Regression with Imprecise Data: A Robust Approach

Marco E. G. V. Cattaneo
Department of Statistics, LMU Munich
cattaneo@stat.uni-muenchen.de

Andrea Wiencierz
Department of Statistics, LMU Munich
andrea.wiencierz@stat.uni-muenchen.de

Abstract

We introduce a robust regression method for imprecise data, and apply it to social survey data. Our method combines nonparametric likelihood inference with imprecise probability, so that only very weak assumptions are needed and different kinds of uncertainty can be taken into account. The proposed regression method is based on interval dominance: interval estimates of quantiles of the error distribution are used to identify plausible descriptions of the relationship of interest. In the application to social survey data, the resulting set of plausible descriptions is relatively large, reflecting the amount of uncertainty inherent in the analyzed data set.

Keywords. Robust regression, imprecise data, nonparametric statistics, likelihood inference, imprecise probability distributions, survey data, informative coarsening, complex uncertainty, interval dominance, identification regions.

1 Introduction

Data are often available only with limited precision. However, only few general methods for analyzing the relationships between imprecisely observed variables have been proposed so far. These approaches seem to fall in two categories. One of them consists of approaches suggesting to apply standard regression methods to all possible precise data compatible with the observations, and to consider the range of outcomes as the imprecise result: see for example [8]. The approaches in the second category consist in representing the imprecise observations by few precise values (for example, intervals by center and width), and in applying standard regression methods to those values: see for instance [7].

In the present paper, we follow another line of approach and suggest a new regression method directly applicable to the imprecise data. This method combines likelihood inference with imprecise probability. It allows to take into account different kinds of uncertainty, that are also reflected in the imprecise results of the regression. The suggested method imposes only very weak assumptions and yields extremely robust results. In particular, it is nonparametric, in the sense that no assumptions about the error distribution are necessary, in contrast, for instance, to the approach of [20]. We describe the regression method in Section 3, which is based on the general methodology for inference with imprecise data introduced in Section 2.

In addition to the theoretical results, in Section 4 we apply the method to analyze an interesting question in the social sciences. We investigate the relationship between age and income on the basis of survey data. The source of data used in this paper is "Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) — German General Social Survey" of 2008. The data is provided by GESIS — Leibniz Institute for the Social Sciences.

2 Imprecise Data

Let $V_1, \dots, V_n$ be n random objects taking values in a set $\mathcal{V}$, and let $V_1^*, \dots, V_n^*$ be n random sets taking values in a set $\mathcal{V}^* \subseteq 2^{\mathcal{V}}$, such that the events $V_i \in V_i^*$ are measurable. We are actually interested in the data V_i, but we can only observe the imprecise data V_i^*. The connection between precise and imprecise data is established by the following assumptions about the probability measures considered as models of the situation.

For each $\varepsilon \in [0,1]$, let $\mathcal{P}_\varepsilon$ be the set of all probability measures[1] P such that the n random objects $(V_1, V_1^*), \dots, (V_n, V_n^*)$ are independent and identically distributed and satisfy

$$P(V_i \in V_i^*) \geq 1 - \varepsilon. \tag{1}$$

[1] Probability measures and random objects are defined on an underlying measurable space.

We assume that the precise and imprecise data can be modeled by a probability measure P included in a particular set $\mathcal{P} \subseteq \mathcal{P}_\varepsilon$, for some $\varepsilon \in [0,1]$. Each $P \in \mathcal{P}$ can be identified with a particular joint distribution for V_i and V_i^* (that is, the precise and imprecise data, respectively) satisfying condition (1). In particular, $\mathcal{P} = \mathcal{P}_\varepsilon$ corresponds to the fully nonparametric assumption that any joint distribution for V_i and V_i^* satisfying condition (1) is a possible model of the situation (this is the assumption we consider in Sections 3 and 4). The usual choice for the value of ε is 0 (see for example [6, 17]), which corresponds to an assumption of correctness of the imprecise data: $V_i^* = A$ implies $V_i \in A$ (a.s.). However, this assumption is often too strong: some imprecise data can be incorrect, in the sense that $V_i^* = A$, but $V_i \notin A$. This is for example the case, when the imprecise data represent the classification of the precise data into categories, and some observations are misclassified. By choosing a positive value for ε, we allow each imprecise observation to be incorrect with probability at most ε.

The set $\mathcal{V}^*$ describes which imprecise data $V_i^* = A$ are considered as possible. As extreme cases we have the actually precise data (when A is a singleton) and the missing data (when $A = \mathcal{V}$). In general, the fully nonparametric assumption $\mathcal{P} = \mathcal{P}_\varepsilon$ does not exclude informative coarsening (see for example [23]): parametric models or uninformative coarsening can be imposed by a stronger assumption $\mathcal{P} \subset \mathcal{P}_\varepsilon$. However, it is important to note that the set $\mathcal{P}_\varepsilon$ depends strongly on the choice of $\mathcal{V}^*$. For example, when $\varepsilon = 0$, the choice of a set $\mathcal{V}^*$ such that its elements build a partition of $\mathcal{V}$ implies the assumption that the coarsening is deterministic and uninformative, because each possible precise data value is contained in exactly one possible imprecise observation $A \in \mathcal{V}^*$.

2.1 Complex Uncertainty

In general, we are uncertain about which of the probability measures in $\mathcal{P}$ is the best model of the reality under consideration. Our uncertainty is composed of two parts. On the one hand, we are uncertain about the distribution of the imprecise data V_i^*: this uncertainty decreases when we observe more and more (imprecise) data. On the other hand, even if we (asymptotically) knew the distribution of the imprecise data V_i^*, we would still be uncertain about the distribution of the (unobserved) precise data V_i: this uncertainty is unavoidable. To formulate this mathematically, let P_V and P_{V^*} be the marginal distributions of V_i and V_i^*, respectively, corresponding to the probability measure $P \in \mathcal{P}$. There is uncertainty about P_{V^*} in the set[2] $\mathcal{P}_{V^*} := \{P'_{V^*} : P' \in \mathcal{P}\}$, but even if P_{V^*} were known, there would still be an unavoidable uncertainty about P_V in the set

$$[P_{V^*}] := \{P'_V : P' \in \mathcal{P},\, P'_{V^*} = P_{V^*}\}.$$

The sets $[P_{V^*}]$ with $P_{V^*} \in \mathcal{P}_{V^*}$ are the identification regions for P_V in the terminology of [12]. Each of them consists of all the distributions for the precise data V_i compatible with a particular distribution for the imprecise data V_i^*. Hence, each set $[P_{V^*}]$ can be interpreted as an imprecise probability distribution on $\mathcal{V}$. By observing the realizations of the imprecise data V_i^*, we learn something about which of the imprecise probability distributions $[P_{V^*}]$ is the best model for the (unobserved) precise data V_i.

