Towards Handling Uncertainty-at-Source in AI – A Review and Next Steps for Interval Regression

Shaily Kabir, Christian Wagner, *Senior Member, IEEE,* and Zack Ellerby

***Abstract*—Most of statistics and AI draw insights through modelling discord or variance between sources (i.e., inter-source) of information. Increasingly however, research is focusing on uncertainty arising at the level of individual measurements (i.e., within- or intra-source), such as for a given sensor output or human response. Here, adopting intervals rather than numbers as the fundamental data-type provides an efficient, powerful, yet challenging way forward—offering systematic capture of uncertainty-at-source, increasing informational capacity, and ultimately potential for additional insight. Following progress in the capture of interval-valued data in particular from human participants, conducting machine learning directly upon intervals is a crucial next step. This paper focuses on linear regression for interval-valued data as a recent growth area, providing an essential foundation for broader use of intervals in AI. We conduct an in-depth analysis of state-of-the-art methods, elucidating their behaviour, advantages, and pitfalls when applied to synthetic and real-world data sets with different properties. Specific emphasis is given to the challenge of preserving mathematical coherence, i.e., models maintain fundamental mathematical properties of intervals. In support of real-world applicability of the regression methods, we introduce and demonstrate a novel visualization approach, the interval regression graph, or *IRG*, which effectively communicates the impact of both position and range of variables within the regression models—offering a leap in their interpretability. Finally, the paper provides practical recommendations concerning regression-method choice for interval data and highlights remaining challenges and important next steps for developing AI with the capacity to handle uncertainty-at-source.**

***Impact Statement*—Capturing information as intervals provides a powerful means for handling uncertainty and inherent range in data. This paper focuses on a basic building block of statistics and AI: linear regression. In recent years, regression for intervals has become a topic of interest in AI and has been applied to domains ranging from marketing to cyber-security. It allows direct modelling of relationships not only between variables per-se, but also their associated uncertainty. For example, we can infer not only how a snack's nutritional benefits impact consumer purchase intention, but also how uncertainty about these benefits impacts purchase intention and its associated uncertainty. Nonetheless, while there are considerable upsides to interval regression and AI, substantial challenges remain. This paper reviews and extends state-of-the-art interval regression methods, presents in-depth experimental results and introduces a novel visualization approach for interval regression—enhances interpretability and facilitates real-world insight. Finally, it provides recommendations for algorithm selection based on key properties of data sets, discusses next steps for interval-valued AI beyond linear regression, and is complemented by an open-source implementation of the algorithms discussed.**

***Index Terms*—Intervals, regression, uncertainty, intra-source, inter-source**

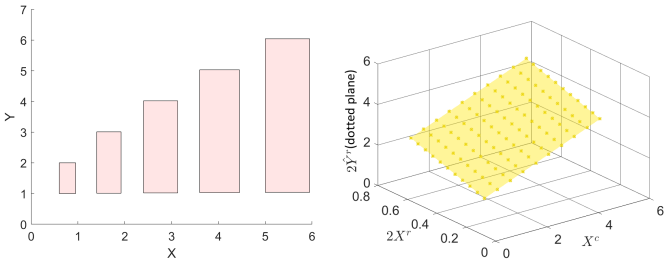
S. Kabir, C. Wagner and Z. Ellerby are with the Lab for Uncertainty in Data and Decision Making (LUCID), School of Computer Science, University of Nottingham, Nottingham, UK (e-mail: shaily.kabir3, christian.wagner, zack.ellerby@nottingham.ac.uk).

I. Introduction

INTERVAL-VALUED (IV) data have gained increasing attention across a variety of contexts as they offer a systematic way to capture information complete with an intrinsic representation of range or uncertainty in each individual 'measurement', which is not possible using point-values, such as numbers or ranks. Such data may arise due to imprecision and uncertainty in measurement (e.g., sensor data), inherent uncertainty of outcome (e.g., estimations of stock prices [1], climate/weather forecasting [2]), or inherent vagueness or nuance (e.g., in linguistic terms [3]). A growing body of evidence also suggests that subjective judgement of humans (e.g., experts, consumers) may be better represented by intervals [4], [5], where the interval width/size captures the response range as a conjunctive set (e.g., the reals between 2 and 4), or degree of uncertainty with respect to an estimate or rating—a disjunctive set (e.g., confidence interval) [6].

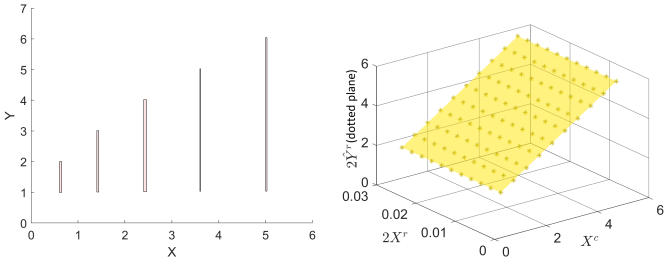
Regression analysis for estimating one or more variables is an important task in data analysis and machine learning under the wider umbrella of AI. While regression is most commonly applied to numeric data, researchers have begun to develop methods of applying (linear) regression analysis to IV data [7]. To date, models designed to handle data sets where *both* independent (regressor) and dependent (regressand) variables are IV have been developed within the frameworks of random sets and symbolic data analysis (SDA) [8], [9]. In the random set framework, intervals are viewed as compact convex sets in $\mathbb{R}$, and their interaction is modelled according to set arithmetic [10]–[14]. On the other hand, in the SDA setting, the classical regression model is expanded to IV data where intervals are treated as bi-variate vectors [15]–[22].

Current linear regression models for intervals use different *reference points*, such as the regressors' center values, their lower and upper bounds, or both their center and range (width), to estimate the IV regressand [21]. Earlier interval regression approaches [15]–[17] risk 'flipping' the interval bounds, respectively generating a negative interval range, thus breaking mathematical coherence (i.e., violating the definition of interval). The more recent approaches [18]–[21] impose either positivity restrictions on parameters, or design the regression model so as to ensure that the lower bound does not exceed the upper bound. For instance, Neto and Carvalho [18] apply the Lawson and Hanson algorithm (LHA) [23], Wang et al. [19] use Moore's linear combination [24], and Souza et al. [21] the Box-Cox transformation [25] to guarantee mathematical coherence. Although Sun and Ralescu [20] consider positivity restrictions on parameters in their model setting, they do not



(a) Set-1 (b) *IRG* for Set-1

Fig. 1: (a) Visualization of IV data Set-1 and (b) *IRG* with respect to range, $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$) using the PM method [21].



(a) Set-2 (b) *IRG* for Set-2

Fig. 2: (a) Visualization of IV data Set-2 and (b) *IRG* with respect to range, $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$) using the PM method [21].

provide a method to maintain these restrictions. A detailed review and analysis of these models is provided in Section II and the general issue is discussed throughout the paper.

An important aspect is that most current approaches, including the above, restrict parameters based on the assumption that an increase in uncertainty[1] of a regressor (i.e., input) leads to greater uncertainty in the dependent variable or regressand. However, there are plausible cases in the real world where this positive correlation does not hold. For example, in cybersecurity, uncertainty around the vulnerability of a system may lead to certainty in the need to invest resources into controls—thus mitigating the risk of the system being compromised, that is, a negative correlation. Similarly, in a multivariate case, consider an economics example; greater uncertainty in future product prices and interest rates may cause more certainty with respect to present consumption choices.

For illustration, Figs. 1(a) and 2(a) show a synthetic example of two IV data sets. Here, in Set-1 (Fig. 1(a)), the position (center) and range (uncertainty) of the regressand, Y increases with the position (center) and range (e.g. the uncertainty) of regressor, X—that is, they vary in unison—this reflects the most common case as outlined above, e.g., the price of cars vs their horsepower. In contrast, for Set-2 (Fig. 2(a)), the range (uncertainty) of the regressand, Y increases irrespective of the range (uncertainty) of the regressor, X. This reflects a less common scenario, e.g., the wealth of people with respect to their age. The relationships between the variables in each figure are captured through Interval Regression Graphs (*IRG*s) in Figs. 1(b) and 2(b) respectively. *IRG*s will be introduced in detail in Section III-B, but for now, note how Fig. 1(b) shows a gentle increase in the range of the dependent variable ($2Y^r$) for an increase in the range of the independent variable ($2X^r$) for Set-1, while Fig. 2(b) shows, as expected, no such change for Set-2. The *IRG*s also highlight the much less drastic increase in range of Set-1's dependent variable that is attributed to an increase in the position of the independent variable (X^c), compared to Set-2, in which variance in regressand range ($2Y^r$) is attributed solely to regressor position (X^c).

Aspects such as the above highlight the complexity but also the potential of IV data in directly capturing uncertainty and range at-source, making it accessible for reasoning. With intervals encoding additional information by comparison to numbers, the choice of a 'universally best' analysis approach, as can often be found for numeric cases, is however highly challenging. Instead, the selection of appropriate techniques for a given IV data set—based on its properties—is essential. Ferson et al. in [26], discuss the complexity of analysing real-world IV data in detail, elucidating the value of focusing on key properties for data sets to inform the choice of methods, such as whether the intervals exhibit strong overlap (i.e., 'puffy' [26]: wider, more uncertain intervals) or very little overlap (i.e., 'skinny' [26]: narrow, less uncertain); whether there are outliers; whether intervals are highly scattered, densely placed or even nested within each other. Beyond these properties, more fundamental questions arise from the semantic meaning of the given intervals, as already alluded to above; whether they represent an epistemic, disjunctive set, with the interval describing the range within which a correct answer is contained (this is arguably the most common use of intervals in science, cf. *confidence intervals*), or indeed an ontic, conjunctive set with a true range, such as the real numbers between 1 and 5, or a time span [5], [27]. For conciseness, we do not discuss the latter further in this manuscript, but will revisit it, and its impact on the analysis of AI and machine learning more generally, in future publications.

