_{12} \cdot \text{age*Abitur}$. The bounds of the four regression coefficients are then determined by transforming the bounds of the four estimators $\hat{\pi}_{00<}$, $\hat{\pi}_{01<}$, $\hat{\pi}_{10<}$ and $\hat{\pi}_{11<}$, hence obtaining

$$\hat{\beta}_0 \in \Big[\ln\Big(\frac{\underline{\hat{\pi}}_{00<}}{1-\underline{\hat{\pi}}_{00<}}\Big),\ \ln\Big(\frac{\overline{\hat{\pi}}_{00<}}{1-\overline{\hat{\pi}}_{00<}}\Big)\Big], \tag{8}$$

$$\hat{\beta}_1 \in \Big[\ln\Big(\frac{\underline{\hat{\pi}}_{10<}}{1-\underline{\hat{\pi}}_{10<}}\Big) - \overline{\hat{\beta}}_0,\ \ln\Big(\frac{\overline{\hat{\pi}}_{10<}}{1-\overline{\hat{\pi}}_{10<}}\Big) - \underline{\hat{\beta}}_0\Big]$$

$$\hat{\beta}_2 \in \Big[\ln\Big(\frac{\underline{\hat{\pi}}_{01<}}{1-\underline{\hat{\pi}}_{01<}}\Big) - \overline{\hat{\beta}}_0,\ \ln\Big(\frac{\overline{\hat{\pi}}_{01<}}{1-\overline{\hat{\pi}}_{01<}}\Big) - \underline{\hat{\beta}}_0\Big],$$

$$\hat{\beta}_{12} \in \Big[\ln\Big(\frac{\underline{\hat{\pi}}_{11<}}{1-\underline{\hat{\pi}}_{11<}}\Big) - \overline{\hat{\beta}}_1 - \overline{\hat{\beta}}_2 - \underline{\hat{\beta}}_0,\ \ln\Big(\frac{\overline{\hat{\pi}}_{11<}}{1-\overline{\hat{\pi}}_{11<}}\Big) - \underline{\hat{\beta}}_1 - \underline{\hat{\beta}}_2 - \overline{\hat{\beta}}_0\Big].$$

For **Example 1** the cautious regression estimates are given in Table 3, where they can also be compared to the results from a traditional procedure[6] assuming uninformative coarsening (in the sense of coarsening at random; more details follow in Section 5). Although the estimates from the traditional procedure are generally included in the result from the cautious estimation, they do not express the lack of knowledge about the coarsening mechanism, also pretending specific signs.

3.2 The non-saturated model

We now study non-saturated regression models, where parametric assumptions are included in the regression model in the sense that certain interactions are set equal to zero. In this way, the number of parameters that have to be estimated is reduced and the regression coefficients are generally no longer able to reproduce the latent variable distribution. We focus on the setting with the binary response variable of **Example 1**, thus choosing the response function in (5) and (6) and link function in (7) again, but now the vector of regression coefficients does not contain any interactions, i.e. $\beta_{12} = 0$. In Section 4, we discuss to which extent the obtained results can be transferred to coarse data, not reducing to the missing data problem.

In this here considered setting, we now present both methods to determine cautious regression estimators, which were already briefly announced in Section 2.3: At first, we turn to the two-step method, which allows for a direct comparison to the procedure and results of the saturated model, and hence we can investigate the impact of the parametric assumption on the regression model. Furthermore, this way gives a first insight into the type of possible situations that have to be distinguished, also including cases where the two-step method is unrewarding. For that reason, we subsequently also present the direct method. The general roles and advantages of the two methods are only then discussed in Section 4.

[6] First of all, we calculated the estimated latent variable distribution under coarsening at random [cf., e.g. 33, Equation (10)] and then transformed it via (8).

Due to the inclusion of parametric assumptions on the regression model, we can no longer rely on a bijective link function, justifying the direct transformation of the bounds of the latent variable distribution (cf. (8)). Nevertheless, a two-step procedure can still be useful, firstly estimating the latent variable distribution (cf. Section 3.1.1), thus applying (3) to e.g. obtain $\underline{\hat{\pi}}_{00<}$ and $\overline{\hat{\pi}}_{00<}$, and secondly trying to minimize/maximize the regression parameters under the condition that this estimated latent variable distribution (cf. Section 3.1.1) can be produced. This leads us to the following optimization problem, here referring to $\pi_{\mathbf{x}<} = h(\beta_0 + \beta_1 \cdot \text{Abitur} + \beta_2 \cdot \text{age})$ of **Example 1** with the response function in (5) and (6) and presented for the determination of the bounds of the effect of Abitur, i.e. $\underline{\beta}_1$ and $\overline{\beta}_1$:

$$
\begin{aligned}
&\beta_1 \to \min/\max \quad \text{given} && (9)\\
&\underline{\hat{\pi}}_{00<} \leq \frac{\exp(\beta_0)}{1+\exp(\beta_0)} \leq \overline{\hat{\pi}}_{00<}, && \underline{\hat{\pi}}_{10<} \leq \frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} \leq \overline{\hat{\pi}}_{10<},\\
&\underline{\hat{\pi}}_{01<} \leq \frac{\exp(\beta_0+\beta_2)}{1+\exp(\beta_0+\beta_2)} \leq \overline{\hat{\pi}}_{01<}, && \underline{\hat{\pi}}_{11<} \leq \frac{\exp(\beta_0+\beta_1+\beta_2)}{1+\exp(\beta_0+\beta_1+\beta_2)} \leq \overline{\hat{\pi}}_{11<}\,.
\end{aligned}
$$

In fact, in more general cases it is not sufficient to include inequalities for the bounds of the estimated latent variable distribution only. This and related consequences will be discussed in Section 4. By using the link function in (7), this optimization problem can be transformed into one with linear constraints:

$$
\begin{aligned}
&\beta_1 \to \min/\max \quad \text{given} && (10)\\
&\ln\Big(\frac{\underline{\hat{\pi}}_{00<}}{1-\underline{\hat{\pi}}_{00<}}\Big) \leq \beta_0 \leq \ln\Big(\frac{\overline{\hat{\pi}}_{00<}}{1-\overline{\hat{\pi}}_{00<}}\Big), && \ln\Big(\frac{\underline{\hat{\pi}}_{10<}}{1-\underline{\hat{\pi}}_{10<}}\Big) \leq \beta_0+\beta_1 \leq \ln\Big(\frac{\overline{\hat{\pi}}_{10<}}{1-\overline{\hat{\pi}}_{10<}}\Big),\\
&\ln\Big(\frac{\underline{\hat{\pi}}_{01<}}{1-\underline{\hat{\pi}}_{01<}}\Big) \leq \beta_0+\beta_2 \leq \ln\Big(\frac{\overline{\hat{\pi}}_{01<}}{1-\overline{\hat{\pi}}_{01<}}\Big), && \ln\Big(\frac{\underline{\hat{\pi}}_{11<}}{1-\underline{\hat{\pi}}_{11<}}\Big) \leq \beta_0+\beta_1+\beta_2 \leq \ln\Big(\frac{\overline{\hat{\pi}}_{11<}}{1-\overline{\hat{\pi}}_{11<}}\Big)\,.
\end{aligned}
$$