Example 1 *Let $\mathcal{V} = \{0,1\}$ and $\mathcal{V}^* = 2^{\{0,1\}}$, and assume $\mathcal{P} = \mathcal{P}_\varepsilon$ for some $\varepsilon \in [0,1]$. Then $\mathcal{P}_{V^*}$ is the set of all probability distributions on $2^{\{0,1\}}$ such that the probability of $\varnothing$ is at most ε. For each $P_{V^*} \in \mathcal{P}_{V^*}$, the identification region $[P_{V^*}]$ is the set of all probability distributions on $\{0,1\}$ such that the probability of 1 lies in the interval $\left[\underline{P}_{V^*}\{1\}, \overline{P}_{V^*}\{1\}\right]$, with*

$$\underline{P}_{V^*}\{1\} = \max\left(P_{V^*}\left\{\{1\}, \varnothing\right\} - \varepsilon,\, 0\right)$$
$$\overline{P}_{V^*}\{1\} = \min\left(P_{V^*}\left\{\{1\}, \{0,1\}\right\} + \varepsilon,\, 1\right).$$

In particular, when $\varepsilon = 0$, the imprecise probability distribution $[P_{V^*}]$ corresponds to the belief function on $\{0,1\}$ with basic probability assignment P_{V^*} (see for example [16]), in the sense that $[P_{V^*}]$ is the set of all probability distributions on $\{0,1\}$ dominating that belief function.

2.2 Likelihood

The likelihood function is a central concept in statistical inference. For parametric probability models, it is usually expressed as a function of the parameters: here we consider the more general formulation (as a function of the probability measures), which is applicable also to nonparametric models (see for example [14]). The observed (imprecise) data $V_1^* = A_1, \ldots, V_n^* = A_n$ induce the (normalized) likelihood function $lik : \mathcal{P} \to [0,1]$ defined by

$$lik(P) = \frac{P(V_1^* = A_1, \ldots, V_n^* = A_n)}{\sup_{P' \in \mathcal{P}} P'(V_1^* = A_1, \ldots, V_n^* = A_n)} = \frac{\prod_{i=1}^n P_{V^*}\{A_i\}}{\sup_{P' \in \mathcal{P}} \prod_{i=1}^n P'_{V^*}\{A_i\}}$$

for all $P \in \mathcal{P}$. The likelihood function describes the relative ability of the probability measures P in predicting the observed (imprecise) data. Therefore, the value $lik(P)$ depends only on the marginal distribution P_{V^*} of the imprecise data V_i^*. The likelihood

[2] The symbol := denotes "is defined to be".

function can be interpreted as the second level of a hierarchical model for imprecise probabilities, with $\mathcal{P}$ as first level (see for example [4, 5]). In particular, for any $\beta \in (0,1)$, the likelihood function can be used to reduce $\mathcal{P}$ to the set

$$\mathcal{P}_{>\beta} := \{P \in \mathcal{P} : lik(P) > \beta\}$$

of all the probability measures that were sufficiently good in predicting the observed (imprecise) data.

Let g be a multivalued mapping[3] from $\mathcal{P}$ to a set $\mathcal{G}$, describing a particular characteristic (in which we are interested) of the models considered. For example, g can be the multivalued mapping from $\mathcal{P}$ to $\mathbb{R}$ assigning to each probability measure P the p-quantile of the distribution of $h(V_i)$ under P, for some $p \in (0,1)$ and some measurable function $h : \mathcal{V} \to \mathbb{R}$. This is the kind of mapping g we consider in Sections 3 and 4: it is multivalued, because in general quantiles are not uniquely defined[4]. For each $\beta \in (0,1)$, the set

$$\mathcal{G}_{>\beta} := \bigcup_{P \in \mathcal{P}_{>\beta}} g(P)$$

is called likelihood-based confidence region with cutoff point β for the values of the multivalued mapping g. This confidence region consists of all values that the characteristic described by g takes on the set $\mathcal{P}_{>\beta}$ of all the probability measures that were sufficiently good in predicting the observed (imprecise) data.

The unique function $lik_g : \mathcal{G} \to [0,1]$ describing these confidence regions, in the sense that

$$\mathcal{G}_{>\beta} = \{\gamma \in \mathcal{G} : lik_g(\gamma) > \beta\}$$

for all $\beta \in (0,1)$, is called (normalized) profile likelihood function induced by the multivalued mapping g. It can be easily checked that[5] for all $\gamma \in \mathcal{G}$,

$$lik_g(\gamma) = \sup_{P \in \mathcal{P} \,:\, \gamma \in g(P)} lik(P).$$

Example 2 *In the situation of Example 1, let $\varepsilon = 0$, and consider the mapping[6] g from $\mathcal{P}$ to $[0,1]$ assigning to each probability measure P the probability $P_V\{1\}$ that a precise data value V_i is* 1 *(before observing the corresponding imprecise data value V_i^*). The induced profile likelihood function[7] lik_g on $[0,1]$ is plotted in Figure 1 for the cases in which the imprecise data $\{0\}$, $\{1\}$, and $\{0,1\}$ have been observed* 11, 21, *and* 6 *times, respectively (solid line), and* 213, 651, *and* 98 *times, respectively (dashed line).*

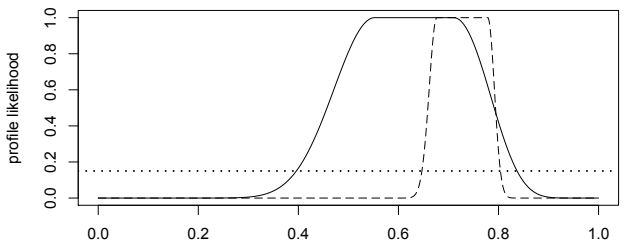

Figure 1: Profile likelihood functions from Examples 2 and 3.

In these two cases, the likelihood-based confidence regions with cutoff point $\beta = 0.15$ for the probability $P_V\{1\}$ are approximately the intervals $[0.39, 0.84]$ *and* $[0.65, 0.80]$*, respectively (the cutoff point $\beta = 0.15$ is represented by the dotted line in Figure 1). They are (conservative) confidence intervals of approximate level* 95% *(see for example [11]).*

2.3 Likelihood for Imprecise Data Models

In the situation we consider, we are actually interested in the (unobserved) precise data V_i. In this case, the characteristic of interest (described by g) depends only on the marginal distribution P_V of the precise data V_i; that is, we can write $g(P) =: g'(P_V)$ for all $P \in \mathcal{P}$. For example, the p-quantile of the distribution of $h(V_i)$ depends only on the distribution of V_i. By contrast, as noted at the beginning of Subsection 2.2, the value $lik(P)$ depends only on the marginal distribution P_{V^*} of the imprecise data V_i^*. By writing $lik(P) = lik^*(P_{V^*})$ for all $P \in \mathcal{P}$, we define a function $lik^* : \mathcal{P}_{V^*} \to [0,1]$, which can be interpreted as the likelihood function on $\mathcal{P}_{V^*}$.