At this stage, we also note that of course one can go beyond intervals to model uncertainty, leveraging distributions with a probabilistic or fuzzy interpretation, depending on the dis- or con-junctive nature of the information. Focusing on intervals a priori provides advantages, such as increasingly established pathways to collecting data (e.g., [5]), while also laying foundations for more advanced methods. For example, in fuzzy set theory, α-cut decomposition provides a direct translation mechanism between intervals and fuzzy sets.

This paper makes four principal contributions:

1) *Review of the state-of-the-art:* In view of the significance and use of intervals as a fundamental data-type in AI, the paper presents a detailed review of all well-known state-of-the-art linear regression approaches based on vector representation of data sets where *both* independent and dependent variables are IV.
2) *Extensions of current methods:* As part of the review of the methods, this paper puts forward extensions to the Linear Model (LM) [20], enhancing its robustness efficiently for real-world applications. Specifically, we propose a dynamically selective approach to applying restrictions, letting the model's parameters vary freely unless the mathematical coherence is being compromised.

[1]Throughout the remainder of this paper, for simplicity, we refer to the range of the interval as representing uncertainty as a general term capturing vagueness, lack of information, etc.

TABLE I: Acronyms and Notation

CM	Center Method [15]
MinMax	MinMax Method [16]
CRM	Center and Range Method [17]
CCRM	Constrained Center and Range Method [18]
CIM	Complete Information Method [19]
LM	Linear Model [20]
PM	Parametrized Model [21]
LM_c	Constrained Linear Model
LM_w	Weakly Constrained Linear Model
IV	Interval-valued
IRG	Interval Regression Graph
$\overline{a}$	Interval $\{\overline{a} \subseteq \mathbb{R} : \overline{a} = [a^-, a^+],\ a^- \leq a^+\}$
a^w	Range of Interval, $a^w = \lvert a^+ - a^- \rvert$
a^c	Center of Interval, $a^c = \frac{(a^+ + a^-)}{2}$
a^r	Radius of Interval $a^r = \frac{a^w}{2}$ (Half Range)
Y	Interval-valued Regressand
$\hat{Y}$	Estimated Y
X	Interval-valued Regressor
μ_{Y^w} and σ_{Y^w}	Mean and Standard Deviation of Range of Y
μ_{X^w} and σ_{X^w}	Mean and Standard Deviation of Range of X
μ_{Y^c} and σ_{Y^c}	Mean and Standard Deviation of Center of Y
μ_{X^c} and σ_{X^c}	Mean and Standard Deviation of Center of X

3) *Practical guidance on algorithm selection and software:* To support the broader adoption of intervals as a fundamental data-type in AI, extant regression approaches and their performance are analysed using synthetic and real world data sets capturing various common and features of intervals, such as 'puffy' and 'skinny' intervals [26]. The resulting insights are distilled, for the first time, to guide readers in selecting the regression algorithms appropriate to *their* data. Further, we provide open-source software at http://lucidresearch.org/software, offering, for the first time, direct access to IV regression approaches, with a view to facilitating adoption and replication.
4) *Explainability and visualization:* The explainability of numeric linear regression models is one of their primary assets as an AI technique. This paper introduces a novel visualization approach for IV regression—the Interval Regression Graph (*IRG*)—which for the first time provides the means to visualize and succinctly communicate the relationship in terms of both position and range between IV regressor and regressand.[2]

The paper is organized as follows: Section II reviews linear regression methods based on vector/matrix representation of intervals in the regression process. Section III introduces the interval regression graph (*IRG*) and discusses adaptations of the LM method [20] for real-world applications, followed by IV data sets (synthetic and real-world) and evaluation metrics. Section IV demonstrates the estimation performance of all regression models with respect to both synthetic and real-world data sets. Section V discusses suitability of the existing regression approaches to salient features of IV data, distilling insights into practical guidance for algorithm-selection. Lastly, Section VI concludes the paper with future works. Table I presents a list of acronyms and notation used in this paper to assist the reader.

[2]Note that in this paper, we focus on establishing and visualizing the relationship between the variables of a given IV data set. We do not focus on *prediction*, a different problem setting with implications for example the need to establish model and prediction robustness, e.g., using cross-validation [28]—offering scope for future research for IV data.

II. Background and State of the Art

A closed interval $\overline{a}$ is characterized by its two endpoints, a^- and a^+ with $a^- \leq a^+$ where a^- and a^+ are respectively the lower and upper bounds of $\overline{a}$ [29]. Generally, $\overline{a}$ is denoted as $\overline{a} = [a^-, a^+]$, and its range is given by $a^w = \lvert a^+ - a^- \rvert$. Alternatively, the same interval can be represented as $\overline{a} = [a^c - a^r, a^c + a^r]$ with its center, $a^c = \frac{(a^+ + a^-)}{2}$ and radius, $a^r = \frac{a^w}{2}$. In this section, we review linear regression methods for *full* IV data sets (i.e. where both dependent and independent variables are intervals) where intervals are treated as bi-variate vectors in the regression process. We note $Y = \{\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_n\}$ as an IV regressand with n intervals, where $\overline{y}_i = [y_i^-, y_i^+]$, $1 \leq i \leq n$. Further, $\{X_1, X_2, \ldots, X_p\}$ presents p (≥ 1) IV regressor(s) where each X_j has also n intervals, $X_j = \{\overline{x}_{j1}, \overline{x}_{j2}, \ldots, \overline{x}_{jn}\}$ with $\overline{x}_{ji} = [x_{ji}^-, x_{ji}^+]$, $1 \leq i \leq n$, $1 \leq j \leq p$. Thus, there are p pairs of (Y, X) in the data set. In addition, the estimated value of Y is given by $\hat{Y} = \{\hat{\overline{y}}_1, \hat{\overline{y}}_2, \ldots, \hat{\overline{y}}_n\}$.

Since about the year 2000, a variety of linear regression approaches for *full* IV data sets have been put forward. We briefly review them below in chronological order.

A. The Center Method (2000)

Billard and Diday [15] proposed a linear regression model, known as the Center Method (CM) in 2000, to fit the center of both regressand and regressors, and then apply this model to the lower and upper bounds of regressors to separately estimate the upper and lower bounds of the regressand. Equation (1) defines the CM regression model,

$$\begin{aligned} y_i^- &= \beta_0^c + \sum_{j=1}^{p} \beta_j^c x_{ji}^- + \epsilon_i^- \\ y_i^+ &= \beta_0^c + \sum_{j=1}^{p} \beta_j^c x_{ji}^+ + \epsilon_i^+ \end{aligned} \tag{1}$$

where y_i^- and y_i^+ are the lower and upper bounds of $\overline{y}_i$. Similarly, x_{ji}^- and x_{ji}^+ are the the lower and upper bounds of $\overline{x}_{ji}$. $\beta_j^c \in \mathbb{R}$ are the regression coefficients for the center estimation, $0 \leq j \leq p$, p is the total number of regressors while ϵ_i^- and ϵ_i^+ are error terms. Both lower and upper bound models of (1) use the same coefficients, therefore, they can be captured together using (2).

$$y_i^c = \beta_0^c + \sum_{j=1}^{p} \beta_j^c x_{ji}^c + \epsilon_i^c, \tag{2}$$

where y_i^c and x_{ji}^c are the center of $\overline{y}_i$ and $\overline{x}_{ji}$ respectively and $\epsilon_i^c = (\epsilon_i^- + \epsilon_i^+)/2$ is error term. Equation (2) can also be expressed for the total set of n observations in the matrix format as $Y^c = X^c\beta^c + \epsilon^c$, where

$$Y^c = (y_1^c \quad y_2^c \quad \ldots \quad y_n^c)^T, \quad \beta^c = (\beta_0^c \quad \beta_1^c \quad \ldots \quad \beta_p^c)^T$$

$$\epsilon^c = (\epsilon_1^c \quad \epsilon_2^c \quad \ldots \quad \epsilon_n^c)^T, \text{ and } X^c = \begin{pmatrix} 1 & x_{11}^c & \ldots & x_{p1}^c \\ 1 & x_{12}^c & \ldots & x_{p2}^c \\ \vdots & \vdots & \ldots & \vdots \\ 1 & x_{1n}^c & \ldots & x_{pn}^c \end{pmatrix}.$$

The coefficients, β^c are estimated using the least squares (LS) method [30], defined in (3).

$$\hat{\beta}^c = ((X^c)^T X^c)^{-1} (X^c)^T Y^c \tag{3}$$

By using $\hat{\beta}^c$, (4) estimates the bounds of regressand, Y.

$$\begin{aligned} \hat{y}_i^- &= \hat{\beta}_0^c + \sum_{j=1}^{p} \hat{\beta}_j^c x_{ji}^- \\ \hat{y}_i^+ &= \hat{\beta}_0^c + \sum_{j=1}^{p} \hat{\beta}_j^c x_{ji}^+. \end{aligned} \tag{4}$$

The CM model is simple, as it follows a standard regression approach applied to the center of the intervals and then the resulting model is applied both to upper and lower endpoints. However, by doing so, it risks and often fails to maintain mathematical coherence [17], that is, the estimated lower bound of $\overline{y}_i$ becomes larger than the estimated upper bound.