Considering optimization problems as in (9) or (10) with the objective function chosen as the respective regression coefficient of interest, the following types of results have to be distinguished, where $\underline{\hat{\pi}}_{\mathbf{x}y}$ and $\overline{\hat{\pi}}_{\mathbf{x}y}$ represent the estimated bounds obtained without parametric assumptions on the regression model (cf. Section 3.1.1), while $\underline{\hat{\pi}}^*_{\mathbf{x}y}$ and $\overline{\hat{\pi}}^*_{\mathbf{x}y}$ denote the bounds achievable under the parametric assumptions[7]:

1. There is a solution.
   (a) Regression estimators are obtainable that are able to produce the estimated bounds of the latent variable distribution calculated without parametric assumptions (i.e. $\hat{\pi}^*_{\mathbf{x}y} \in [\underline{\hat{\pi}}_{\mathbf{x}y},\ \overline{\hat{\pi}}_{\mathbf{x}y}]$).
   (b) The resulting regression estimators can only represent tighter bounds of the estimated latent variable distribution (i.e. $\hat{\pi}^*_{\mathbf{x}y} \in [\underline{\hat{\pi}}^*_{\mathbf{x}y},\ \overline{\hat{\pi}}^*_{\mathbf{x}y}]$ with $\underline{\hat{\pi}}^*_{\mathbf{x}y} > \underline{\hat{\pi}}_{\mathbf{x}y}$ and/or $\overline{\hat{\pi}}^*_{\mathbf{x}y} < \overline{\hat{\pi}}_{\mathbf{x}y}$), hence the inequalities are not satisfied with equality.
2. There is no solution.[8]

[7] For instance, the bounds $\underline{\hat{\pi}}^*_{10<}$ and $\overline{\hat{\pi}}^*_{10<}$ are determined by choosing $\beta_0 + \beta_1$ as objective function in the optimization problem (10). Generally, we use the superscript "*" only when we explicitly want to distinguish the respective parameter/estimator from the one without parametric assumptions on the regression model.

[8] In general, corresponding optimization problems are not solvable in the precise case either: Here, the parametric assumption on the regression model is too strong and hence prevents that the estimated response probability can be reproduced by means of the regression estimators.

Table 4: Regression estimates with parametric assumptions (Example 1).

cautious estimation	$\hat{\beta}_0 \in [-0.53,\ 1.15]$	$\hat{\beta}_1 \in [-1.84,\ 0.92]$	$\hat{\beta}_2 \in [-1.68,\ 1.08]$
traditional procedure	$\hat{\beta}_0 = 0.35$	$\hat{\beta}_1 = 0.05$	$\hat{\beta}_2 = 0.00$

By rearranging the system of inequalities in (10), we can derive the following necessary and sufficient condition for the existence of a solution of the linear optimization problem (situation 1):

$$\begin{aligned} \ln\Big(\frac{\hat{\underline{\pi}}_{11<}}{1-\hat{\underline{\pi}}_{11<}}\Big) + \ln\Big(\frac{\hat{\underline{\pi}}_{00<}}{1-\hat{\underline{\pi}}_{00<}}\Big) &\leq \ln\Big(\frac{\hat{\overline{\pi}}_{10<}}{1-\hat{\overline{\pi}}_{10<}}\Big) + \ln\Big(\frac{\hat{\overline{\pi}}_{01<}}{1-\hat{\overline{\pi}}_{01<}}\Big) \\ \ln\Big(\frac{\hat{\underline{\pi}}_{10<}}{1-\hat{\underline{\pi}}_{10<}}\Big) + \ln\Big(\frac{\hat{\underline{\pi}}_{01<}}{1-\hat{\underline{\pi}}_{01<}}\Big) &\leq \ln\Big(\frac{\hat{\overline{\pi}}_{11<}}{1-\hat{\overline{\pi}}_{11<}}\Big) + \ln\Big(\frac{\hat{\overline{\pi}}_{00<}}{1-\hat{\overline{\pi}}_{00<}}\Big) \end{aligned} \tag{11}$$

It turns out that **Example 1** is classified to situation 1, and more specifically to 1(a); the corresponding results for cautious regression estimates are given in Table 4. Again, we can conclude that results from a traditional procedure (assuming coarsening at random) have to be treated with caution: While this approach would suggest no effect of age, avoiding specific coarsening assumptions could also indicate a negative or positive effect. Comparing the results with the ones from Table 3 gives some indication about the impact of the parametric assumption on the regression model.

In the saturated model, the regression estimators are obtainable by a simple transformation (cf. Section 3.1), hence they reveal the same information as the estimated parameters determining the latent variable distribution. This is not the case in the non-saturated model, where further restrictions are included induced by the loss of flexibility from the lack of several interactions. Consequently, under parametric assumptions on the regression model tighter bounds for the regression estimators may result, but they are never wider. In this way, there is a synergy of the uncertainty associated to the coarse data problem and the one due to the parametric assumption on the regression model, which we study next for situation 1 by comparing the estimation of the coarsening parameter from the saturated model to the one from the non-saturated model.

For this purpose, we can exploit the relation between the parameters of the observed and the latent world expressed by (1). When the optimization problem in (10) is solvable (i.e. in situation 1), then the estimators of the latent variable distribution fit to the data in the sense that the estimators for the parameters of the observed world, i.e. $\hat{p}_{\mathbf{x}\mathfrak{y}}$, $\mathbf{x} \in \Omega_X,\ \mathfrak{y} \in \Omega_\mathcal{Y}$, are unaffected by the parametric assumptions and are still calculated by the (conditional) relative frequency (cf. Section 3.1.1). Since in situation 1(a) also the estimated bounds of the latent variable distribution coincide with the ones obtained without parametric assumptions, the estimated bounds of the coarsening parameters remain unchanged by the parametric assumption as well. Thus, for **Example 1**, $\hat{\underline{q}}_{\mathfrak{y}|\mathbf{x}y}$ and $\hat{\overline{q}}_{\mathfrak{y}|\mathbf{x}y}$ can still be inferred from Table 2, even if parametric assumptions are included. In situation 1(b), by applying the relation in (1) for the binary case and solving for the coarsening parameters the following estimated bounds are achievable:

$$\hat{q}_{na|\mathbf{x}<} \in \left[1 - \frac{\hat{p}_{\mathbf{x}<}}{\hat{\underline{\pi}}^*_{\mathbf{x}<}},\ \frac{\hat{\overline{\pi}}^*_{\mathbf{x}<} - \hat{p}_{\mathbf{x}<}}{\hat{\overline{\pi}}^*_{\mathbf{x}<}}\right] \quad \text{and} \quad \hat{q}_{na|\mathbf{x}\geq} \in \left[1 - \frac{\hat{p}_{\mathbf{x}\geq}}{\hat{\underline{\pi}}^*_{\mathbf{x}\geq}},\ \frac{\hat{\overline{\pi}}^*_{\mathbf{x}\geq} - \hat{p}_{\mathbf{x}\geq}}{\hat{\overline{\pi}}^*_{\mathbf{x}\geq}}\right], \tag{12}$$

with $\frac{0}{0} := 1$. Whenever $\hat{\underline{\pi}}^*_{\mathbf{x}<} = \hat{\underline{\pi}}_{\mathbf{x}<}$, then $\hat{\underline{\pi}}^*_{\mathbf{x}<} = \hat{p}_{\mathbf{x}<}$ is valid, such that the lower bound of $\hat{q}_{na|\mathbf{x}<}$ stays zero and is thus not refined (while analogous conclusions can be made for