In order to obtain the profile likelihood function lik_g, it can be useful to consider the multivalued mapping g^* from $\mathcal{P}_{V^*}$ to $\mathcal{G}$ defined by

$$g^*(P_{V^*}) = \bigcup_{P_V \in [P_{V^*}]} g'(P_V)$$

for all $P_{V^*} \in \mathcal{P}_{V^*}$. The multivalued mapping g^* assigns to each P_{V^*} all the values that the characteristic described by g' takes on the set $[P_{V^*}]$ of all distributions for the precise data V_i compatible with the distribution P_{V^*} for the imprecise data V_i^*. That is, g^* can be interpreted as an imprecise version of g', assigning to each imprecise probability distribution $[P_{V^*}]$ the corresponding imprecise value of g'.

The multivalued mapping g^* can be useful to obtain the profile likelihood function lik_g because, as can be

[3] Mathematically, $g : \mathcal{P} \to 2^{\mathcal{G}} \setminus \{\varnothing\}$, but g is interpreted as an "imprecise" mapping from $\mathcal{P}$ to $\mathcal{G}$.

[4] A p-quantile of the distribution of $h(V_i)$ is any value $q \in \mathbb{R}$ such that $P\left(h(V_i) < q\right) \leq p \leq P\left(h(V_i) \leq q\right)$.

[5] In this paper, $\sup \varnothing = 0$.

[6] As a multivalued mapping, g is defined by $g(P) = \{P_V\{1\}\}$ for all $P \in \mathcal{P}$.

[7] The details of the calculation of lik_g are not of primary interest at this point.

easily checked,

$$lik_g(\gamma) = \sup_{P_{V^*} \in \mathcal{P}_{V^*} : \gamma \in g^*(P_{V^*})} lik^*(P_{V^*})$$

for all $\gamma \in \mathcal{G}$. The right-hand side of this expression can be interpreted as the value $lik^*_{g^*}(\gamma)$ of the profile likelihood function $lik^*_{g^*}$ induced by the multivalued mapping g^*, when lik^* is considered as the likelihood function on $\mathcal{P}_{V^*}$.

The profile likelihood function $lik^*_{g^*}$ is particularly interesting, because lik^* describes the uncertainty about the distribution P_{V^*} of the imprecise data V_i^*, which decreases when we observe more and more (imprecise) data, while g^* describes the unavoidable uncertainty about the values of the multivalued mapping g'. In the terminology of [12], the values of g^* are the identification regions for the values of the multivalued mapping g.

Example 3 *The imprecise version g^* of the mapping g of Example 2 is the multivalued mapping from $\mathcal{P}_{V^*}$ to $[0,1]$ assigning to each P_{V^*} the interval*

$$\left[\underline{P}_{V^*}\{1\},\, \overline{P}_{V^*}\{1\}\right] = \left[P_{V^*}\left\{\{1\}\right\},\, P_{V^*}\left\{\{1\},\{0,1\}\right\}\right].$$

That is, $g^*(P_{V^*})$ is the interval probability that a precise data value V_i is 1 (before observing the corresponding imprecise data value V_i^*) according to the imprecise probability distribution $[P_{V^*}]$ (i.e., the belief function on $\{0,1\}$ with basic probability assignment P_{V^*}).

The profile likelihood function $lik_g = lik^*_{g^*}$ on $[0,1]$ is plotted in Figure 1 for the two cases considered in Example 2. In the case with 38 data (solid line) there is uncertainty also about the distribution P_{V^*} of the imprecise data V_i^*, while in the case with 962 data (dashed line) almost only the unavoidable uncertainty described by g^* remains, in the sense that $lik^*_{g^*}$ is almost equal to the indicator function of an identification region for $P_V\{1\}$ (i.e., of a probability interval $\left[\underline{P}_{V^*}\{1\},\, \overline{P}_{V^*}\{1\}\right]$).

3 Regression

Now consider that the (unobservable) precise data are pairs $V_i = (X_i, Y_i)$, where $X_1, \ldots, X_n$ are n random objects taking values in a set $\mathcal{X}$, and $Y_1, \ldots, Y_n$ are n random variables, with $\mathcal{V} = \mathcal{X} \times \mathbb{R}$. For some $\mathcal{V}^* \subseteq 2^{\mathcal{X}\times\mathbb{R}}$ and some $\varepsilon \in [0,1]$, we consider the fully nonparametric assumption $\mathcal{P} = \mathcal{P}_\varepsilon$. In the remainder of the paper, we focus on this setting.

We want to describe the relation between X_i and Y_i by means of a function $f \in \mathcal{F}$, where $\mathcal{F}$ is a particular set of (measurable) functions $f : \mathcal{X} \to \mathbb{R}$. In order to assess the quality of the description by means of f, we define the (absolute) residuals

$$R_{f,i} := |Y_i - f(X_i)|.$$

The n random variables $R_{f,1}, \ldots, R_{f,n} \in [0, +\infty)$ are independent and identically distributed: the more their distribution is concentrated near 0, the better is the description by means of f.

In order to compare the quality of the descriptions by means of different functions $f \in \mathcal{F}$, we need to compare the concentration near 0 of the distributions of the corresponding residuals $R_{f,i}$. Usual choices of measures for this concentration are the second and first moments $E(R_{f,i}^2)$ and $E(R_{f,i})$, respectively. However, the moments of the distribution of the residuals cannot be really estimated in the fully nonparametric setting we consider, because moments are too sensitive to small variations in the distribution (see also Subsection 4.2). In fact, if $\varepsilon > 0$ or the set

$$\mathcal{R}_f := \{|y - f(x)| : (x,y) \in A,\, A \in \mathcal{V}^*\}$$

is unbounded, then the likelihood-based confidence region for any particular moment of the distribution of the residuals is unbounded (even when only the distributions with finite moments are considered), independently of the cutoff point and of the observed (imprecise) data.

By contrast, the quantiles of the distribution of the residuals can in general be estimated even in the fully nonparametric setting we consider. Therefore, we propose to use the p-quantile of the distribution of the residuals $R_{f,i}$ as a measure of the concentration near 0 of this distribution, for some $p \in (0,1)$. The technical details of the estimation of such quantiles are given in Subsections 3.1 and 3.2.

The minimizations of the second and first moments of the distribution of the residuals can be interpreted as the theoretical counterparts of the methods of least squares and least absolute deviations, respectively. In the same sense, the minimization of the p-quantile of the distribution of the residuals can be interpreted as the theoretical counterpart of the method of least quantile of squares (or absolute deviations), introduced in [15] as a generalization of the method of least median of squares (corresponding to the choice $p = 0.5$). The method of least quantile of squares leads to robust regression estimators, with breakdown point $\min\{p, 1-p\}$ (that is, the highest possible breakdown point 50% is reached when $p = 0.5$). By contrast, the methods of least squares and least absolute deviations lead to regression estimators with breakdown point 0, since they cannot even handle a single outlier (including leverage points).