B. The MinMax Method (2002)

Billard and Diday [16] presented the MinMax method in 2002 which directly uses the lower and upper bounds of the regressors to separately estimate the lower and upper bounds of the regressand. Equation (5) defines the MinMax regression model for the lower and upper bounds of the regressand,

$$\begin{aligned} y_i^- &= \beta_0^- + \sum_{j=1}^{p} \beta_j^- x_{ji}^- + \epsilon_i^- \\ y_i^+ &= \beta_0^+ + \sum_{j=1}^{p} \beta_j^+ x_{ji}^+ + \epsilon_i^+ \end{aligned} \tag{5}$$

where β_j^- and β_j^+ are regression coefficients for the lower and upper bounds respectively ($\beta_j^-, \beta_j^+ \in \mathbb{R}, 0 \leq j \leq p$). In matrix notation, the lower bound model of (5) with all n observations is expressed as $Y^- = X^-\beta^- + \epsilon^-$, where

$$Y^- = (y_1^- \quad y_2^- \quad \dots \quad y_n^-)^T, \quad \beta^- = (\beta_0^- \quad \beta_1^- \quad \dots \quad \beta_p^-)^T,$$

$$\epsilon^- = (\epsilon_1^- \quad \epsilon_2^- \quad \dots \quad \epsilon_n^-)^T, \text{ and } X^- = \begin{pmatrix} 1 & x_{11}^- & \dots & x_{p1}^- \\ 1 & x_{12}^- & \dots & x_{p2}^- \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n}^- & \dots & x_{pn}^- \end{pmatrix}.$$

Like the CM method, (6) estimates the coefficient (β^-) for the lower bound model of (5) using the LS method [30].

$$\hat{\beta}^- = ((X^-)^T X^-)^{-1} (X^-)^T Y^- \tag{6}$$

Now, using the estimated $\hat{\beta}^-$, (7) estimates the lower bounds for the regressand, Y.

$$\hat{y}_i^- = \hat{\beta}_0^- + \sum_{j=1}^{p} \hat{\beta}_j^- x_{ji}^- \tag{7}$$

Similarly, the upper bound model of (5) is expressed as $Y^+ = X^+\beta^+ + \epsilon^+$, where

$$Y^+ = (y_1^+ \quad y_2^+ \quad \dots \quad y_n^+)^T, \quad \beta^+ = (\beta_0^+ \quad \beta_1^+ \quad \dots \quad \beta_p^+)^T,$$

$$\epsilon^+ = (\epsilon_1^+ \quad \epsilon_2^+ \quad \dots \quad \epsilon_n^+)^T, \text{ and } X^+ = \begin{pmatrix} 1 & x_{11}^+ & \dots & x_{p1}^+ \\ 1 & x_{12}^+ & \dots & x_{p2}^+ \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n}^+ & \dots & x_{pn}^+ \end{pmatrix}.$$

Equation (8) estimates the coefficients for the upper bound of (5) by the LS method [30].

$$\hat{\beta}^+ = ((X^+)^T X^+)^{-1} (X^+)^T Y^+ \tag{8}$$

Now, using $\hat{\beta}^+$, (9) estimates the upper bounds of the regressand, Y.

$$\hat{y}_i^+ = \hat{\beta}_0^+ + \sum_{j=1}^{p} \hat{\beta}_j^+ x_{ji}^+ \tag{9}$$

In the MinMax method, the use of two separate models to estimate the lower and upper bounds of the regressand improves the model fitness and interpretation compared to the CM method [17]. However, it still does not guarantee mathematical coherence on the estimated bounds [17]. Further, its estimation performance can be reduced if there is no dependency between the bounds of regressand and regressors [21].

C. The Center and Range Method (2008)

Neto and Carvalho [17] extended the CM method [15] in 2008 to the Center and Range Method (CRM), by considering both interval center and half the 'range' (i.e., the radius) of the regressor and regressand variables. They build two separate regression models—one for the centers and other for the ranges (or radii). Equation (10) is used to model the interval centers.

$$y_i^c = \beta_0^c + \sum_{j=1}^{p} \beta_j^c x_{ji}^c + \epsilon_i^c \tag{10}$$

where y_i^c and x_{ji}^c are the centers $\overline{y}_i$ and $\overline{x}_{ji}$ respectively and ϵ_i^c is error term. The CRM method follows the same matrix format of the CM method to represent the center model of (10) and applies the same strategy of the CM method to estimate the coefficients β_j^c (see (3)). It then estimates the center values for the regressand, Y using the estimated $\hat{\beta}_j^c$ by (11).

$$\hat{y}_i^c = \hat{\beta}_0^c + \sum_{j=1}^{p} \hat{\beta}_j^c x_{ji}^c \tag{11}$$

Similarly, the interval ranges are modelled by the CRM method using (12),

$$y_i^r = \beta_0^r + \sum_{j=1}^{p} \beta_j^r x_{ji}^r + \epsilon_i^r, \tag{12}$$

where y_i^r and x_{ji}^r are the radius of $\overline{y}_i$ and $\overline{x}_{ji}$ respectively. $\beta_j^r \in \mathbb{R}$ are the coefficients for the range estimations, $0 \leq j \leq p$ and ϵ_i^r is error term. Again, following the same matrix notation, the coefficients, β_j^r are estimated by (13),

$$\hat{\beta}^r = ((X^r)^T X^r)^{-1} (X^r)^T Y^r \tag{13}$$

and using the estimated $\hat{\beta}^r$, (14) computes the ranges of the regressand, Y.

$$\hat{y}_i^r = \hat{\beta}_0^r + \sum_{j=1}^{p} \hat{\beta}_j^r x_{ji}^r \tag{14}$$

Finally, the CRM method estimates the bounds of the regressand from the estimated $\hat{y}_i^c$ and $\hat{y}_i^r$ using (15).

$$\hat{y}_i^- = \hat{y}_i^c - \hat{y}_i^r \text{ and } \hat{y}_i^+ = \hat{y}_i^c + \hat{y}_i^r. \tag{15}$$

The CRM model provides better estimation when there is a linear dependency between the ranges of regressand and regressors [18], [21]. If this is not the case, it does not ensure mathematical coherence as (14) can produce negative y_i^r [21].

D. The Constrained Center and Range Method (2010)

Neto and Carvalho [18] later refined the CRM model [17] (see Section II-C) in 2010 to ensure mathematical coherence. The resulting approach is known as the Constrained Center and Range Method (CCRM), where a positivity restriction is enforced on the coefficients which are estimated with respect to the relationship of the radii of regressand and regressor variables. In other words, the overall estimation process in the CCRM method remains the same as for the CRM model (see (10) to (15)) with positivity constraints on the range coefficients β_j^r. The CCRM model uses an iterative algorithm proposed by Lawson and Hanson in [23] to ensure $\beta_j^r \geq 0$. For more detail, we refer the reader to [18]. Finally, the CCRM method estimates the bounds of the regressand using (15).

Neto and Carvalho recommend to apply the CRM model in all cases, only adopting the CCRM method as a suitable strategy when the CRM method fails to maintain mathematical coherence, as the use of the CCRM model can lead to biased estimation outcomes [18]. In particular, the positivity restriction within the CCRM method forces any negative range coefficient to 0 (at the same time updating the remaining range coefficients), in turn leading to potentially biased estimation outcomes and poor model fitness [18].

E. The Complete Information Method (2012)

Wang et al. [19] presented the Complete Information Method (CIM) in 2012 with a focus on considering all the information contained within the intervals. The CIM model considers each interval observation of regressand and regressor variables as a hyper-cube (if a single-valued regressor and associated regressand are considered, each observation effectively reflects a rectangular, as visualized in Fig. 4), and the regression model is built on these hyper-cubes. Further, it adopts Moore's linear combination algorithm [24] to ensure the mathematical consistency of the bounds. In this model, Y is presented by a linear combination of X_j,

$$Y = \beta_0 1_n + \sum_{j=1}^{p} \beta_j X_j + \epsilon_i, \tag{16}$$

where β_j represents the coefficients with $0 \leq j \leq p$ and 1_n is an n-dimensional column constant vector of ones. Equation (17) defines the lower and upper bound estimations of Y.

$$\begin{aligned} y_i^- &= \beta_0 + \sum_{j=1}^{p} \beta_j \left(\tau_j x_{ji}^- + (1-\tau_j) x_{ji}^+\right) \\ y_i^+ &= \beta_0 + \sum_{j=1}^{p} \beta_j \left((1-\tau_j) x_{ji}^- + \tau_j x_{ji}^+\right) \end{aligned} \tag{17}$$

where,

$$\tau_j = \begin{cases} 0, & \text{if } \beta_j \leq 0 \\ 1, & \text{otherwise.} \end{cases}$$

Using (16), the LS method [30] of (17) generates the following linear system,

$$M\beta = b \tag{18}$$

where,

$$M = \begin{pmatrix} \langle 1,1\rangle & \langle 1,X_1\rangle & \dots & \langle 1,X_p\rangle \\ \langle X_1,1\rangle & \langle X_1,X_1\rangle & \dots & \langle X_1,X_p\rangle \\ \vdots & \vdots & \dots & \vdots \\ \langle X_p,1\rangle & \langle X_p,1\rangle & \dots & \langle X_p,X_p\rangle \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \text{ and } b = \begin{pmatrix} \langle 1,Y\rangle \\ \langle X_1,Y\rangle \\ \vdots \\ \langle X_p,Y\rangle \end{pmatrix}.$$

Here, $\langle X_a, X_b\rangle$ is the inner product between X_a and X_b ($0 \leq a,b \leq p$ and $X_0 = 1$) which is computed using (19).

$$\langle X_a, X_b\rangle = \begin{cases} \frac{1}{3}\sum_{i=1}^{n}\left(x_{ai}^{-2} + x_{ai}^- x_{ai}^+ + x_{ai}^{+2}\right), & \text{if } X_a = X_b \\ \frac{1}{4}\sum_{i=1}^{n}\left(x_{ai}^- x_{bi}^- + x_{ai}^- x_{bi}^+ + x_{ai}^+ x_{bi}^- + x_{ai}^+ x_{bi}^+\right), & \text{if } X_a \neq X_b \end{cases} \tag{19}$$

From (18), the coefficients β can be estimated by the inverse of matrix M as defined in (20).

$$\hat{\beta} = (M)^{-1} b \tag{20}$$

Thus, using $\hat{\beta}$ and $\hat{\tau}_j$, the lower and upper bounds of the regressand, Y can be estimated by (17).