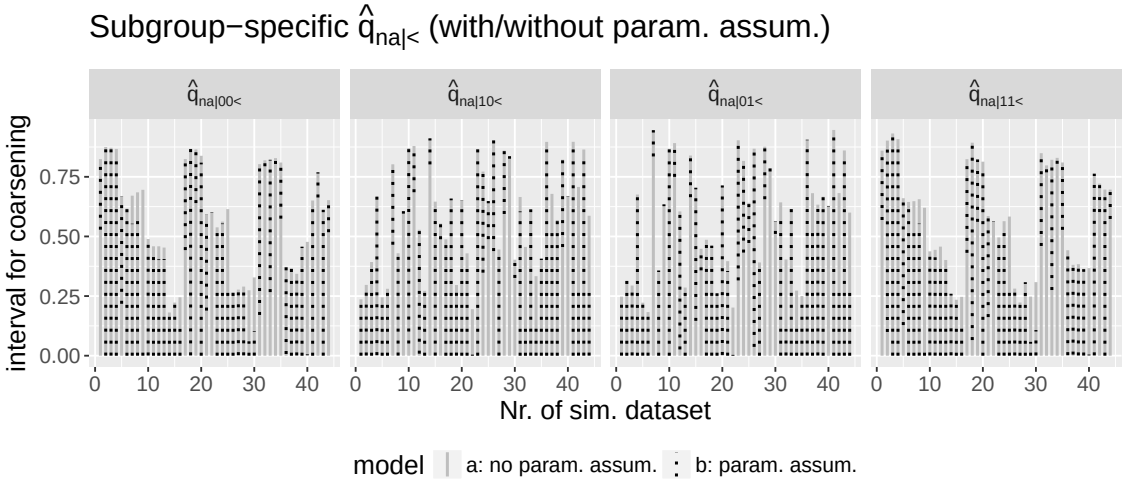

Figure 3: We restrict ourselves to data situations 1(b): In some cases the parametric assumption on the regression model induces a noticeable reduction of the coarsening intervals, while in others that are close to situation 1(a) the refinement is hardly recognizable.

the lower bound of $\hat{q}_{na|\mathbf{x}\geq}$). Due to $\underline{\hat{\pi}}^*_{\mathbf{x}y} \geq \underline{\hat{\pi}}_{\mathbf{x}y}$ and/or $\overline{\hat{\pi}}^*_{\mathbf{x}y} \leq \overline{\hat{\pi}}_{\mathbf{x}y}$, the bounds in (12) are generally not wider than those received without parametric assumptions. This is in line with the tenor in [17], who holds the view that model selection and the "disambiguation" of the incomplete data should go "hand in hand" in the sense that precise values that are consistent with the observation, but appear to be implausible under the model assumption, should no longer be under consideration. However, on the other hand from taking the model assumptions seriously several difficulties may occur, as the problem of possible ill-conditioning of the obtained set-valued estimators under such strong parametric assumptions, shortly discussed for the case of linear regression in **(author?)** [35, Section 6.1 and Appendix A therein].
We further investigate how the parametric assumption on the regression model may affect the estimated coarsening parameters in situation 1(b) by simulating different data situations, arising from assuming the same marginal distribution for the covariates as in **Example 1** and then varying the parameters of the observed variable distribution on a grid of values. Figure 3 shows the development of the intervals for the estimated coarsening $\hat{q}_{na|\mathbf{x}<}$ under the parametric assumption for those datasets that are classified into situation 1(b). As a by-product of this simulation study, we gain a first insight about the frequency of the different situations: From the 100 data sets we considered, 35 were classified into situation 1(a), 44 into situation 1(b) and 21 into situation 2. This already indicates that the number of cases where the optimization problem is not solvable is not negligible, which leads us to continue with investigating the direct approach next.

We consider the (relative) profile log-likelihood again, now not in dependence of a specific $\pi_{\mathbf{x}y}$ (cf. (4)), but of the regression coefficient of interest. The global log-likelihood $l(\beta_0, \beta_1, \beta_2, q_{na|00<}, q_{na|00\geq}, \ldots, q_{na|11\geq})$ is obtained from the one depending on the parameters determining the latent variable distribution and the coarsening parameters by replacing all parameters $\pi_{\mathbf{x}y}$ by the chosen response function. The profile log-likelihood

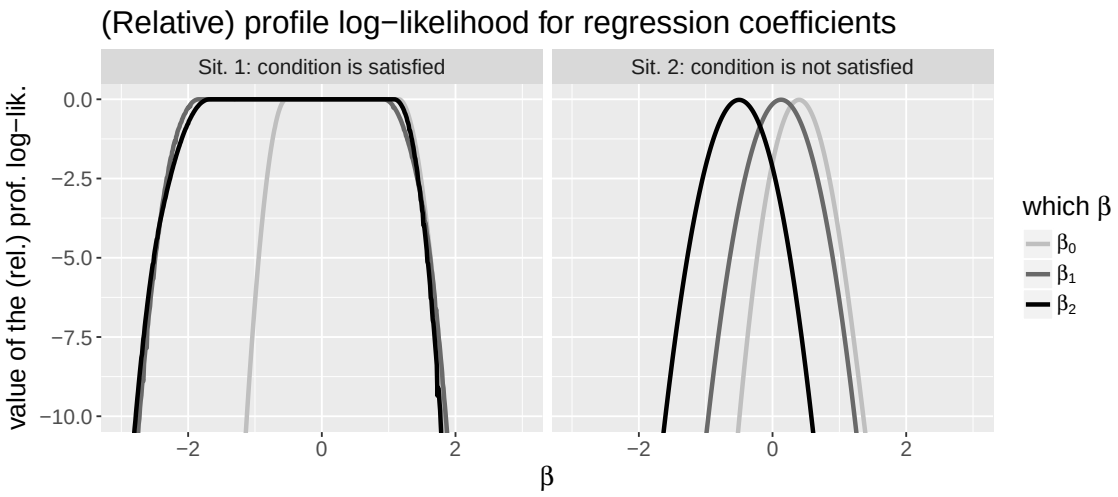

Figure 4: The left-hand part refers to the data situation of **Example 1** classified into Situation 1(a). An arbitrary data situation where the condition is not satisfied is underlying the right-hand part.

function of e.g. β_1 is then given by

$$l(\beta_1) = \max_{\xi} l(\beta_1, \xi) \, , \tag{13}$$

taking β_0, β_2, and all coarsening parameters as nuisance parameters ξ.
Figure 4 gives the (relative) profile log-likelihood for two data situations, one corresponds to the one in **Example 1** and is thus in accordance with the condition in (11), while the other is not ($n_{00<} = 60,\ n_{00\geq} = 10,\ n_{00na} = 10,\ n_{10<} = 30,\ n_{10\geq} = 40,\ n_{10na} = 5,\ n_{01<} = 20,\ n_{01\geq} = 50,\ n_{01na} = 2, n_{11<} = 40,\ n_{11\geq} = 10,\ n_{11na} = 5$). The ranges of the plateaus within the left plot corroborate the respective intervals for the regression estimators presented in Table 4.[9] It appears that precise maximum likelihood estimators are obtained, when the condition is not satisfied, while otherwise imprecision is still inherent (cf. Figure 3).
This systematic difference with regard to the nature of the result (imprecise versus precise results in situation 1 and 2, respectively) represents a particularity ascribable to the interaction of the parametric assumption on the regression model and the coarse data problem: While the parametric assumption on the regression model generally brings us into situation 2, whenever all data are precisely observed, the availability of coarse data and the associated flexibility due to the variety of possible underlying precise data scenarios can allow to "repair" the incompatibility with the observed data. This gives us the opportunity not only to assess whether the observed data fit to the model assumptions, but also to actively decide about the inclusion of additional coarsening or model assumptions, when the solvability of the optimization problem represents our claim.