In the location problem (that is, when $\mathcal{F}$ is the set of all constant functions $f : \mathcal{X} \to \mathbb{R}$), the values of the constant functions f minimizing the second and first moments of the distribution of the residuals $R_{f,i}$ are the mean and median of the distribution of Y_i, respectively (when these exist and are unique). The value of the constant function f minimizing the p-quantile of the distribution of the residuals $R_{f,i}$ is the p-center of the distribution of Y_i (that is, the center of the shortest interval containing Y_i with probability at least p), when this exists and is unique. The p-center can be interpreted as a generalization of the mode of a distribution, since under some regularity conditions the mode corresponds to the limit of the p-center when p tends to 0. The p-center of a symmetric, strictly unimodal distribution corresponds to its median and mean (when this exists), independently of p. Therefore, the minimizations of the p-quantile, first moment, and second moment of the distribution of the residuals lead to the same (correct) regression function, under the usual assumptions for the error distribution: see for example [18].

3.1 Determination of Profile Likelihood Functions for Quantiles of Residuals

We want to determine the likelihood-based confidence regions for the quantiles of the distribution of the residuals: to this purpose, we calculate the profile likelihood function for such quantiles. Let $p \in (0,1)$, and for each function $f \in F$, let $\mathcal{Q}_f := \mathcal{L}_f \cap \mathcal{U}_f$, with

$$\mathcal{L}_f = \bigcup_{r \in \mathcal{R}_f} [r, +\infty)$$

when $p > \varepsilon$ and $\mathcal{L}_f = [0, +\infty)$ otherwise, while

$$\mathcal{U}_f = \bigcup_{r \in \mathcal{R}_f} [0, r]$$

when $p < 1-\varepsilon$ and $\mathcal{U}_f = [0, +\infty)$ otherwise. It can be easily checked that $\mathcal{Q}_f$ is the set of all possible values for the p-quantile of the distribution of the residuals $R_{f,i}$, since $P(R_{f,i} \notin \mathcal{R}_f) \leq \varepsilon$. In particular, if $\varepsilon < p < 1-\varepsilon$, then $\mathcal{Q}_f$ is the smallest interval containing $\mathcal{R}_f$.

For each $f \in F$, let Q_f be the multivalued mapping from $\mathcal{P}$ to $\mathcal{Q}_f$ assigning to each probability measure P the p-quantile of the distribution of the residuals $R_{f,i}$ under P. As noted in Subsection 2.2, the mapping Q_f is multivalued, because in general quantiles are not uniquely defined. We want to determine the profile likelihood function $lik_{Q_f} : \mathcal{Q}_f \to [0,1]$ induced by the multivalued mapping Q_f. It is important to note that we would obtain the same results by considering only the distributions for which the p-quantile is unique (that is, the vagueness in the definition of quantiles has no influence on the resulting likelihood-based confidence regions).

Assume that the (imprecise) data $V_1^* = A_1, \ldots, V_n^* = A_n$ are observed, where $A_1, \ldots, A_n \in \mathcal{V}^* \backslash \{\varnothing\}$. In order to obtain the profile likelihood function lik_{Q_f} for the p-quantile of the distribution of the residuals $R_{f,i}$, we define for each function $f \in \mathcal{F}$ and each distance $q \in [0, +\infty)$ the bands

$$\overline{B}_{f,q} := \{(x,y) \in \mathcal{V} : |y - f(x)| \leq q\}$$
$$\underline{B}_{f,q} := \{(x,y) \in \mathcal{V} : |y - f(x)| < q\}$$

and the functions $\overline{k}_f, \underline{k}_f$ on $[0, +\infty)$ such that[8]

$$\overline{k}_f(q) = \#\left\{i \in \{1, \ldots, n\} : A_i \cap \overline{B}_{f,q} \neq \varnothing\right\}$$
$$\underline{k}_f(q) = \#\left\{i \in \{1, \ldots, n\} : A_i \subseteq \underline{B}_{f,q}\right\}$$

for all $q \in [0, +\infty)$. That is, $\overline{k}_f(q)$ is the number of imprecise data intersecting $\overline{B}_{f,q}$, while $\underline{k}_f(q)$ is the number of imprecise data completely contained in $\underline{B}_{f,q}$. Therefore, in particular, $\underline{k}_f(q) \leq \overline{k}_f(q)$ for all $q \in [0, +\infty)$.

Thanks to the results of Subsection 2.3 and the above definitions, we can now express the profile likelihood function lik_{Q_f} for the p-quantile of the distribution of the residuals $R_{f,i}$ as follows (a sketch of the proof is given in the Appendix):

$$lik_{Q_f}(q) = \begin{cases} \left[\lambda\left(\frac{\overline{k}_f(q)}{n}, p - \varepsilon\right)\right]^n & \text{if } \overline{k}_f(q) < (p-\varepsilon)\, n \\ \left[\lambda\left(\frac{\underline{k}_f(q)}{n}, p + \varepsilon\right)\right]^n & \text{if } \underline{k}_f(q) > (p+\varepsilon)\, n \\ 1 & \text{otherwise} \end{cases}$$

for all $q \in \mathcal{Q}_f$, where λ is the function on $[0,1] \times (0,1)$ defined by[9]

$$\lambda(s,t) = \left(\frac{s}{t}\right)^{-s} \left(\frac{1-s}{1-t}\right)^{s-1}$$

for all $s \in [0,1]$ and all $t \in (0,1)$. Hence, lik_{Q_f} is a piecewise constant function, which can take at most $n+2$ different values.

Example 4 *Consider the (imprecise) data described in Subsection 4.1 and depicted in Figure 4, and the regression function f represented by the upper curve (blue) in Figure 5. The corresponding profile likelihood function lik_{Q_f} for the 0.5-quantile of the distribution of the residuals $R_{f,i}$ is plotted in Figure 2 for the cases with $\varepsilon = 0$ (solid line) and $\varepsilon = 0.05$ (dashed line).*

[8] The cardinality of a set A is denoted by $\#A$.

[9] In this paper, $0^0 = 1$.

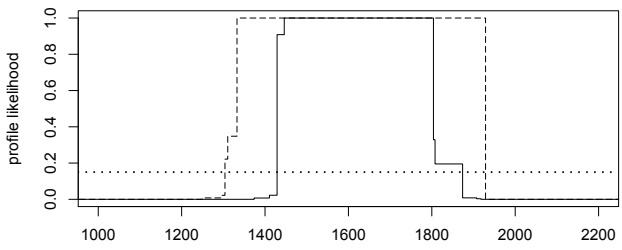

Figure 2: Profile likelihood functions from Examples 4 and 5.