The CIM method satisfies mathematical coherence of the estimated bounds of the regressand using τ_j where its value (1 or 0) depends on the sign (+ or -) of $\hat{\beta}_j$ [19].

F. The Linear Model (2015)

Sun and Ralescu [20] developed the Linear Model (LM) in 2015 based on the affine operator in the cone $\mathcal{C} = \{(x,y) \in \mathbb{R}^2 | x \leq y\}$, aiming to maximize model flexibility while preserving its interpretability. This model considers both lower and upper bounds of regressors and their ranges for estimating the bounds of the regressand. In this model, $\overline{y}_i$ is expressed as a linear transformation of $\overline{x}_{ji}$ such that,

$$\begin{aligned} y_i^- &= \sum_{j=1}^{p} \left(\alpha_j x_{ji}^- + \beta_j x_{ji}^+\right) + \eta + \epsilon_i^-, \\ y_i^+ &= \sum_{j=1}^{p} \left((\alpha_j - \gamma_j) x_{ji}^- + (\beta_j + \gamma_j) x_{ji}^+\right) + \eta + \theta + \epsilon_i^+, \end{aligned} \tag{21}$$

where the regressor coefficients, $\alpha_j, \beta_j, \eta \in \mathbb{R}$ and $\gamma_j, \theta \geq 0$, $j = 1, 2, ..., p$. Equation (21) can also be written in the following matrix form as $Y = X\beta + \epsilon$, where

$$Y = \begin{pmatrix} Y^- \\ Y^+ \end{pmatrix} \text{ with } Y^- = \begin{pmatrix} y_1^- \\ \vdots \\ y_n^- \end{pmatrix} \text{ and } Y^+ = \begin{pmatrix} y_1^+ \\ \vdots \\ y_n^+ \end{pmatrix},$$

$$X = \begin{pmatrix} X^* & 0 \\ X^* & X^w \end{pmatrix} \text{ with } X^* = \begin{pmatrix} 1 & x_{11}^- & x_{11}^+ & \dots & x_{p1}^+ \\ 1 & x_{12}^- & x_{12}^+ & \dots & x_{p2}^+ \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n}^- & x_{1n}^+ & \dots & x_{pn}^+ \end{pmatrix}$$

$$\text{and } X^w = \begin{pmatrix} 1 & x_{11}^w & \dots & x_{p1}^w \\ 1 & x_{12}^w & \dots & x_{p2}^w \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n}^w & \dots & x_{pn}^w \end{pmatrix}, \text{ and } \beta = \begin{pmatrix} \beta^* \\ \beta^w \end{pmatrix} \text{ with}$$

$$\beta^* = \begin{pmatrix} \eta \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_p \\ \beta_p \end{pmatrix} \text{ and } \beta^w = \begin{pmatrix} \theta \\ \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix}.$$

Again, the LS estimate of the coefficients, $\hat{\beta}$ are computed by (22),

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{22}$$

and (23) estimates the bounds of the regressand, Y.

$$\begin{aligned} \hat{y}_i^- &= \sum_{j=1}^{p} \left(\hat{\alpha}_j x_{ji}^- + \hat{\beta}_j x_{ji}^+ \right) + \hat{\eta}, \\ \hat{y}_i^+ &= \sum_{j=1}^{p} \left((\hat{\alpha}_j - \hat{\gamma}_j) x_{ji}^- + (\hat{\beta}_j + \hat{\gamma}_j) x_{ji}^+ \right) + \hat{\eta} + \hat{\theta} \end{aligned} \tag{23}$$

Even though the authors assume positive constraints on range coefficients (i.e., $\gamma_j \geq 0$ and $\theta \geq 0$) in (21) to maintain mathematical coherence, the actual model does not ensure compliance with these constraints. As a result, (22) can result in negative range coefficients ($\hat{\gamma}_j, \hat{\theta} < 0$), which may lead to flipped interval bounds ($\hat{y}_i^- > \hat{y}_i^+$) through (23). The authors do not discuss how to maintain these constraints in practice, though they expect that if any of the estimates of $\hat{\gamma}_j$ and $\hat{\theta}$ turns out to be negative, forcing it to be positive may lead to poor fitness of the LM model [20].

G. The Parametrized Model (2017)

Souza et al. [21] proposed the Parametrized Model (PM) in 2017, which also builds two different models for the regressand bounds. This approach extracts the best reference points from the regressors and uses them to build linear regression models for both lower and upper bounds of the regressand, unlike earlier approaches that use specific reference points on the regressors, such as center, range, interval bounds.

In this method, an interval is considered as a line segment, useful to reach all points within it. For instance, given an interval $\overline{a}$, any point $q \in \overline{a}$ can be computed as $q = a^-(1-\lambda) + a^+\lambda$, $0 \leq \lambda \leq 1$. By setting λ, $\overline{a}$ is turned into a single point. Hence, when $\lambda = 0$, $q = a^-$ (lower bound of $\overline{a}$) and when $\lambda = 1$, $q = a^+$ (upper bound of $\overline{a}$). Similarly, $q = a^c$ (center of $\overline{a}$) when $\lambda = 0.5$. Utilizing this concept, the PM method specifies the linear regression models for the lower and upper bounds of Y in (24).

$$\begin{aligned} y_i^- &= \beta_0^- + \sum_{j=1}^{p} \beta_j^-(1-\lambda_j) x_{ji}^- + \beta_j^- \lambda_j x_{ji}^+ + \epsilon_i^-, \\ y_i^+ &= \beta_0^+ + \sum_{j=1}^{p} \beta_j^+(1-\lambda_j) x_{ji}^- + \beta_j^+ \lambda_j x_{ji}^+ + \epsilon_i^+. \end{aligned} \tag{24}$$

Equation (25) simplifies (24) by replacing $\beta_j^-(1-\lambda_j)$ by α_j^- and $\beta_j^-\lambda_j$ by ω_j^- for lower bounds, and $\beta_j^+(1-\lambda_j)$ and $\beta_j^+\lambda_j$ by α_j^+ and ω_j^+ respectively.

$$\begin{aligned} y_i^- &= \beta_0^- + \sum_{j=1}^{p} \alpha_j^- x_{ji}^- + \omega_j^- x_{ji}^+ + \epsilon_i^- \\ y_i^+ &= \beta_0^+ + \sum_{j=1}^{p} \alpha_j^+ x_{ji}^- + \omega_j^+ x_{ji}^+ + \epsilon_i^+. \end{aligned} \tag{25}$$

In matrix notation, the lower bound model can be expressed for all n observations as $Y^- = X^*\beta^- + \epsilon^-$, where

$$Y^- = (y_1^- \quad y_2^- \quad \dots \quad y_n^-)^T, \quad \beta^- = (\beta_0^- \quad \alpha_1^- \quad \omega_1^- \quad \dots \quad \alpha_p^- \quad \omega_p^-)^T,$$
$$\epsilon^- = (\epsilon_1^- \quad \epsilon_2^- \quad \dots \quad \epsilon_n^-)^T, \text{ and } X^* = \begin{pmatrix} 1 & x_{11}^- & x_{11}^+ & \dots & x_{p1}^+ \\ 1 & x_{12}^- & x_{12}^+ & \dots & x_{p2}^+ \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n}^- & x_{1n}^+ & \dots & x_{pn}^+ \end{pmatrix}.$$

The LS estimate of the coefficients for the lower bound model, β^- is computed by (26).

$$\hat{\beta}^- = ((X^*)^T X^*)^{-1} (X^*)^T Y^- \tag{26}$$

The matrix expression follows the same pattern for the upper bound model, $Y^+ = X^*\beta^+ + \epsilon^+$, and the LS estimate of the coefficients for the upper bound model, β^+ in defined in (27).

$$\hat{\beta}^+ = ((X^*)^T X^*)^{-1} (X^*)^T Y^+ \tag{27}$$

Finally, using $\hat{\beta}^-$ and $\hat{\beta}^+$, the lower and upper bounds of Y are estimated using (28).

$$\begin{aligned} \hat{y}_i^- &= \hat{\beta}_0^- + \sum_{j=1}^{p} \hat{\alpha}_j^- x_{ji}^- + \hat{\omega}_j^- x_{ji}^+ \\ \hat{y}_i^+ &= \hat{\beta}_0^+ + \sum_{j=1}^{p} \hat{\alpha}_j^+ x_{ji}^- + \hat{\omega}_j^+ x_{ji}^+. \end{aligned} \tag{28}$$

The PM method does not automatically guarantee the mathematical coherence of the bounds [21]. To avoid flipping the interval bounds, the approach estimates the range of Y using (29) before performing the regression.

$$\hat{Y}^w = X^*((X^*)^T X^*)^{-1} (X^*)^T Y^w \tag{29}$$

If all estimated ranges are positive ($\hat{y}^w \in \hat{Y}^w$), the model automatically ensures mathematical coherence. However, if at least one of the estimated ranges is negative, it applies the Box-Cox transformation [25], extended to intervals by the authors [21] (defined in (30)), to transform the regressand so that the desirable coherence is achieved by the PM method.

$$\overline{y}_i^k = \begin{cases} \left[\frac{(y_i^- + k_2)^{k_1} - 1}{k_1}, \frac{(y_i^+ + k_2)^{k_1} - 1}{k_1} \right], & \text{if } k_1 \neq 0. \\ \left[log(y_i^- + k_2), log(y_i^+ + k_2) \right], & \text{if } k_1 = 0. \end{cases} \tag{30}$$

where $\overline{y}_i^k$ is the transformed interval of $\overline{y}_i$. k_1 is any real value and k_2 maintains the following restriction: $y_i^- + k_2 > 0$.