3.3 Likelihood-based confidence intervals

Taking the cautious analysis seriously, the recognition of the sampling error induced by the absence of an infinite sample is crucial. There have already been several proposals to

[9] This is invisible to the naked eye, but the results from numerical optimization are quite exact.

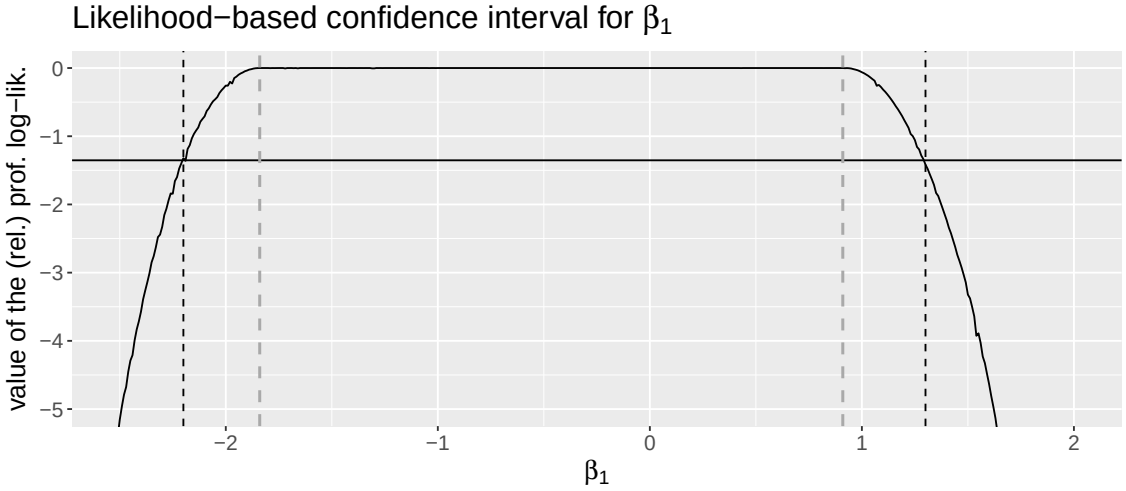

Figure 5: While the δ-cut is symbolized by the solid line, the black dashed lines mark the bounds of the confidence interval, here with $\alpha = 0.1$. The extent of the sampling uncertainty is visible by comparing these bounds with the bounds of the maximum-likelihood estimator characterized by the gray lines.

Table 5: Likelihood-based confidence intervals for the regression coefficients (**Example 1**).

for β_0 : $[-0.75,\ 1.40]$	for β_1 : $[-2.20,\ 1.29]$	for β_2 : $[-2.11,\ 1.35]$

attach value to both sources of uncertainty and confidence intervals for the latent variable distribution have been constructed [also cf. 18, 16, 37, 43]. To give confidence intervals for the regression parameters, we can tie on one of the here presented methods and either rely on a two-step method by reparametrizing the confidence intervals for the latent variable distribution via the relation formalized by the link function or base our considerations on the profile-log likelihood, where we here decide for the second option. These likelihood-based confidence intervals are appealing due to their (compared to Wald intervals) better performance in case of a small sample size [cf. e.g. 27].

Generally, likelihood-based confidence intervals are constructed by cutting the (relative) profile (log-)likelihood function at level δ with $\delta = (-0.5\chi^2_{1,1-\alpha})$ [cf., e.g. 44]. The confidence interval is then specified by regarding all parameters of interest whose value of the profile likelihood is larger than the value of δ. Likelihood-based confidence intervals in the presence of coarse data are already studied for $\pi_{\mathbf{x}y}$, $\mathbf{x} \in \Omega_X,\ y \in \Omega_Y$, relying on the profile likelihood presented in Section 3.1 [cf. 6, 47]. By referring to the (relative) profile (log-)likelihood for the regression coefficients, we can analogously proceed and define asymptotic $(1 - \alpha)$ confidence intervals by using these δ-cuts.

In Figure 5, we exemplify the construction of likelihood-based confidence intervals for the Abitur effect β_1 by using the data in **Example 1**. The result with regard to the other coefficients can be inferred from Table 5. By comparing these intervals with the ones in Table 4 an impression about the magnitude of the sampling uncertainty can be gained.

4 Studying the data application with coarse data in the strict sense (Example 2)

Since we up to now focused on a setting reducing to the missing data situation, a discussion from a more general viewpoint and an illustrative study of a situation with coarse data as present in **Example 2** is of interest. In the saturated model, the cautious regression estimators can generally be determined by a two-step procedure that gives us the cautious regression estimators by transforming the bounds of the latent variable distributions in a direct and easy way. In the non-saturated model, the preferable method (cf. Section 2.3) is not that clear. Thus, we now address the advantages and limitations of both ways, throughout turning to a non-saturated model.

To account for the ordinal structure of the response variable in **Example 2**, we base our analysis on the cumulative logit model [cf., e.g. 11, p. 334–337]. This model is based on the notion that the ordinal response categories are received due to the impossibility to collect the values of a latent continuous variable $\tilde{Y}$, thus introducing a second layer of latency. For this variable a regression model $\tilde{Y} = -d(\mathbf{x})^T\boldsymbol{\beta} + \epsilon$ with $\epsilon \sim F$ is assumed, where F is the logistic distribution function. The connection to our categorical variable of interest Y is given by $Y = y^{(l)} \Longleftrightarrow \beta_{0y^{(l-1)}} < \tilde{Y} \leq \beta_{0y^{(l)}}$, $l = 1, \ldots, m$, where $y^{(l)}$ is the lth category within the ordered categories $y^{(1)}, \ldots, y^{(l)}, \ldots, y^{(m)}$, and $-\infty = \beta_{0y^{(0)}} < \beta_{0y^{(1)}} < \cdots < \beta_{0y^{(m)}} = \infty$. In this way, the intercepts are increasing with the order of the respective category. While the intercepts are category-specific, the regression coefficients $\boldsymbol{\beta}$ are not in this model, also referred to as proportional-odds assumption. The ordinal structure is included by basing the analysis on the cumulative probabilities describing the distribution function $F(\cdot)$, hence considering the response function

$$P(Y \leq y^{(l)} \,|\mathbf{x}) = F(\beta_{0y^{(l)}} + d(\mathbf{x})^T\boldsymbol{\beta}), \quad \text{with} \tag{14}$$

$$F(\beta_{0y^{(l)}} + d(\mathbf{x})^T\boldsymbol{\beta}) = \frac{\exp(\beta_{0y^{(l)}} + d(\mathbf{x})^T\boldsymbol{\beta})}{1 + \exp(\beta_{0y^{(l)}} + d(\mathbf{x})^T\boldsymbol{\beta})}, \quad \text{and with}$$

$$\pi_{\mathbf{x}y^{(l)}} = P(Y = y^{(l)} \,|\mathbf{x}) = F(\beta_{0y^{(l)}} + d(\mathbf{x})^T\boldsymbol{\beta}) - F(\beta_{0y^{(l-1)}} + d(\mathbf{x})^T\boldsymbol{\beta}), \quad l = 1, \ldots, m,$$

[cf., e.g. 11, p. 335].