3.2 Determination of Confidence Intervals for Quantiles of Residuals

Thanks to the above expression for the profile likelihood function lik_{Q_f}, we can now calculate the likelihood-based confidence regions for the quantiles of the distribution of the residuals $R_{f,i}$. Choose $\beta \in (0,1)$ and assume that

$$(\max\{p, 1-p\} + \varepsilon)^n \leq \beta \tag{2}$$

(that is, $\varepsilon < p < 1-\varepsilon$, and n is sufficiently large). Let $\mathcal{K} := \{0, \ldots, n\}$, and define

$$\underline{k} := \max\left\{k \in \mathcal{K} : k < (p-\varepsilon)\, n,\ \lambda(\tfrac{k}{n}, p-\varepsilon) \leq \sqrt[n]{\beta}\right\}$$
$$\overline{k} := \min\left\{k \in \mathcal{K} : k > (p+\varepsilon)\, n,\ \lambda(\tfrac{k}{n}, p+\varepsilon) \leq \sqrt[n]{\beta}\right\}.$$

Then $\underline{k} < \overline{k}$, and for each $f \in \mathcal{F}$, the interval

$$\mathcal{C}_f := \left\{q \in [0, +\infty) : \underline{k} < \overline{k}_f(q),\ \underline{k}_f(q) < \overline{k}\right\}$$

is the likelihood-based confidence region with cutoff point β for the p-quantile of the distribution of the residuals $R_{f,i}$. The interval $\mathcal{C}_f$ consists of all $q \in [0, +\infty)$ such that the band $\overline{B}_{f,q}$ intersects at least $\underline{k}+1$ imprecise data, and the band $\underline{B}_{f,q}$ contains at most $\overline{k}-1$ imprecise data. When $\varepsilon = 0$, the interval $\mathcal{C}_f$ is asymptotically a (conservative) confidence interval of level $F_{\chi^2}(-2\log\beta)$ for the p-quantile of the distribution of the residuals $R_{f,i}$, where F_{χ^2} is the cumulative distribution function of the chi-square distribution with 1 degree of freedom (see for example [13]). The exact level of the (conservative) confidence interval $\mathcal{C}_f$ can be obtained directly from its definition, by means of simple combinatorial arguments (also when $\varepsilon > 0$).

It is important to note that the confidence intervals $\mathcal{C}_f$ do not depend on the choice of the set $\mathcal{V}^*$ of possible imprecise data (as far as the observed ones, $A_1, \ldots, A_n$, are contained in it). This can be surprising, since the set $\mathcal{P} = \mathcal{P}_\varepsilon$ of probability measures considered depends strongly on $\mathcal{V}^*$, as noted at the beginning of Section 2. However, the independence of the confidence intervals $\mathcal{C}_f$ from the choice of the set $\mathcal{V}^*$ is not so surprising when one considers that the intervals $\mathcal{C}_f$ are likelihood-based confidence regions, and that likelihood inference is always conditional on the data (that is, independent of considerations about which other data could have been observed). This can be considered as a sort of robustness against misspecification of the set $\mathcal{V}^*$ of possible imprecise data. The practical advantage is that it is not necessary to think about which other imprecise data could have been observed, besides the ones that were actually observed (that is, $A_1, \ldots, A_n$).

Example 5 *In the situation of Example 4, the confidence interval $\mathcal{C}_f$ with $\beta = 0.15$ is approximately $[1429, 1874]$ when $\varepsilon = 0$, and $[1304, 1929]$ when $\varepsilon = 0.05$ (the cutoff point $\beta = 0.15$ is represented by the dotted line in Figure 2).*

3.3 Regression as a Decision Problem

The problem of minimizing the p-quantile of the distribution of the residuals $R_{f,i}$ can be described as a statistical decision problem: the set of probability measures considered is $\mathcal{P} = \mathcal{P}_\varepsilon$, the set of possible decisions is $\mathcal{F}$, and the loss function $L : \mathcal{P} \times \mathcal{F} \to [0, \infty)$ is defined by

$$L(P, f) = Q_f(P)$$

for all $P \in \mathcal{P}$ and all $f \in \mathcal{F}$. That is, the p-quantile of the distribution of the residuals $R_{f,i}$ is interpreted as the loss we incur when we choose the function f. In fact, the loss function L is multivalued, since in general the p-quantile is not unique: $L(P, f)$ could be reduced to a single value by taking for example the upper p-quantile of the distribution of the residuals $R_{f,i}$.

The information provided by the observed (imprecise) data is described by the likelihood function lik on $\mathcal{P}$. A very simple way of using this information consists in reducing $\mathcal{P}$ to the set $\mathcal{P}_{>\beta}$ for some cutoff point $\beta \in (0,1)$. The resulting set $\mathcal{P}_{>\beta}$ can be interpreted as an imprecise probability measure, on which we can base our choice of f. For each $f \in \mathcal{F}$, the set of all possible values of the loss $L(P, f)$ when P varies in $\mathcal{P}_{>\beta}$ can be interpreted as the imprecise p-quantile of the residuals $R_{f,i}$ under the imprecise probability measure $\mathcal{P}_{>\beta}$. It corresponds to the interval $\mathcal{C}_f$, when condition (2) is satisfied.

Assume that condition (2) is satisfied. In order to choose a function f, we can minimize the supremum of $\mathcal{C}_f$. This approach is similar to the Γ-minimax decision criterion with respect to the imprecise probability measure $\mathcal{P}_{>\beta}$, and is called LRM (likelihood-based region minimax) criterion in [4]. When there

is a unique $f \in \mathcal{F}$ minimizing $\sup \mathcal{C}_f$, it can be denoted by f_{LRM}, and $\sup \mathcal{C}_f$ can be denoted by $\overline{q}_{LRM}$. In this case, f_{LRM} is characterized geometrically by the fact that $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ is the thinnest band of the form $\overline{B}_{f,q}$ containing at least $\overline{k}$ imprecise data, for all $f \in \mathcal{F}$ and all $q \in [0, +\infty)$. Finding the function f_{LRM} is an interesting computational problem: see for example [2, 15, 22].

An interesting description of the uncertainty about the optimal choice of $f \in \mathcal{F}$ is obtained by considering interval dominance for the imprecise p-quantiles of the residuals $R_{f,i}$ under the imprecise probability measure $\mathcal{P}_{>\beta}$. When f_{LRM} exists, the undominated functions $f \in \mathcal{F}$ are those such that $\mathcal{C}_f$ intersects $\mathcal{C}_{f_{LRM}}$. In particular, when $\overline{q}_{LRM} \in \mathcal{C}_{f_{LRM}}$ (that is, $\mathcal{C}_{f_{LRM}}$ is right-closed), the undominated functions $f \in \mathcal{F}$ are characterized geometrically by the fact that $\overline{B}_{f,\overline{q}_{LRM}}$ intersects at least $\underline{k}+1$ imprecise data. In general, the set of undominated functions f tends to get smaller when we observe more and more (imprecise) data, but it does not necessarily tend to reduce to a singleton, because of the unavoidable uncertainty discussed in Subsection 2.1.