III. Data and Methodology

In Section II, we discussed linear regression models for intervals considering their vector/matrix representation. Table II provides a summary of these models and their features. In this section, we initially put forward an extension to the LM method [20] which ensures mathematical coherence, thus making it more suitable for real-world data. Next, we introduce

TABLE II: Summary of well-known linear regression models for IV data

Methods	Interval reference point(s) used in the regression process	Also applicable to numeric data (interval range = 0)	Ensure mathematical coherence on interval bounds
CM [15]	Center	Yes	No
MinMax [16]	Lower and upper bounds	Yes	No
CRM [17]	Center and radius (half of range)	No	No
CCRM [18]	Center and radius (half of range)	No	Yes with positivity restrictions on range coefficients
CIM [19]	All points within intervals	Yes	Yes with Moore's linear combination algorithm [24]
LM [20]	Lower and upper bounds, and range	No	Yes with positivity restrictions as proposed in this paper
PM [21]	Any point within intervals	Yes	Yes with the interval Box-Cox transformation [25]

a novel approach to intuitively visualizing IV regression—the Interval Regression Graph (*IRG*). *IRG*s are designed to facilitate the interpretation and communication of the regression model and thus the relationship between an independent and dependent variable of interest. The latter is a crucial asset of traditional (numeric) regression models, accounting for a substantial part of their utility and popularity. Going beyond the traditional visualization of numeric regression, *IRG*s communicate the relationships of both position and range (uncertainty) of IV regressor and regressand. Following the introduction of *IRG*s, we introduce both synthetic and real-world data sets to support the empirical exploration and analysis of the regression approaches in Section II. Finally, we discuss evaluation metrics employed for performance assessment during the remainder of the paper.

A. Extension of the Linear Model [20]

As pointed out in Section II, the LM method [20] considers positivity restrictions on the range coefficients to consistently estimate the regressand bounds. However, it neither guarantees these restrictions through its model setting, nor provides algorithmic solutions to maintaining them. Thus, to make the model suitable for practical real-world deployment, we put forward two alternative extensions to the LM method: (1) universally forcing positive restrictions on parameters as [20] suggested—that is proposing a constrained LM method, and (2) enforcing restrictions only when needed to avoid unnecessary estimation bias [20]—that is, proposing a weakly constrained LM method.

Algorithm 1 A constrained LM method (LM_c)

Input: Let $Y = \{\overline{y}_1, \overline{y}_2, ..., \overline{y}_n\}$ be the regressand variable's samples and $\{X_1, X_2, ..., X_p\}$ are the p (≥ 1) be the regressor variables.

Output: Let $\hat{Y} = \{\hat{\overline{y}}_1, \hat{\overline{y}}_2, ..., \hat{\overline{y}}_n\}$ be the estimated Y.

1: Apply the LM method at (22) on Y and X_j $(1 \leq j \leq p)$ to estimate $\hat{\beta}^*$ and $\hat{\beta}^w$
2: **if** $\hat{\theta} < 0$ or $\hat{\gamma}_j < 0$ where $\hat{\theta}, \hat{\gamma}_j \in \hat{\beta}^w$, **then**
3: Apply LHA [23] to make $\hat{\theta}$ and $\hat{\gamma}_j$ positive
4: **end if**
5: Estimate $\hat{Y}^-$ and $\hat{Y}^+$ using (23)
6: **return** $\hat{Y} = [\hat{Y}^-, \hat{Y}^+]$

(1) We propose a constrained LM method (LM_c) following the notion of the CCRM method [18] where the Lawson and Hanson algorithm (LHA) [23] is applied to universally impose the positivity restriction on the range coefficients, γ_j and θ. Algorithm 1 presents this constrained LM variant, LM_c. Note that universally (or blindly) applying the restrictions for maintaining mathematical coherence risks poor performance in terms of model fit [20].

(2) To minimize the impact on performance, we also develop a more weakly constrained LM method (LM_w), in line with the rationale of the PM approach [21] to avoid unnecessary bias in the estimation results due to universally restricting parameters. Here, we first regress the range of the regressand variable (y_i^w) on the range of regressors (x_{ji}^w) before performing the estimation using (29) as the estimated range values of the regressand guarantee mathematical coherence of the estimated bounds *if* they are positive [21]. Therefore, if all estimated range values are positive ($\hat{y}_i^w \geq 0, \forall i$), we compute the least squares estimation of (22) without any parameter restrictions. In contrast, if any of the estimated range values are negative ($\hat{y}_i^w < 0, \exists i$), we enforce positivity restrictions on $\hat{\gamma}_j$ and $\hat{\theta}$ using the Lawson and Hanson algorithm (LHA) [18], [23]. *Algorithm 2 summarizes this adapted approach, LM_w.*

Algorithm 2 A weakly-constrained LM method (LM_w)

Input: Let $Y = \{\overline{y}_1, \overline{y}_2, ..., \overline{y}_n\}$ be the regressand variable's samples and $\{X_1, X_2, ..., X_p\}$ are the p (≥ 1) regressor variables.

Output: Let $\hat{Y} = \{\hat{\overline{y}}_1, \hat{\overline{y}}_2, ..., \hat{\overline{y}}_n\}$ be the estimated Y.

1: Estimate the range values of regressand variable, $\hat{Y}^w = \{\hat{y}_1^w, ..., \hat{y}_n^w\}$ using (29)
2: **if** all $\hat{y}_i^w \geq 0$ **then**
3: Apply the LM method at (22) on Y and X_j $(1 \leq j \leq p)$ to generate $\hat{\beta}^*$ and $\hat{\beta}^w$
4: Estimate $\hat{Y}^-$ and $\hat{Y}^+$ using (23)
5: **else**
6: Apply **Algorithm 1** to estimate $\hat{Y}^-$ and $\hat{Y}^+$
7: **end if**
8: **return** $\hat{Y} = [\hat{Y}^-, \hat{Y}^+]$

B. The Interval Regression Graph (IRG)

Visualization of traditional regression models provides a powerful mechanism for interpretability and communication of the relationship between numeric variables. While intervals provide a richer model, enabling the capture of uncertain information, the crucial interpretation and communication of any insights is complex, yet just as vital as for the numeric case. We introduce a novel 3D visualization approach, the interval regression graph (*IRG*) which succinctly and clearly captures the change in a regressand's key features (center and

Algorithm 3 Interval Regression Graph (*IRG*) Generation

Input: An IV regression model (IVRM). The IV regressor's (e.g., from the original data set) minimum range and maximum range are range$X_{\min}$ and range$X_{\max}$, as well as its minimum center and maximum center coordinates are center$X_{\min}$ and center$X_{\max}$.

Output: Two *IRG* plots mapping the regressor to the regressand's center, $\hat{Y}^c$ and range, $\hat{Y}^w$.

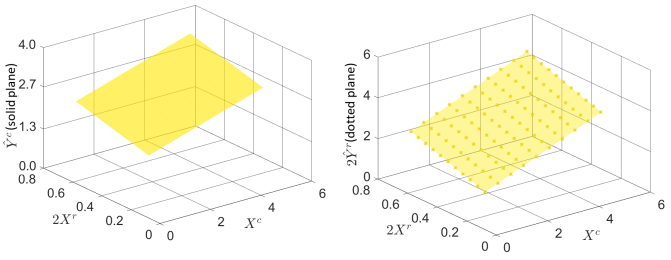
1: Generate the set $\boldsymbol{X^w}$ of p discretizations of the interval $[\text{range}X_{\min}, \text{range}X_{\max}]$[a]
2: Generate the set $\boldsymbol{X^c}$ of q discretizations of the interval $[\text{center}X_{\min}, \text{center}X_{\max}]$
3: **for** each discretized $X_i^w, 1 \leq i \leq p$ **do**
4: **for** each discretized $X_j^c, 1 \leq j \leq q$ **do**
5: Compute $X_{ij}^- = X_j^c - \frac{X_i^w}{2}$
6: Compute $X_{ij}^+ = X_j^c + \frac{X_i^w}{2}$
7: Compute $\hat{Y}_{ij}^-$, $\hat{Y}_{ij}^+$ with X_{ij}^-, X_{ij}^+ using IVRM
8: Compute $\hat{Y}_{ij}^c = \frac{\hat{Y}_{ij}^- + \hat{Y}_{ij}^+}{2}$
9: Compute $\hat{Y}_{ij}^w = \hat{Y}_{ij}^+ - \hat{Y}_{ij}^-$
10: **end for**
11: **end for**
12: Generate $\hat{Y}^c$ *IRG* plot with $\hat{\boldsymbol{Y}}^{\boldsymbol{c}}$ on the vertical axis, $\boldsymbol{X^w}$ on the bottom left axis and $\boldsymbol{X^c}$ on the bottom right axis
13: Generate $\hat{Y}^w$ *IRG* plot with $\hat{\boldsymbol{Y}}^{\boldsymbol{w}}$ on the vertical axis, $\boldsymbol{X^w}$ on the bottom left axis and $\boldsymbol{X^c}$ on the bottom right axis
14: **return** *IRG* plots for $\hat{Y}^c$ and $\hat{Y}^w$

[a]Note that the *IRG* generation algorithm is designed to allow for the visualization of both linear and non-linear relationships between variables. In this paper, we focus on linear relationships/regression models only.

range) for given changes in a regressor's key features (center and range). The *IRG* visualizes the correlation between a regressor's and regressand's position (or range), based on a given regression model [31]. Note that in this paper, for conciseness, we only introduce and focus on *IRG*s generated for visualizing the relationship between one regressor and one regressand. Algorithm 3 presents the pseudocode for constructing *IRG*s for a regressand's center and range with respect to a regressor's center and range—for a given regression model.

To illustrate the *IRG* and its utility, consider the IV Set-1 (Fig. 1(a)), presented in the introductory section. Figures 3(a) and (b) individually show the two different aspects of the *IRG* for Set-1 based on the PM regression method. The bottom-left and bottom-right axes show the regressor's center (X^c) and range ($X^w \simeq 2\hat{X}^r$) respectively. In Fig. 3(a), the vertical axis denotes the regressand's estimated center ($\hat{Y}^c$ (solid plane)), while in Fig. 3(b), it reflects the regressand's estimated range ($\hat{Y}^w \simeq 2\hat{Y}^r$(dotted plane)).