In the context of **Example 1** we already noticed that the proposed two-step method is unrewarding, whenever we are in situation 2. Now, we will additionally find that even when we are in situation 1, this procedure not necessarily simplifies the calculation as it did in **Example 1**. For given values of the covariates $\mathbf{x} \in \Omega_X$, the optimization problem considered in connection with **Example 1** only included estimated bounds for one parameter, i.e. only for $\pi_{\mathbf{x}<}$. Since a given $\hat{\pi}_{\mathbf{x}<}$, $\mathbf{x} \in \Omega_X$, refers to a specific precise scenario uniquely determining the compatible coarsening estimators $\hat{q}_{na|\mathbf{x}<}$ and $\hat{q}_{na|\mathbf{x}<}$, in situation 1 we can be sure that the cautious regression estimators obtained by the two-step method (cf. (9) and (10)) indeed maximize the respective profile log-likelihood. To fully determine the distribution in generalized probability theory, it is not sufficient to have the probability assessments on each elementary event only, but knowledge for all subsets is needed [cf., e.g., 36]. Thus, relying on the cumulative logit model, we have to include inequalities for each subset Q of Ω_Y, where the lower and upper bounds (of the confidence in Q in a given group $\mathbf{x} \in \Omega_X$) can (again) be calculated by the estimated belief and plausibility of Q, respectively. [10] While the theory behind is out of the scope of this

[10] In this way, we calculate the estimated belief of a specific Q, by including all respondents that report categories in $\mathcal{P}(\Omega_Y)$ that support Q for sure and hence are fully contained within Q, while the estimated

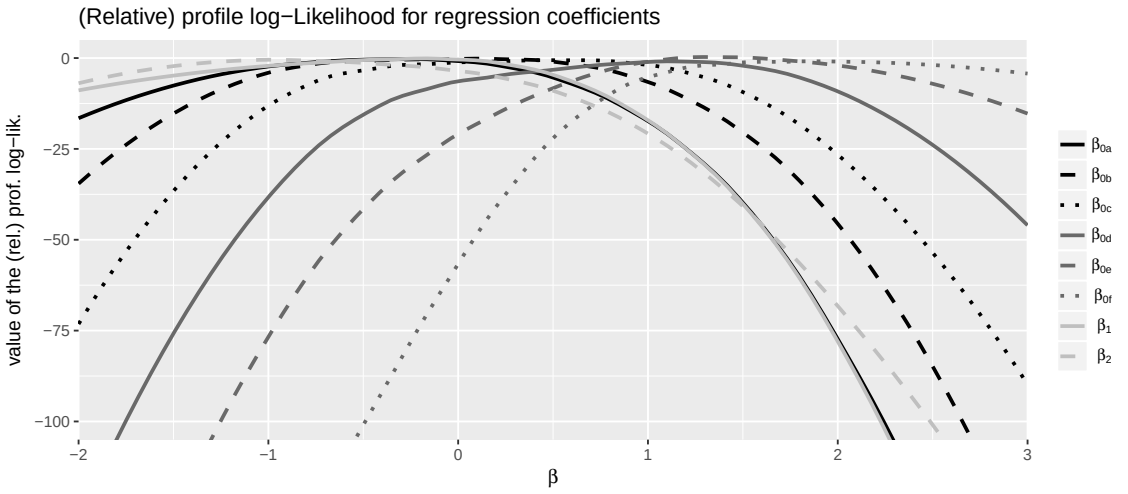

Figure 6: Relying on the data in Table **??**, for all regression coefficients the respective profile-likelihood is shown.

paper, a quick look at **Example 2**, where this leads to $2^7 \cdot 4 \cdot 2 + 5 = 1029$ inequalities[11], already clarifies that a way through the optimization problem may no longer simplify the calculation.[12] Additionally, it is not possible anymore to transform the obtained constraints, such as

$$
\begin{aligned}
\underline{\hat{\pi}}_{00c} \leq \quad & \frac{\exp(\beta_{0c})}{1+\exp(\beta_{0c})} - \frac{\exp(\beta_{0b})}{1+\exp(\beta_{0b})} - \frac{\exp(\beta_{0a})}{1+\exp(\beta_{0a})} && \leq \overline{\hat{\pi}}_{00c} \\
\underline{\hat{\pi}}_{10c} \leq \quad & \frac{\exp(\beta_{0c}+\beta_1)}{1+\exp(\beta_{0c}+\beta_1)} - \frac{\exp(\beta_{0b}+\beta_1)}{1+\exp(\beta_{0b}+\beta_1)} - \frac{\exp(\beta_{0a}+\beta_1)}{1+\exp(\beta_{0a}+\beta_1)} && \leq \overline{\hat{\pi}}_{10c} \\
\underline{\hat{\pi}}_{01c} \leq \quad & \frac{\exp(\beta_{0c}+\beta_2)}{1+\exp(\beta_{0c}+\beta_2)} - \frac{\exp(\beta_{0b}+\beta_2)}{1+\exp(\beta_{0b}+\beta_2)} - \frac{\exp(\beta_{0a}+\beta_2)}{1+\exp(\beta_{0a}+\beta_2)} && \leq \overline{\hat{\pi}}_{01c} \\
\underline{\hat{\pi}}_{11c} \leq \quad & \frac{\exp(\beta_{0c}+\beta_1+\beta_2)}{1+\exp(\beta_{0c}+\beta_1+\beta_2)} - \frac{\exp(\beta_{0b}+\beta_1+\beta_2)}{1+\exp(\beta_{0b}+\beta_1+\beta_2)} - \frac{\exp(\beta_{0a}+\beta_1+\beta_2)}{1+\exp(\beta_{0a}+\beta_1+\beta_2)} && \leq \overline{\hat{\pi}}_{11c} \,,
\end{aligned}
$$

when choosing Q to be "c" as example[13] (cf. (14)), into linear ones, further preventing a facilitation of computation.

Next, we turn to the direct method. The log-likelihood for the regression coefficients can again be written down by relying on the log-likelihood $l(\pi_{00a}, \ldots, \pi_{11f}, q_{\{abc\}|00a}, \ldots, q_{\{abcdefg\}|g})$ and replacing the latent variable distribution by the respective connection to the regression

plausibility accounts for all respondents giving answers that possibly support and thus intersect Q [cf. 36, 7]. This only extends the special case of **Example 1**, where only singletons Q were considered, but the calculation of the lower and upper bound corresponded to the estimated belief and plausibility (of query set "$<$" in the respective group $\mathbf{x} \in \Omega$), cf. Footnote 5.