3.4 Prediction

Consider the case in which (instead of n) we have $n+1$ pairs (V_i, V_i^*) of precise and imprecise data $V_i = (X_i, Y_i)$ and V_i^*, respectively. We want to predict the realization of the precise data value V_{n+1} on the basis of the realization of the n imprecise data $V_1^*, \ldots, V_n^*$. Choose $k \in \{1, \ldots, n\}$, and assume that for each possible realization of the $n+1$ imprecise data $V_1^*, \ldots, V_{n+1}^*$, there is a distance $q' \in [0, +\infty)$ such that for some $f' \in \mathcal{F}$ (not necessarily unique), $\overline{B}_{f',q'}$ is a thinnest band of the form $\overline{B}_{f,q}$ containing at least k of the $n+1$ imprecise data, for all $f \in \mathcal{F}$ and all $q \in [0, +\infty)$. Because of symmetry, the probability that V_{n+1}^* is included in a band $\overline{B}_{f,q'}$ containing at least k of the $n+1$ imprecise data (for some $f \in \mathcal{F}$) is at least $\frac{k}{n+1}$. Hence, when $\overline{B}_{f'',q''}$ is a thinnest band of the form $\overline{B}_{f,q}$ containing at least k of the n imprecise data $V_1^*, \ldots, V_n^*$ (for all $f \in \mathcal{F}$ and all $q \in [0, +\infty)$), the probability that V_{n+1}^* is included in the union $\mathcal{B}$ of all bands $\overline{B}_{f,q''}$ containing at least $k-1$ of the n imprecise data $V_1^*, \ldots, V_n^*$ (for all $f \in \mathcal{F}$) is at least $\frac{k}{n+1}$. That is, $\mathcal{B}$ is a (conservative) prediction region of level $\frac{k}{n+1} - \varepsilon$ for the precise data value V_{n+1}.

In particular, when condition (2) is satisfied and f_{LRM} exists, the union $\mathcal{B}$ of all bands $\overline{B}_{f,\overline{q}_{LRM}}$ containing at least $\overline{k}-1$ of the n imprecise data $V_1^*, \ldots, V_n^*$ (for all $f \in \mathcal{F}$) is a (conservative) prediction region of level $\frac{\overline{k}}{n+1} - \varepsilon$ for the precise data value V_{n+1}. Prediction regions of this form can sometimes be reduced to smaller regions thanks to the assumption that V_{n+1}^* takes values in $\mathcal{V}^*$. When besides the realization of the n imprecise data $V_1^*, \ldots, V_n^*$, also the (precise or imprecise) realization of X_{n+1} has been observed, the realization of Y_{n+1} can be predicted for example by using the idea of conformal prediction (see [21]), but this goes beyond the scope of the present paper.

4 Example of Application

In this section, we apply the proposed regression method to socioeconomic data from the ALLBUS (German General Social Survey). Data collection in surveys is subject to many different influences that may cause various biases in the data set (see for example [3]). Therefore, it is often reasonable to assume that the actual value lies rather in some interval around the observed value. Furthermore, data on sensitive quantities is sometimes only available in categories that form a partition of the space of possible values. A simple approach to analyze this kind of data is to reduce the intervals to their central values and to apply usual regression methods to the reduced, precise data. In contrast to this, we suggest to analyze directly the interval-valued data by means of the regression method proposed in Section 3.

We want to investigate the age-income profile, which is a fundamental relationship in the social sciences and a typical example in textbooks on social research methods (see for example [1]).

Income is a key demographic variable for socioeconomic research questions. But asking for income in an interview is a sensitive question that some respondents refuse to answer. Thus, survey data on income often include missing values. One way to make the question less sensitive is to present predefined income categories according to which the income of the respondent shall be classified. In the ALLBUS, income data is collected with a two-step design with the open question for income as first step and the presentation of a category scheme as second step. As a result, the data set contains at the same time precise values for some individuals and interval-valued observations for others. Yet, even if the respondents are willing to give their exact income, limited remembrance usually prevents them from doing so. Instead, they will give rounded and heaped values (see [9]), where heaping refers to irregular rounding behavior (see for example [10]). Therefore, it is more reliable to regard also the precise income values as interval-valued observations.

Data on the age of respondents is more easily obtained, but it is always measured with limited precision, e.g. in years. In this case, it might be useful to

consider intervals $[age, age+1)$ instead. Furthermore, age data might be available as age classes only.

4.1 ALLBUS Data and Regression Model

We analyze the ALLBUS data set of 2008 containing 3 247 interviews. The considered variables are *personal income* (on average per month) and *age*. Here, we consider the worst case, where both variables are available in categories only ($v389$ and $v155$ of the data set with 22 possible income categories and six age classes; see [19]), although the proposed regression method could be applied to the data set with some precise and some imprecise observations, too. Thus, for each individual $i \in \{1, \ldots, n\}$ we consider observations $V_i^* = X_i^* \times Y_i^*$, where $X_i^* = [\underline{x}_i, \overline{x}_i)$ is the corresponding age class and $Y_i^* = [\underline{y}_i, \overline{y}_i)$ is the category into which the income of respondent i falls. In the given data set, there are 620 missing income values and 11 missing age values. Missing values are replaced by intervals that cover the entire observation space of each variable. In this case, $X_i^* = [18, 100)$ or $Y_i^* = [0, +\infty)$, respectively. A two-dimensional histogram of the data set is given in Figure 4.

The relationship between age and income is usually modeled by a quadratic function in age (see for example [1]). Thus, the set of regression functions we consider here is

$$\mathcal{F} = \{f_{a,b_1,b_2} : \ a, b_1, b_2 \in \mathbb{R}\},$$

where each function f_{a,b_1,b_2} is defined by

$$f_{a,b_1,b_2}(x) = a + b_1\, x + b_2\, x^2$$

for all $x \in \mathcal{X} := [18, 100)$. We choose to minimize the 0.5-quantile of the distribution of the residuals (i.e., $p = 0.5$), and we take the cutoff point $\beta = 0.15$. Furthermore, we want to compare the results obtained by the proposed method with those from an ordinary least squares (OLS) regression based on the interval centers. Since the latter implies the assumption $P(V_i \in V_i^*) = 1$, we also set $\varepsilon = 0$ here.

We conduct the regression analysis as follows: First, the likelihood-based confidence regions $\mathcal{C}_{f_{a,b_1,b_2}}$ are computed for reasonable parameter values (a, b_1, b_2). Then, we identify the parameter combination among these that minimizes the upper bound of $\mathcal{C}_{f_{a,b_1,b_2}}$. The function corresponding to this parameter combination is the function f_{LRM} which is optimal according to the LRM criterion (see Subsection 3.3). Finally, the upper bound $\overline{q}_{LRM}$ of $\mathcal{C}_{f_{LRM}}$ is used to determine the set of undominated functions.