Interpreting these figures, we can see how Fig. 3(a) visualizes that the regressand's center, $\hat{Y}^c$ increases with respect to increasing values of both the regressor's range, $2X^r$ (or X^w) and its center, X^c. Fig. 3(b) shows how the regressand's range, $2\hat{Y}^r$ also increases with respect to both increasing values of the regressor's range, $2X^r$ and center, X^c, and that it does so at a greater rate in each case than does $\hat{Y}^c$. For



(a) *IRG* as to $\hat{Y}^c$ (b) *IRG* as to $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$)

Fig. 3: Relationship between regressand and regressor with respect to center (position) (a) and range (b) using the PM method for Set-1 (Fig. 1(a)).

compact and complete visualization, *IRGs* commonly combine the visualizations with respect to both the regressand's center and range as shown in Fig. 18(f). We will leverage *IRGs* throughout the remainder with additional examples.

C. Data Sets

We use both synthetic and real-world IV data sets to investigate the performance of linear regression methods. The synthetic data sets are used to explore and visualize the sensitivity of the regression models with respect to different properties of the interval data, whereas the real-world data sets are used to investigate their suitability to real-world scenarios. All data sets are described below.

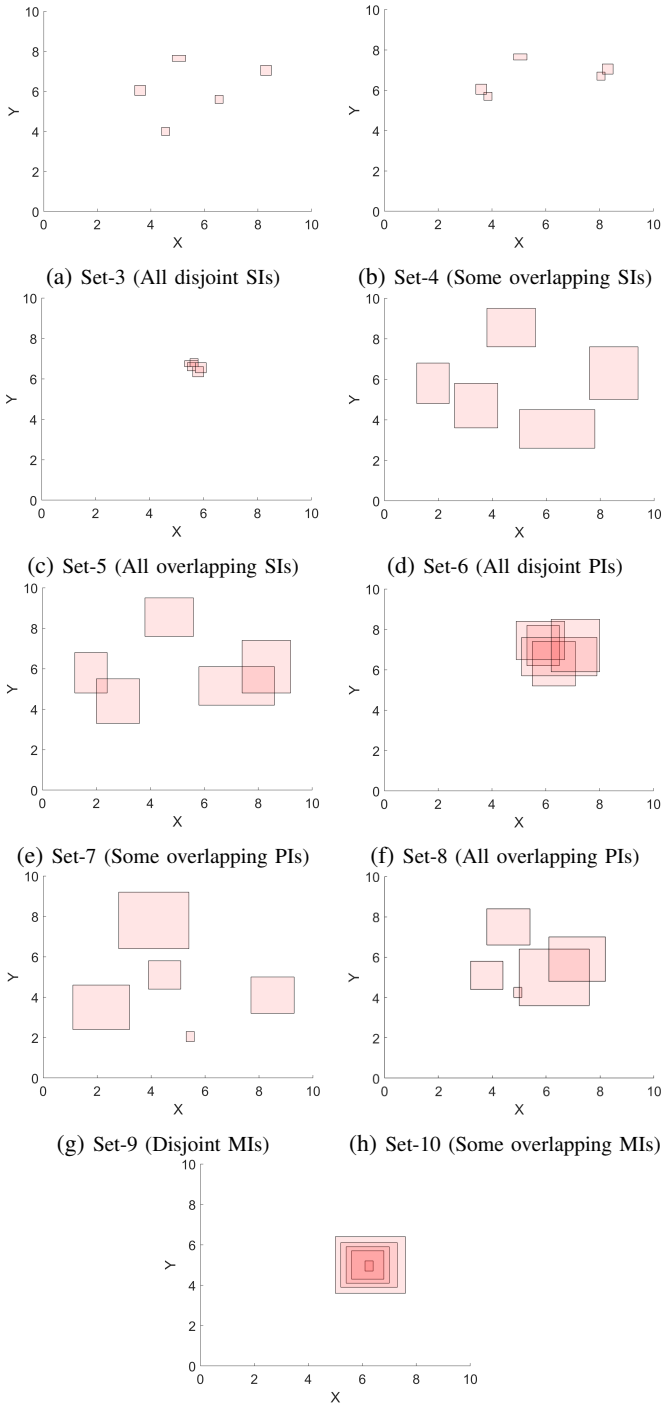
1) Synthetic Data Sets: We specifically design a series of synthetic IV data sets to investigate the behavior of the regression models. First, Set-1 and Set-2, shown at the beginning of the paper (Figs. 1(a) and 2(a)), are used to explore the relationship between regressor and regressand with respect to the primary features (i.e., center and range). Second, an additional nine sets of intervals (Set-3 to Set-11), exhibiting different properties are used: we use three sets of *skinny* intervals (Set-3 to Set-5) with smaller range[3] and three *puffy* sets with wider range (Set-6 to Set-8) having features of disjointedness, some, and full overlap. We also consider three mixed cases (Set-9 to Set-11), containing puffy and skinny intervals in separated, partially overlapping and nested states. All of these data sets are produced within the range of 0 to 10. They are given in Table VII (a) to (k) in the Appendix including the mean (μ) and standard deviation (σ) of the center and range of intervals. Figure 4 graphically presents the data sets, Set-3 to Set-11.

2) Real-world Data Sets: We focus on three real-world IV data sets with different properties and arising in different application contexts, i.e., medical, consumer food ratings/marketing, and cyber-security, reflecting the broad applicability of IV data in real-world applications.

The *first* data set represents medical observations and includes interval-valued *systolic* and *diastolic* blood pressure (BP) data of 59 patients of the Hospital Valle del Nalón in Asturias, Spain [11], [14].

The *second* data set presents IV consumer ratings of eight different (UK market) snack-food products [32]. In this case,

[3]The 'skinny' and 'puffy' designs have been inspired by Ferson et al. [26].



(a) Set-3 (All disjoint SIs) (b) Set-4 (Some overlapping SIs)

(c) Set-5 (All overlapping SIs) (d) Set-6 (All disjoint PIs)

(e) Set-7 (Some overlapping PIs) (f) Set-8 (All overlapping PIs)

(g) Set-9 (Disjoint MIs) (h) Set-10 (Some overlapping MIs)

(i) Set-11 (Nested/all overlapping MIs)

Fig. 4: 2D visualization of synthetic data sets, Set-3 to Set-11. SI=skinny interval, PI=puffy interval, and MI=mixed interval.

40 consumers rated each product—using 'DECSYS' interval survey software [33]—on six different attributes: *visual appeal*, *value for money*, *healthiness*, *taste*, *branding*, and *ethics*, along with their *overall purchase intention* (i.e., how likely they were to buy each product).

The *third* and final data set captures assessments by cyber-security experts of the vulnerability of system components in a real-world scenario. Specifically, 38 cyber-security experts at the UK CESG (Communications-Electronics Security Group) assessed a range of components (referred to as hops) that are commonly encountered during a cyber-attack [34], [35]. They rated the hops on overall difficulty for an attacker to either attack or evade them, as well as rating each on several attributes (seven for attack, and three for evade) that might affect this difficulty. The experts provided their IV ratings on a scale from 0 to 100. For more details of these data sets, see [11], [14], [32], [34], [35].

D. Evaluation Metrics

Evaluation poses one of the key challenges for IV data analysis and AI more broadly. While 'quality of fit' for numeric data is straightforward, the same is not the case for intervals. Comparing the latter requires application-led assumptions (on the nature of the intervals and whether for example interval size is more important than position), which in turn feed into similarity measures or arithmetic means of comparison. We do not delve further into this challenge in this paper, instead following three fairly standard metrics to evaluate the performance of the linear regression models. Each of the evaluation metrics is defined with respect to the regressand (samples), $Y = \{\overline{y}_1, \overline{y}_2, ..., \overline{y}_n\}$ and its estimated value(s), $\hat{Y} = \{\hat{\overline{y}}_1, \hat{\overline{y}}_2, ..., \hat{\overline{y}}_n\}$.

1) Root Mean Squared Error (RMSE) estimates error by considering the average deviation of the estimated values from the actual ones [18]. The *RMSE* for the lower and upper bounds of Y is defined as,

$$RMSE^- = \sqrt{\tfrac{1}{n}\sum_{i=1}^{n}(y_i^- - \hat{y}_i^-)^2} \text{ and } RMSE^+ = \sqrt{\tfrac{1}{n}\sum_{i=1}^{n}(y_i^+ - \hat{y}_i^+)^2},$$

where a smaller *RMSE* value indicates a better fitted model.

2) Mean Absolute Error (MAE) calculates the average of absolute difference between the actual and estimated values [36]. For the upper and lower bounds of Y, the *MAE* is defined as,

$$MAE^- = \tfrac{1}{n}\sum_{i=1}^{n}\left|y_i^- - \hat{y}_i^-\right| \text{ and } MAE^+ = \tfrac{1}{n}\sum_{i=1}^{n}\left|y_i^+ - \hat{y}_i^+\right|,$$

where $0 \leq MAE^-, MAE^+ \leq \mathbb{R}^+$. Again, a lower value of *MAE* implies a better fitted model.

3) Mean Magnitude of Relative Error (MMRE) measures the mean of dispersion of the actual and estimated values with respect to the actual values [37]. It is calculated as follows:

$$MMRE = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\left\{\left|\frac{y_i^- - \hat{y}_i^-}{y_i^-}\right| + \left|\frac{y_i^+ - \hat{y}_i^+}{y_i^+}\right|\right\},$$

where $0 \leq MMRE \leq \mathbb{R}^+$. In this case, the lower the *MMRE*, the better the model.