[11] Cross-classifying the two covariates gives us four groups ("00", "10", "01" and "11") and hence we consider four inequalities (as in **Example 1**) for every of the 2^7 subsets of Y. Additionally we obtain inequalities for the lower and upper bounds, respectively and five further inequalities are given by $\beta_{0a} < \beta_{0b} < \cdots < \beta_{0f}$ induced by the cumulative logit model.

[12] Even if some constraints may be eliminated – theory [cf., e.g. 36] tells us that we e.g. do not need inequalities for the empty set and the ones for a set and its complement are equivalent – a high number of inequalities remains.

[13] This can be similarly written down for the other subsets Q.

coefficients, for the cumulative model given by the response function in (14). In Figure 6 the (smoothed) (relative) profile log-likelihood functions for all regression parameters are depicted, where we here refer to one possible data scenario that is compatible with the data in Table **??**.[14] From a substance matter view this is sufficient, since the results from all data scenarios closely resemble each other. The maximum likelihood estimators for the regression coefficients are again received by considering the maxima/maximum of the respective function. Due to numerical problems that occur in the optimization we can again not be sure about the kind of results, i.e. whether the optimum is indeed unique – as Figure 6 suggests – or not. Solving these computational challenges should be part of further research.

To sum up, whenever a saturated model is of interest, basing considerations on a two-step method gives us direct formulas to calculate the cautious regression estimates. Referring to non-saturated models, the (nonparametrich) latent variable distribution and the regression coefficients do not bear the same information anymore; however, we could indeed find a way to rely on a two-step method. Although the two-step method of that kind showed to be helpful to investigate the role of the parametric assumption on the regression model in the "disambiguation" of the coarse data, it should and can be applied only in particular situations: In a setting with a binary response variable (as in **Example 1**), the two-step method turned out to be very simple – in the sense that we obtain a manageable number of linear constraints. However, when we are in situation 2 (for setting of **Example 1**, we could derive a proper criterion), we have to draw on the direct method also in these simple cases. Depending on the setting and the chosen response function, the direct method may lead to technical difficulties (as already met in context of **Example 2**), here left as an open problem.

5 Incorporation of auxiliary information

Although results obtained from a cautious analysis as described in Section 3 and Section 4 at a first glance may be regarded as practically unappealing due to an unsatisfactory information content, one should generally avoid conjuring information just to force an ability to act. However, there are frequently situations where some tenable auxiliary information about the incompleteness is obtainable, refining the results in the spirit of partial identification and sensitivity analysis [e.g. 21, 23]. For the missing-data problem, literature already reveals some possibilities to incorporate (partial) knowledge, mostly by restricting either the distribution of the incompleteness or the response propensities [e.g. 24]. By formulating constraints on $q_{\mathfrak{y}|\mathbf{x}y}$, we concentrate on the first option in the context of coarse data. For this purpose, we start by considering two specific, quite strict, assumptions: Coarsening at random (CAR) and subgroup independence (SI). Afterwards, we look at generalizations to have a medium to include also other kind of knowledge, including weak knowledge about the coarsening process.

(author?) [13] introduced the concept of CAR, which requires constant coarsening probabilities $q_{\mathfrak{y}|y}$ regardless of the true underlying value y as long as it matches with the fixed observed value $\mathfrak{y}$. Adapting this assumption for our contingency table framework, the requirement has to be valid for all subgroups split by the considered covariates. An alternative type of coarsening is characterized by the independence from the corresponding covariate values. In [33] we called this assumption subgroup independence (SI) and stud-

[14] We attribute a higher selection probability to scenarios that are similar to the true one.

Table 6: Reliable regression estimates and confidence intervals under $q_{na|\mathbf{x}<} < q_{na|\mathbf{x}\geq}$ (**Example 1**).

point estimation	$\hat{\beta}_0 \in [-0.53,\ 0.35]$	$\hat{\beta}_1 \in [-0.73,\ 0.05]$	$\hat{\beta}_2 \in [-0.85,\ 0.00]$
confidence interval	for $\beta_0:\ [-0.74,\ 0.64]$	for $\beta_1:\ [-1.17,\ 0.52]$	for $\beta_2:\ [-1.35,\ 0.34]$

ied it in more detail in the setting considered there.
(author?) [29] suggests a possibility to generalize the MAR assumption by including the ratio between missing mechanisms into the analysis of non-randomly missing and misclassified data. In [32] we applied this idea by making assumptions about the coarsening ratios

$$R_{\mathbf{x},y,y',\mathfrak{y}} = \frac{q_{\mathfrak{y}|\mathbf{x}y}}{q_{\mathfrak{y}|\mathbf{x}y'}}, \quad \mathfrak{y} \in \Omega_{\mathcal{Y}},\ y,\ y' \in \mathfrak{y},\ \mathbf{x} \in \Omega_X \ , \tag{15}$$

defined for all pairs of directly successive categories y and y', where the special case of CAR is expressed by setting all these ratios equal to 1. Analogously, assumptions about the ratios

$$R_{\mathbf{x},\mathbf{x}',y,\mathfrak{y}} = \frac{q_{\mathfrak{y}|\mathbf{x}y}}{q_{\mathfrak{y}|\mathbf{x}'y}}, \quad \mathfrak{y} \in \Omega_{\mathcal{Y}},\ y \in \mathfrak{y},\ \mathbf{x}, \mathbf{x}' \in \Omega_X \ , \tag{16}$$

defined for all $\mathbf{x}$ and $\mathbf{x}'$ with two directly successive covariate values and equal other covariate values may be imposed, with $R_{\mathbf{x},\mathbf{x}',y\mathfrak{y}} = 1,\ \forall \mathbf{x},\ \mathbf{x}' \in \Omega_X,\ \mathfrak{y} \in \Omega_{\mathcal{Y}},\ y \in \Omega_Y$ representing the case of SI [cf. 33]. If all coarsening ratios in (15) were known, the parameter of interest, i.e. all parameters determining the latent variable distribution, would be point-identified, hence a particular coarsening scenario would be considered. In this way, these coarsening ratios can be regarded as sensitivity parameters in the sense of [43]. In specific cases this is also valid for the coarsening ratios in (16), studied in more detail in [33].
In most practical cases it is unrealistic to claim knowledge about the exact value of the ratios. Nevertheless, it seems quite realistic that former studies or substance-matter considerations allow rough statements about the magnitude of the ratios. In order to investigate how to include such weak knowledge about the coarsening process into the cautious estimation of the regression coefficients presented in the previous sections, we start by taking a closer look at some results under a specific partial assumption in the setting of **Example 1**, before we discuss some more general partial assumptions in context of **Example 2**.

Example 1: Frequently, there are situations, where assumptions as "respondents with a high income rather tend to give no answer compared to the ones with a low income" might be justified from an application standpoint. This weak knowledge about the missingness can be formalized as $q_{na|\mathbf{x}<} < q_{na|\mathbf{x}\geq}$ or $R_{\mathbf{x},<,\geq,na} \in [0, 1[$. Consequently, we can still rely on the consideration of the (relative) profile log-likelihood by simply adding this linear constraint on the coarsening parameters into the original optimization problem. Figure 7 shows the obtained (relative) profile likelihood functions, also indicating the δ-cut for the construction of asymptotic 90% confidence intervals. By comparing the results in Table 6, giving the estimated regression coefficients and respective confidence intervals under the auxiliary information about the missingness, to the ones without auxiliary information in Table 4 and Table 5, one notes a remarkable refinement of the results.