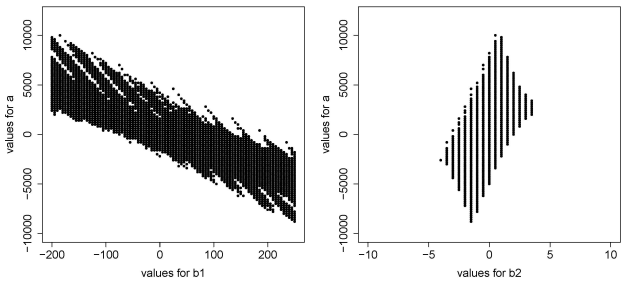

Figure 3: Two-dimensional projections of the set of undominated parameter values.

4.2 Results

We considered a grid of combinations of parameter values where $a \in [-10\,000,\ 12\,000]$, $b_1 \in [-200,\ 250]$, and $b_2 \in [-10,\ 10]$. Corresponding to the set of undominated functions, we find the set of undominated parameter combinations displayed in Figure 3. This set is clearly not convex. Moreover, in the case considered here, the parameters are not independent from each other, in the sense that many different combinations of parameter values (a, b_1, b_2) may lead to very similar functions f_{a,b_1,b_2} over $\mathcal{X}$. Thus, there are actually infinitely many undominated parameter combinations, but the associated curves are similar to those we find within the considered grid.

The parameter combination implying the smallest upper endpoint of the confidence interval for the 0.5-quantile of the residuals is (850, 0, 0) with $\mathcal{C}_{f_{850,0,0}} = [525,\ 650]$. The function f_{LRM} is thus a constant line: this is due to the rectangular shape and the locations of the observations in our data set. Hence, the value 850 can be interpreted as an estimate of the p-center (with $p = 0.5$) of the income distribution (see the beginning of Section 3). A further interpretation of the function f_{LRM} is given by the band $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ limited by the functions $f_{LRM} - \overline{q}_{LRM}$ and $f_{LRM} + \overline{q}_{LRM}$: Among all bands constructed around all considered functions, this band is the thinnest one that contains at least $\overline{k} = 1\,679$ imprecise observations (see Subsection 3.3).

The function f_{LRM} and the band $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ are presented in Figure 5, besides the undominated functions. It can be seen that within the set of undominated functions there is a large variety of shapes of the age-income profile, including straight lines, convex parabolic curves as well as concave ones. From a social scientist's point of view this result may be unsatisfying because it doesn't support only one form of the relationship between age and income. However, given the imprecision of the data, it is reasonable to consider all shapes consistent with the data as possi-

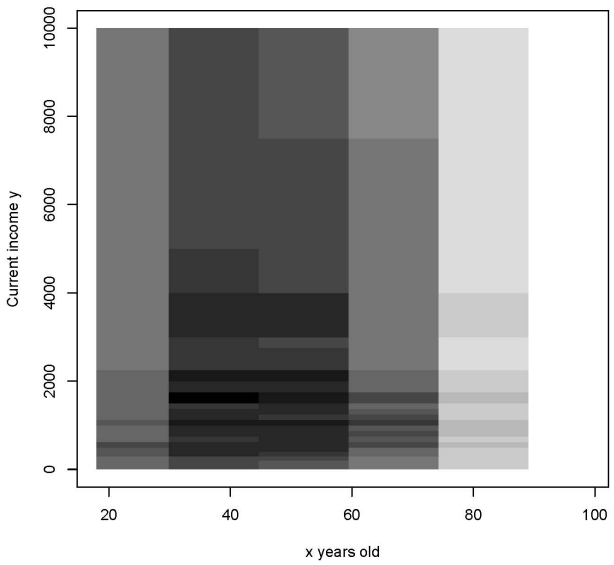

Figure 4: Two-dimensional histogram of the data set.

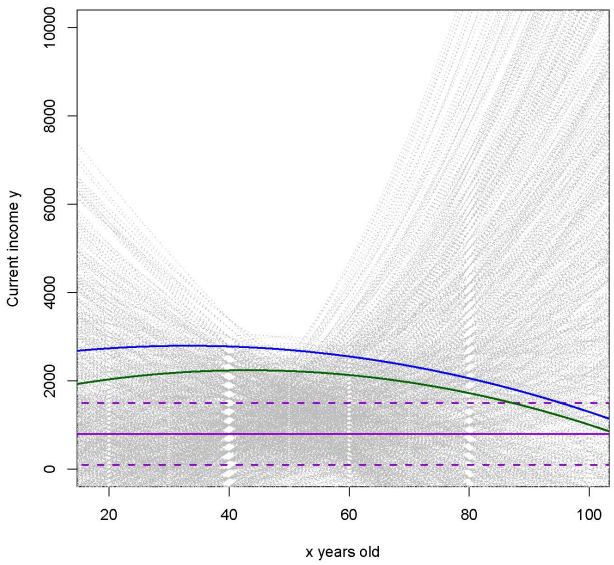

Figure 5: Undominated functions (dotted curves, gray), interval data-based f_{LRM} (solid line, violet) and band $\overline{B}_{f_{LRM},\overline{q}_{LRM}}$ (dashed lines, violet) versus OLS regressions on interval centers with upper income limit 15 000 (upper curve, blue) and upper income limit 10 000 (lower curve, green).

ble age-income profiles. If the observed intervals were overlapping or if they constituted a finer partition of the space of possible observations, the set of undominated functions would be smaller. Hence, the set of undominated functions can be interpreted as the set of plausible descriptions of the age-income profile that reflects at the same time the uncertainty inherent in the imprecise data.

The usual method to analyze this kind of interval data is to conduct a quadratic OLS regression based on the interval centers ignoring the imprecision of the data. In this case, one has to give an upper limit for the highest income class [7 500, $+\infty$) in order to compute the interval centers. Of course, the choice of this upper limit has an impact on the estimates of the OLS regression. The effect of two different choices of the upper income limit is illustrated in Figure 5. The OLS curves displayed there are based on interval centers with upper income limits 15 000 and 10 000, respectively. In contrast to the OLS approach, the regression method proposed in this paper is not sensitive to the extremes, since the regression functions are evaluated on the basis of confidence regions for the 0.5-quantile of the residuals' distribution.

The proposed regression method permits to identify plausible descriptions of the relationship between the socioeconomic characteristics *age* and *income*. Given the imprecise data, many different shapes of the age-income profile are plausible. Further computations indicated that our findings hold for transformed income data on the logarithmic scale, too. The results are not very informative, but reliable. To obtain more informative, but less reliable results, it suffices to increase the cutoff point β (that is, to decrease the confidence level of the intervals $\mathcal{C}_{f_{a,b_1,b_2}}$). One idea to obtain more informative results without sacrificing reliability could be to use many different category schemes during the income data collection and thereby obtain a data set with overlapping categories.