IV. Results

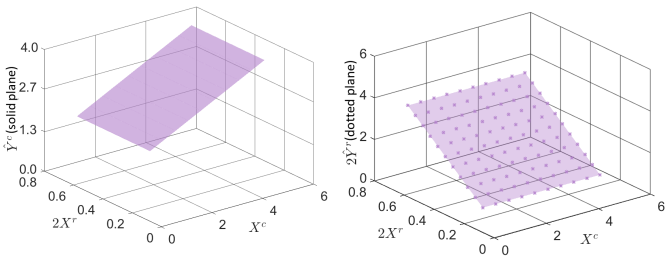
We explore the performance of existing regression approaches together with two proposed variants of the LM method (LM_c and LM_w) for synthetic and real-world IV data sets (see Section III-C). Their performance is assessed using different metrics described in Section III-D. All these regression methods are implemented in Matlab and the full source code will be available at http://lucidresearch.org/software with the publication of the paper.

A. Performance Analysis with Synthetic Data Sets

In Section III-B, we briefly introduced the *IRG*s using IV Set-1 (Fig. 1(a)). We expand on this example here, reporting performance of all regression models covered above with this set, as well as for Set-2 (Fig. 2(a)).

As discussed earlier, Set-1 shows the more commonly considered and intuitive scenario where uncertainty and position in regressor and regressand vary in unison. On the other hand, Set-2 presents a less commonly considered scenario, where the uncertainty of the regressand increases irrespective of the uncertainty of the regressor. Said differently, for Set-2, uncertainty and position of the regressand vary quasi exclusively with respect to the *position* of the regressor variable.

Tables III and IV present the fitness results for all regression approaches for Sets 1 and 2, respectively. These show that the PM method provides the best fitness in each case, though the MinMax method provides almost equal fitness for Set-2 (they do differ with respect to MAE^{+}). For conciseness, we focus further discussion of the regression results on two of the regression approaches: the CRM—representing an earlier and more basic approach which does not guarantee mathematical coherence, and the PM—the most recent approach, which does. *IRG*s for all methods are shown in the appendix (Figs. 18 and 19). We independently present the relationship of the center and range of the test regressor on the center (subfigures (a)) and the range (subfigures (b)) of the regressand, for the CRM and PM methods in Figs. 3, 5, 6, and 7. Note that, Figs. 3(b) and 6(b) are repeated for clarity from Figs. 1(b) and 2(b) in Section I. The *IRG*s in Figs. 3(a) and (b) show that, according to the PM method, both changes in the estimated center and range of Set-1's regressand are associated with the changes in the center and range of the regressor—in-line with our expectation. On the other hand, the *IRG*s in Figs. 5(a) and (b) show that for the CRM method, the estimated center of Set-1's regressand is influenced only by the center of the regressor, and the estimated range of Set-1 is influenced only by the range of the regressor. Thus, the CRM method is not in line with our expectation for the Set-1 data set. The rest of the methods also do not perform according to the expectation (see Appendix, Fig. 18).



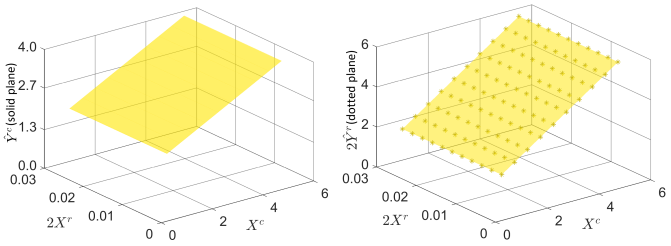
(a) *IRG* as to $\hat{Y}^c$ (b) *IRG* as to $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$)

Fig. 5: Relationship between regressand and regressor with respect to center (position) (a) and range (b) using the CRM method for Set-1 (Fig. 1(a)).

TABLE III: Regression performance for Set-1

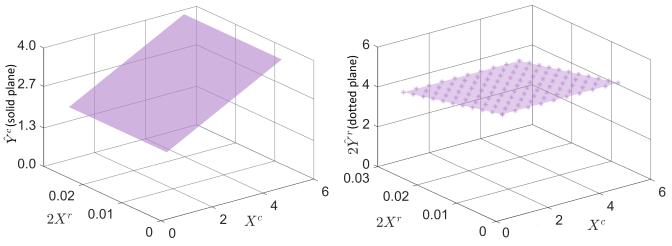
Methods	$RMSE^{-}$	$RMSE^{+}$	MAE^{-}	MAE^{+}	$MMRE$
CM	1.506	1.510	1.354	1.354	0.816
MinMax	0.002	0.113	0.001	0.109	0.017
CRM	0.160	0.055	0.140	0.047	0.075
CCRM	0.251	0.227	0.218	0.198	0.140
CIM	1.503	1.517	1.356	1.356	0.817
PM	**0.0**	**0.040**	**0.0**	**0.031**	**0.006**
LM_c	0.097	0.428	0.085	0.381	0.104
LM_w	0.097	0.105	0.085	0.088	0.053

Bold = Best



(a) *IRG* as to $\hat{Y}^c$ (b) *IRG* as to $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$)

Fig. 6: Relationship between regressand and regressor with respect to center (position) (a) and range (b) using the PM method for Set-2 (Fig. 2(a)).



(a) *IRG* as to $\hat{Y}^c$ (b) *IRG* as to $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$)

Fig. 7: Relationship between regressand and regressor with respect to their center (position) (a) and range (b) using the CRM method for Set-2 (Fig. 2(a)).

TABLE IV: Regression performance for Set-2

Methods	$RMSE^{-}$	$RMSE^{+}$	MAE^{-}	MAE^{+}	$MMRE$
CM	1.650	1.657	1.491	1.491	0.90
MinMax	**0.002**	**0.153**	**0.001**	0.148	**0.023**
CRM	0.575	0.592	0.53	0.535	0.344
CCRM	0.703	0.719	0.61	0.625	0.398
CIM	1.65	1.657	1.491	1.491	0.90
PM	**0.002**	**0.153**	**0.001**	**0.145**	**0.023**
LM_c	0.573	1.007	0.528	0.930	0.398
LM_w	0.573	0.593	0.528	0.537	0.343

Bold = Best

Further, the *IRG*s in Figs. 6(a) and (b) show how, according to the PM model, both center and range of the regressand of Set-2 vary solely with respect to the center of the regressor—in line with expectation. The *IRG* in Fig. 7(a) visualizes how, using the CRM method, the estimated center of Set-2's regressand is similarly positively influenced by the changes in the center, but not the range of the regressor—as expected. However, in contrast to the PM model outputs, the CRM-derived *IRG* for the same data, shown in Fig. 7(b), indicates no strong relationship between regressand range and either facet of the regressor. If any relationship is present at all, then this is a negative effect of the range of the regressor on that of the regressand—these results do not fit those expected when viewing the data. Fig. 19 (in the Appendix) visualizes that the MinMax method shows similar performance to the PM method. The remaining methods perform less well.

Next, we will briefly discuss equivalent analysis and model visualizations, using the *IRG*s, for an additional nine synthetic data sets (Set-3—Set-11), each designed to reflect a key

This article has been accepted for publication in IEEE Transactions on Artificial Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAI.2023.3234930

TABLE V: Performance of different regression models for IV synthetic data sets (Set-3 to Set-11)

(a) Set-3 (All disjoint skinny intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.215	1.181	0.972	1.035	0.185
MinMax	1.204	1.169	1.004	1.002	0.185
CRM	1.185	1.189	0.986	1.021	0.185
CCRM	1.195	1.179	0.995	1.011	0.185
CIM	1.215	1.181	0.972	1.035	0.185
PM	**0.343**	**0.355**	**0.291**	0.312	**0.059**
LM_c	**0.343**	0.356	0.293	0.317	**0.059**
LM_w	**0.343**	**0.355**	0.293	**0.310**	**0.059**

(b) Set-4 (Some overlapping skinny intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	0.656	0.601	0.592	0.407	0.074
MinMax	0.631	0.574	0.483	0.432	0.068
CRM	0.616	0.591	0.475	0.451	0.069
CCRM	0.626	0.581	0.482	0.443	0.069
CIM	0.656	0.601	0.592	0.407	0.074
PM	**0.263**	**0.192**	**0.249**	**0.181**	**0.032**
LM_c	**0.263**	0.193	**0.249**	0.185	0.033
LM_w	**0.263**	**0.192**	**0.249**	**0.181**	**0.032**

(c) Set-5 (All overlapping skinny intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	0.370	0.362	0.341	0.341	0.052
MinMax	0.141	**0.115**	**0.120**	**0.107**	**0.017**
CRM	0.143	0.120	0.125	0.108	0.018
CCRM	0.143	0.121	0.128	0.111	0.018
CIM	0.204	0.177	0.159	0.149	0.024
PM	**0.139**	**0.115**	0.128	**0.107**	0.018
LM_c	0.143	0.122	0.126	0.113	0.018
LM_w	0.143	0.120	0.126	0.109	0.018

(d) Set-6 (All disjoint puffy intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.996	1.999	1.796	1.568	0.350
MinMax	1.681	1.685	1.299	1.396	0.268
CRM	1.692	1.673	1.314	1.389	0.270
CCRM	1.679	1.687	1.297	1.391	0.268
CIM	1.976	1.986	1.777	1.564	0.347
PM	**1.456**	**1.344**	**1.143**	1.061	0.188
LM_c	1.460	1.350	1.147	1.059	0.188
LM_w	1.460	1.349	1.147	**1.057**	**0.187**

(e) Set-7 (Some overlapping puffy intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.751	1.697	1.673	1.220	0.266
MinMax	1.434	1.367	1.070	1.034	0.180
CRM	1.442	1.357	1.1067	1.036	0.184
CCRM	1.437	1.364	1.090	1.038	0.183
CIM	1.753	1.698	1.674	1.220	0.267
PM	**1.411**	**1.283**	**1.067**	0.984	0.173
LM_c	1.415	1.289	1.071	0.983	0.173
LM_w	1.415	1.288	1.071	**0.981**	0.172