Example 2: Assumptions of that kind can also be included in the presence of coarse

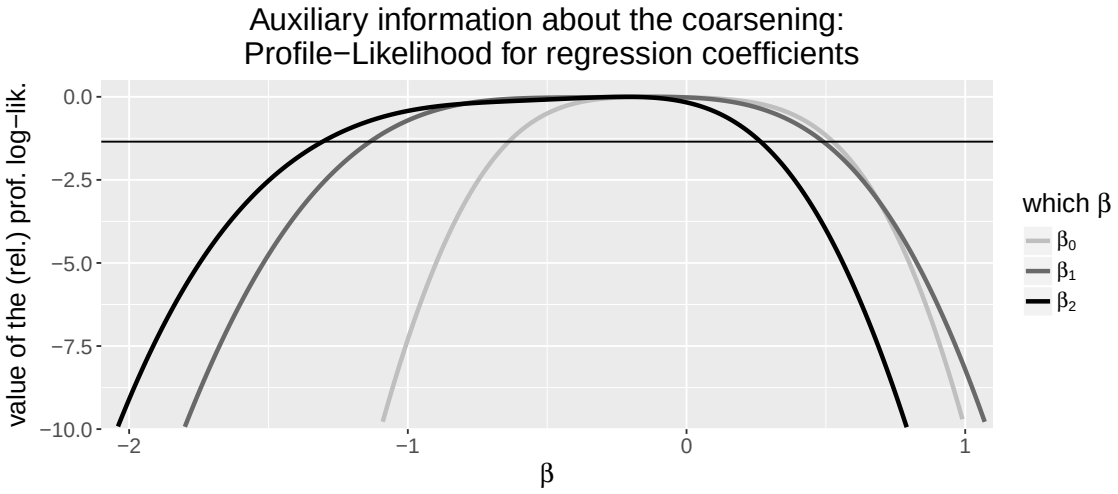

Figure 7: Based on the auxiliary information $q_{na|\mathbf{x}<} < q_{na|\mathbf{x}\geq}$ and the data of **Example 1**, the (relative) profile log-likelihood is determined. The δ-cut is marked by the horizontal line.

data in the strict sense, hence incorporating for instance $q_{\{a,b,c\}|\mathbf{x}a} < q_{\{a,b,c\}|\mathbf{x}b} < q_{\{a,b,c\}|\mathbf{x}c}$. More generally, $R_{\mathbf{x},y,y',\mathfrak{y}}$ (or analogously $R_{\mathbf{x},\mathbf{x}',y,\mathfrak{y}}$) can be assumed to be in the interval $[\underline{R},\ \overline{R}]$ with $\underline{R}, \overline{R} \in \mathbb{R}_0^+$, where one can practically incorporate this information by adding the linear constraints $q_{\mathfrak{y}|\mathbf{x}y} \geq q_{\mathfrak{y}|\mathbf{x}y'} \cdot \underline{R}$ and $q_{\mathfrak{y}|\mathbf{x}y} \leq q_{\mathfrak{y}|\mathbf{x}y'} \cdot \overline{R}$ into the optimization problem. As a special case, there are several practical situations where CAR or SI is principally conceivable, but their exact satisfaction is rather questionable. Then the inclusion of specific neighborhood assumptions [as e.g. addressed in 24, for MAR] is desirable, requiring that the coarsening probabilities lie in the environment of the CAR or SI case. This corresponds to choosing $R_{\mathbf{x},y,y',\mathfrak{y}}$ or $R_{\mathbf{x},\mathbf{x}',y,\mathfrak{y}}$ to lie within the interval $[\frac{1}{\tau_1},\ \tau_2]$, where $\tau_1,\ \tau_2 \geq 1$ specify the neighborhood. Further research should be devoted to the incorporation of auxiliary information in terms of comparable statements about the ratios (as e.g. $R_{\mathbf{x},a,b,\{a,b,c\}} \leq R_{\mathbf{x},b,c,\{a,b,c\}}$) leading to bilinear constraints and the investigation of the impact of auxiliary information under the three situations (situation 1(a), (b), 2).

6 Concluding remarks

Most reports containing survey results, also including publications in official statistics, at best point to the fact that non-sampling errors occured, but totally neglect to quantify them [cf. 24]. This practice is especially undesirable since it not only bluffs certainty leading to misinterpretation of results, but may also conduce to a substantial bias. Consequently, communication of the underlying uncertainty should be part of every trustworthy data analysis. Frequently, a considerable contribution to the non-sampling error is ascribable to the item nonresponse problem, which we tackled here by addressing the more general situation of coarse data.

We explicitly departed from the goal of forcing a particular coarsening scenario to achieve point-identified parameters. Allowing for partially identified parameters enables the user to make an analysis driven by the available information about the coarsening process, instead of – maybe unfoundedly chosen – optimization criteria or point-identifying coarsening assumptions. By generalizing the coarsening at random and the subgroup independence

assumptions, we could reveal a practical possibility how the user can include frequently available rough statements about the coarsening to refine the results obtained from an analysis based on no assumptions about the coarsening at all.
Aiming at a reliable categorical regression analysis in the presence of coarse data, two different methods to determine cautious regression coefficients have been discussed in the light of data examples: The first one is based on a two-step procedure, which turned out to simplify things only in specific situations, such as cases with a binary response variable, and is even then not always rewarding. Studying this procedure gave rise to various types of results (situation 1(a), 1(b), 2). In this way, we figured out that the parametric assumption on the regression model can induce a principally differing impact on the estimated coarsening parameters, from no effect, via tighter bounds, through to point-identified parameters. The second method, here called direct method, relies on the (relative) profile log-likelihood, where the estimated bounds of the regression coefficients are given by considering the set of all maxima. This procedure is natural, always applicable – although the computation of the (relative) profile log-likelihood may be challenging – and offers a simple way to construct confidence intervals. Having a closer look at response functions of further categorical regression models and discussing the appropriateness of both methods in this context should be part of further research.
We applied all findings to the PASS data. A comparison of the results of our cautious approach to the ones of a traditional method relying on coarsening at random showed that sometimes even certainty about the sign of the regression estimates would be pretended by the latter procedure. Depending on the research question, our results might be assessed as too little informative, especially if the confidence intervals are the focus of interest. But this does not justify to return to traditional methods, which here would pretend certainty about even the sign of the regression coefficients in some cases. Thus, a possibly small content of information should not be regarded as a weakness of an approach based on the methodology of partial identification, but associated to sparse additional knowledge. Although the gain of information achieved by the explicit collection of coarse data is comparably small in our case, which is ascribable to the low proportion of coarse compared to missing answers, the used questionnaire design for requesting the income of the PASS study is recommendable, especially for sensitive topics.
The cautious likelihood approach for the latent variable distribution turns out to be a fruitful field of study for further research: The connection between the latent and the observed world gives the opportunity to transfer already existing likelihood-based methods for precise categorical data [as e.g. statistical tests, as e.g. in 33] to the setting of coarse data. Another promising topic is the application of our cautious approach to other problems relying on strong assumptions. A direct reference is conjectured for misclassification, propensity score matching and statistical matching, where starting points are already provided in [26], [38] (who studied an approach based on partial identification to estimate treatment effects without considering propensity scores), [9], respectively. Propensity score matching and statistical matching traditionally rely on strict assumptions, namely the strongly ignorable treatment assignment as well as the conditional independence assumption, respectively, where a cautious strategy would allow for a relaxation of these prerequisites.

acknowledgements

We are grateful to the Research Data Center at the Institute for Employment Research, Nuremberg, especially Mark Trappmann and Anja Wurdack, for the access to the PASS data and their support in practical matters. Furthermore, we appreciate the help of Paul Fink, who discussed data disclosure issues with us. The first author thanks the LMUMentoring program, providing financial support for young, female researchers.