5 Conclusion

In this paper, we introduced a robust approach to regression with imprecise data, in which the error distribution is not constrained to a particular parametric family. The method was presented within a very general framework and it can be adapted to a wide range of practical settings, since it can be applied to all kinds of imprecise data covering e.g. interval data, precise data, and missing data. In our method, the imprecise data are interpreted as the result of a coarsening process which can be informative, and even wrong with a certain probability.

In future work, the statistical properties of the proposed regression method shall be studied in more detail. In particular, we plan to investigate the impact of stronger assumptions about the error distribution and the coarsening process. Moreover, the performance of the regression method shall be compared to those of alternative approaches to regression with imprecise data, also with regard to computational aspects.

Acknowledgements

The authors wish to thank Thomas Augustin and the anonymous referees for their helpful comments.

Appendix

The expression for the profile likelihood function lik_{Q_f} given in Subsection 3.1 can be proved as follows. In Subsection 2.3, we have seen that $lik_{Q_f} = lik^*_{Q^*_f}$, where lik^* and Q^*_f are defined on the set $\mathcal{P}_{V^*}$ of all possible distributions P_{V^*} for the imprecise data V^*_i. The function lik^* assigns to each P_{V^*} the corresponding likelihood value: in particular, it has a unique maximum in the empirical distribution (of the imprecise data) $\widehat{P}_{V^*}$. The multivalued mapping Q^*_f assigns to each P_{V^*} all p-quantiles of the residuals $R_{f,i}$ for all distributions of the precise data V_i compatible with P_{V^*}. Consider in particular $Q^*_f(\widehat{P}_{V^*})$: if $\varepsilon = 0$, then $q \in \mathcal{Q}_f$ is a p-quantile of the residuals $R_{f,i}$ for some distribution of the precise data V_i compatible with $\widehat{P}_{V^*}$ if and only if $\underline{k}_f(q) \leq p\,n \leq \overline{k}_f(q)$. The case with $\varepsilon > 0$ corresponds to the case with $\varepsilon = 0$ when $Q^*_f(\widehat{P}_{V^*})$ is enlarged to all p'-quantiles of the residuals $R_{f,i}$ such that $p - \varepsilon \leq p' \leq p + \varepsilon$. This proves the "otherwise" part of the expression for lik_{Q_f} given in Subsection 3.1, since $lik^*(\widehat{P}_{V^*}) = 1$.

Now assume that $q \in \mathcal{Q}_f$ satisfies $\overline{k}_f(q) < (p - \varepsilon)\,n$. Let $P'_{V^*} \in \mathcal{P}_{V^*}$ be the empirical distribution obtained when only the $n - \overline{k}_f(q)$ imprecise data not intersecting $\overline{B}_{f,q}$ are considered, and let $P''_{V^*} \in \mathcal{P}_{V^*}$ be the empirical distribution obtained when only the $\overline{k}_f(q)$ imprecise data intersecting $\overline{B}_{f,q}$ are considered. The latter is not well-defined when $\overline{k}_f(q) = 0$: in this case, let $P''_{V^*} \in \mathcal{P}_{V^*}$ be the Dirac distribution assigning probability 1 to a set $A \in \mathcal{V}^*$ intersecting $\overline{B}_{f,q}$ (such a set A exists, since $q \in \mathcal{Q}_f$). Then $q \in Q^*_f(P'''_{V^*})$ with $P'''_{V^*} = (p - \varepsilon)\,P''_{V^*} + (1 - p + \varepsilon)\,P'_{V^*}$, and it can be easily checked that

$$lik^*_{Q^*_f}(q) = lik^*(P'''_{V^*}) = \left[\lambda\left(\frac{\overline{k}_f(q)}{n},\, p - \varepsilon\right)\right]^n .$$

This proves the first case of the expression for lik_{Q_f} given in Subsection 3.1, and the second one can be proved analogously.

References

[1] Allison, P. D. (1998). *Multiple Regression*. Pine Forge Press.

[2] Bernholt, T. (2005). Computing the least median of squares estimator in time $O(n^d)$. In *Computational Science and Its Applications — ICCSA 2005*. Springer, 697–706.

[3] Biemer, P. P., and Lyberg, L. E. (2003). *Introduction to Survey Quality*. Wiley.

[4] Cattaneo, M. (2007). *Statistical Decisions Based Directly on the Likelihood Function*. PhD thesis, ETH Zurich. doi:10.3929/ethz-a-005463829.

[5] Cattaneo, M. (2008). Fuzzy probabilities based on the likelihood function. In *Soft Methods for Handling Variability and Imprecision*. Springer, 43–50.

[6] de Cooman, G., and Zaffalon, M. (2004). Updating beliefs with incomplete observations. *Artif. Intell.* 159, 75–125.

[7] Domingues, M. A. O., de Souza, R. M. C. R., and Cysneiros, F. J. A. (2010). A robust method for linear regression of symbolic interval data. *Pattern Recognit. Lett.* 31, 1991–1996.

[8] Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., and Ginzburg, L. (2007). *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*. Technical Report SAND2007-0939. Sandia National Laboratories.

[9] Hanisch, J. U. (2005). Rounded responses to income questions. *Allg. Stat. Arch.* 89, 39–48.

[10] Heitjan, D. F., and Rubin, D. B. (1991). Ignorability and coarse data. *Ann. Stat.* 19, 2244–2253.

[11] Hudson, D. J. (1971). Interval estimation from the likelihood function. *J. R. Stat. Soc. B* 33, 256–262.

[12] Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer.

[13] Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC.

[14] Pawitan, Y. (2001). *In All Likelihood*. Oxford University Press.

[15] Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley.

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

[17] Strassen, V. (1964). Meßfehler und Information. *Z. Wahrscheinlichkeitstheorie* 2, 273–305.

[18] Tasche, D. (2003). Unbiasedness in least quantile regression. In *Developments in Robust Statistics*. Physica-Verlag, 377–386.

[19] Terwey, M., and Baltzer, S. (2009). *ALLBUS Datenhandbuch 2008*. GESIS.

[20] Utkin, L., Zatenko, S., and Coolen, F. (2009). Combining imprecise Bayesian and maximum likelihood estimation for reliability growth models. In *ISIPTA '09*. SIPTA, 421–430.

[21] Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer.

[22] Watson, G. A. (1998). On computing the least quantile of squares estimate. *SIAM J. Sci. Comput.* 19, 1125–1138.

[23] Zaffalon, M., and Miranda, E. (2009). Conservative inference rule for uncertain reasoning under incompleteness. *J. Artif. Intell. Res. (JAIR)* 34, 757–821.