References

[1] B. Arpino, E. De Cao, and F. Peracchi. Using panel data for partial identification of human immunodeficiency virus prevalence when infection status is missing not at random. *J. R. Statist. Soc. A*, 177:587–606, 2014.

[2] T. Augustin, G. Walter, and F. Coolen. Statistical inference. In T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 135–189. Wiley, 2014.

[3] O. Beyer, A. Hüser, K. Rudloff, and M. Rüst. Statistische Geheimhaltung: Rechtliche Grundlagen und fachliche Regelungen der Statistik der Bundesagentur für Arbeit, 2014. Accessed: 2017-10-13.

[4] G. Brack. Wie viele Flüchtlinge sind ohne Schulabschluss?, 2017. Accessed: 2017-10-13.

[5] H. Brücker and J. Schupp. Annähernd zwei Drittel der Geflüchteten haben einen Schulabschluss, 2017. Accessed: 2017-10-13.

[6] M. Cattaneo and A. Wiencierz. Likelihood-based Imprecise Regression. *International Journal of Approximate Reasoning*, 53:1137–1154, 2012.

[7] I. Couso, D. Dubois, and L. Sánchez. *Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables*. Springer, 2014.

[8] T. Denœux. Likelihood-based belief function: justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55:1535–1547, 2014.

[9] M. D'Orazio, M. Di Zio, and M. Scanu. Statistical matching for categorical data: Displaying uncertainty and using logical constraints. *Journal of Official Statistics*, 22:137–157, 2006.

[10] J. Drechsler, H. Kiesl, and M. Speidel. MI double feature: Multiple imputation to address nonresponse and rounding errors in income questions. *Austrian Journal of Statistics*, 44:59–71, 2015.

[11] L. Fahrmeir, T. Kneib, S. Lang, and B. Marx. *Regression: Models, Methods and Applications*. Springer, 2013.

[12] J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5:475–492, 1976.

[13] D. Heitjan and D. Rubin. Ignorability and coarse data. *Ann. Statist.*, 19:2244–2253, 1991.

[14] C. Heumann. *Monte Carlo Methods for Missing Data in Generalized Linear and Generalized Linear Mixed Models*. 2004. Habilitation (post-doctoral thesis). Ludwig-Maximilians Universität, München.

[15] D. Hoeren. 59 Prozent der Flüechtlinge haben keinen Schulabschluss, 2017. Accessed: 2017-10-13.

[16] J. Horowitz and C. Manski. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Am. Statist. Ass.*, 95:77–84, 2000.

[17] E. Hüllermeier. Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization. *International Journal of Aproximate Reasoning*, 55:1519–1534, 2014.

[18] G. Imbens and C. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

[19] D. Jackson, I. White, and M. Leese. How much can we learn about missing data?: an exploration of a clinical trial in psychiatry. *J. R. Statist. Soc. A*, 173:593–612, 2010.

[20] M. Jaeger. On testing the missing at random assumption. In J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, editors, *ECML '06, Proceedings of the 17th European Conference on Machine Learning*, Lecture Notes in Artificial Intelligence, pages 671–678. Springer, 2006.

[21] M. Kenward, E. Goetghebeur, and G. Molenberghs. Sensitivity analysis for incomplete categorical data. *Statistical Modelling*, 1:31–48, 2001.

[22] R. Little and D. Rubin. *Statistical Analysis with Missing Data*. 2nd edition, Wiley, 2014.

[23] C. Manski. *Partial Identification of Probability Distributions*. Springer, 2003.

[24] C. Manski. Credible interval estimates for official statistics with survey nonresponse. *Journal of Econometrics*, 191:293–301, 2015.

[25] G. McLachlan and D. Peel. *Finite mixture models*. Wiley, 2004.

[26] F. Molinari. Partial identification of probability distributions with misclassified data. *Journal of Econometrics*, 144:81–117, 2008.

[27] M. Neale and M. Miller. The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*, 27:113–120, 1997.

[28] H. Nguyen. *An Introduction to Random Sets*. CRC, 2006.

[29] E. Nordheim. Inference from nonrandomly missing categorical data: An example from a genetic study on Turner's syndrome. *J. Am. Statist. Ass.*, 79:772–780, 1984.

[30] K. Olson. Do non-response follow-ups improve or reduce data quality?: a review of the existing literature. *J. R. Statist. Soc. A*, 176:129–145, 2013.

[31] Y. Pawitan. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, 2001.

[32] J. Plass, T. Augustin, M. Cattaneo, and G. Schollmeyer. Statistical modelling under epistemic data imprecision: Some results on estimating multinomial distributions and logistic regression for coarse categorical data. In T. Augustin, S. Doria, E. Miranda, and E. Quaeghebeur, editors, *ISIPTA '15, Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 247–256. SIPTA, 2015.

[33] J. Plass, M. Cattaneo, G. Schollmeyer, and T. Augustin. On the testability of coarsening assumptions: A hypothesis test for subgroup independence. *International Journal of Approximate Reasoning*, 90:292–306, 2017.

[34] D. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976.

[35] G. Schollmeyer and T. Augustin. Statistical modeling under partial identification: Distinguishing three types of identification regions in regression analysis with interval data. *International Journal of Approximate Reasoning*, 56:224–248, 2015.

[36] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[37] J. Stoye. More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315, 2009.

[38] J. Stoye. Partial identification and robust treatment choice: an application to young offenders. *Journal of Statistical Theory and Practice*, 3:239–254, 2009.

[39] E. Tamer. Partial identification in econometrics. *Annual Review Economics*, 2:167–195, 2010.

[40] G. Tanna. Missing data analysis in practice t. Raghunathan, 2016 Boca Raton, Chapman and Hall–CRC 230 pp.,£ 52.99 ISBN 978-1-482-21192-4. *J. R. Statist. Soc. A*, 180:684–685, 2017.

[41] R. Tourangeau and T. Yan. Sensitive questions in surveys. *Psychological Bulletin*, 133:859–883, 2007.

[42] M. Trappmann, S. Gundert, C. Wenzig, and D. Gebhardt. PASS: A household panel survey for research on unemployment and poverty. *Schmollers Jahrbuch*, 130:609–623, 2010.

[43] S. Vansteelandt, E. Goetghebeur, M. Kenward, and G. Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16:953–979, 2006.

[44] D. Venzon and S. Moolgavkar. A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 37:87–94, 1988.

[45] G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, 2009.

[46] A. Wood, I. White, and M. Hotopf. Using number of failed contact attempts to adjust for non-ignorable non-response. *J. R. Statist. Soc. A*, 169(3):525–542, 2006.

[47] Z. Zhang. Profile likelihood and incomplete data. *International Statistical Review*, 78:102–116, 2010.