(f) Set-8 (All overlapping puffy intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.308	1.315	1.238	1.238	0.184
MinMax	0.416	0.437	0.387	0.398	0.059
CRM	0.416	0.445	0.365	0.387	0.056
CCRM	0.410	0.452	0.348	0.404	0.056
CIM	1.073	1.084	0.986	0.986	0.147
PM	**0.396**	**0.406**	0.321	**0.328**	**0.049**
LM_c	0.415	0.429	**0.313**	0.369	0.051
LM_w	0.415	0.425	**0.313**	0.348	**0.049**

(g) Set-9 (Disjoint mixed intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.871	2.449	1.616	1.882	0.509
MinMax	1.596	2.236	1.349	1.686	0.430
CRM	1.886	1.911	1.497	1.526	0.433
CCRM	1.886	1.911	1.497	1.526	0.433
CIM	1.828	2.298	1.550	1.796	0.487
PM	**1.210**	**1.238**	**0.065**	1.089	0.251
LM_c	**1.210**	**1.238**	1.066	**1.087**	**0.250**
LM_w	**1.210**	**1.238**	1.066	**1.087**	**0.250**

(h) Set-10 (Some overlapping mixed intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.361	1.572	1.293	1.307	0.152
MinMax	0.994	1.235	0.906	0.891	0.105
CRM	1.107	1.105	0.938	0.918	**0.10**
CCRM	1.107	1.105	0.938	0.918	**0.10**
CIM	1.344	1.549	1.278	1.281	0.165
PM	**0.988**	**0.977**	**0.883**	0.869	**0.098**
LM_c	**0.988**	**0.977**	0.886	**0.866**	**0.098**
LM_w	**0.988**	**0.977**	0.886	**0.866**	**0.098**

(i) Set-11 (Nested mixed intervals)

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.008	1.022	0.926	0.926	0.194
MinMax	0.048	0.048	0.046	0.045	0.010
CRM	0.027	0.028	0.019	0.026	**0.004**
CCRM	0.027	0.028	0.019	0.026	**0.004**
CIM	0.945	0.958	0.870	0.870	0.183
PM	**0.018**	**0.020**	0.014	**0.016**	**0.003**
LM_c	**0.018**	**0.020**	**0.013**	**0.016**	**0.003**
LM_w	**0.018**	**0.020**	**0.013**	**0.016**	**0.003**

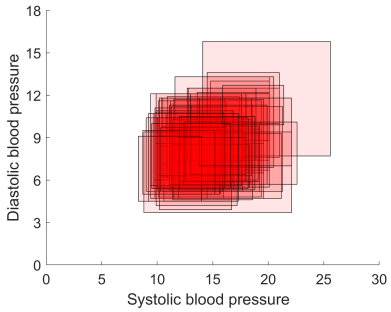
Bold = Best

property regularly encountered in IV data sets. This will highlight how different data sets—with different features—are differently suitable for different regression methods, as later results will expand upon in more detail. For data Set-3 to Set-11, all shown in Fig. 4, we briefly summarize individual performance results across methods below before discussing performance more broadly. Table V(a) – (i) present the fitness of regression methods for each of these data sets.

- *Set-3 (All disjoint skinny intervals)*: The PM and LM_w methods produce the best fit for all disjoint skinny intervals. The LM_c method results in equally good fit well with respect to two metrics ($RMSE^-$ and $MMRE$).
- *Set-4 (Some overlapping skinny intervals)*: Again, the PM and LM_w methods produce the best fit. The LM_c method also gives the same fitness with respect to three metrics ($RMSE^-$, MAE^-, and $MMRE$). The fitness of other methods improves with some overlapping of intervals.
- *Set-5 (All overlapping skinny intervals)*: The PM and MinMax methods come out as the best approach. Among others, CRM, CCRM, LM_c and LM_w methods produce quite similar results with respect to some of the evaluation metrics.
- *Set-6 (All disjoint puffy intervals)*: The PM method gives the best fit, though the LM_c and LM_w methods also produce quite similar results for puffy intervals.
- *Set-7 (Some overlapping puffy intervals)*: The PM and LM_w emerge as the best approaches with respect to different metrics, though the MinMax,CRM, CCRM and LM_c methods also produce similarly good fit in terms of $MMRE$.
- *Set-8 (All overlapping puffy intervals)*: For all overlapping puffy intervals, the PM method gives the best fit with respect to three metrics and the LM_w method is for two evaluation metrics. Except CM and CIM methods, growing overlap among intervals appears to lower estimation errors by the CRM, CCRM and MinMax methods.
- *Set-9 (Disjoint mixed intervals)*: The LM_w, LM_c and PM methods perform best. All other methods result in relatively low fitness. In particular, the CM and CIM methods achieve the worst results.
- *Set-10 (Some overlapping mixed intervals)*: The PM, LM_c and LM_w methods result in similar model fit for almost all metrics. The CRM and CCRM methods give better fitness results with respect to one metric ($MMRE$).
- *Set-11 (Nested/all overlapping mixed intervals)*: The LM_c, LM_w, and PM methods give the best fit. The CRM and CCRM methods also closely fit the data as to $MMRE$.

Overall, the PM, and LM_w (and often LM_c) methods perform best and result in reliably good—and almost identical—fitness for disjoint skinny or puffy intervals. With growing overlap among intervals (skinny or puffy), model fitness by the CRM, CCRM and MinMax methods generally improves. Nevertheless, they consistently fail to reach the performance levels of the PM and LM methods. For the mixed sets with puffy and skinny intervals, where these are disjoint or exhibit some overlap, both PM and both variants of the LM method result in similarly good (best) fit. Others, particularly the CRM and CCRM approaches, give nearly identical results for some metrics, for all overlapping, respectively nested, mixed cases. In all cases, the estimates by the CIM and CM methods deviate substantially from the actual values and hence they come out as poor performers.

As well as the empirical analyses reported above, we



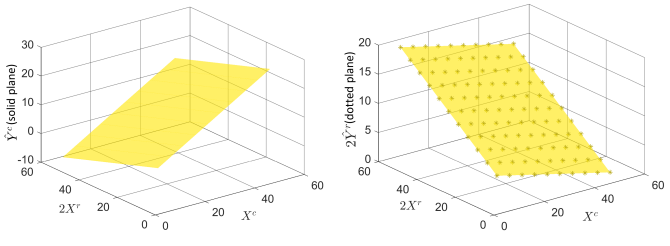
(a) Blood pressure data set

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	1.259 [0.937 1.806]	1.617 [1.302 2.039]	1.097 [0.772 1.658]	1.395 [1.087 1.756]	0.167 [0.121 0.238]
MinMax	0.793 [0.631 0.905]	1.275 [1.046 1.438]	0.651 [0.499 0.758]	1.058 [0.846 1.216]	0.106 [0.086 0.120]
CRM	0.881 [0.729 1.151]	1.256 [1.006 1.377]	0.653 [0.525 0.888]	1.029 [0.814 1.173]	0.105 [0.088 0.129]
CCRM	0.881 [0.728 1.148]	1.256 [1.008 1.377]	0.653 [0.525 0.884]	1.029 [0.815 1.174]	0.105 [0.088 0.129]
CIM	2.104 [1.869 2.386]	2.249 [1.986 2.522]	1.951 [1.712 2.240]	1.907 [1.681 2.199]	0.268 [0.232 0.314]
PM	**0.780** [0.616 0.887]	**1.208** [0.977 1.356]	0.638 [0.483 0.739]	**0.999** [0.799 1.149]	**0.102** [0.083 0.115]
LM_c	0.784 [0.620 0.897]	1.211 [1.310 3.232]	**0.631** [0.480 0.736]	1.008 [1.006 2.917]	**0.102** [0.10 0.194]
LM_w	0.784 [0.620 0.897]	1.211 [1.307 3.232]	**0.631** [0.480 0.736]	1.008 [1.004 2.917]	**0.102** [0.10 0.194]

Bold = Best; Bootstrapped 95% confidence intervals at 10000 simulations are provided.

(b) Model Performance

Fig. 8: Graphical presentation of *systolic* and *diastolic* BP data set [11] and performance of different regression models as to evaluation metrics using this set.



(a) with respect to $\hat{Y}^c$ (b) with respect to $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$)

Fig. 9: *IRG*s showing the relationship between *systolic* and *diastolic* BP in terms of center and range for the PM method.

carried out additional experiments going beyond the scope of this paper—exploring the impact of differences in mean interval size/range, differences in the standard deviation of the mean size/range, and the variation in interval centers (i.e., dispersion), on model fitness or estimation errors [38]. Results suggest that estimation errors tend to grow for all regression methods with respect to both higher standard deviation of ranges and more distant placement of intervals (centers) [38]. We will review this aspect with real-world data in future.

B. Real-world Data Sets

In this section, three real-world IV data sets from the application contexts described in Section III-C, are used to explore and evaluate the behaviour of different regression methods with respect to the mixed set of features of the data (in the sense of the previous section) present in these sets. The fitness of each method is compared using the same evaluation metrics (i.e., *RMSE*, *MAE*, *MMRE*) as before.

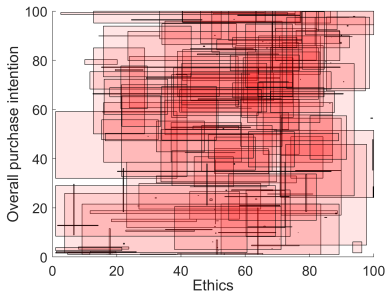
In addition, the larger degrees of freedom—compared to the synthetic data—of these real world data sets, permit us to empirically determine a measure of the variability associated with each performance value, utilizing a non-parametric resampling approach [39]. We therefore report bootstrapped (percentile) 95% confidence intervals [39] [40] associated with each error measure, for each method and all three real-world cases. Note here that although many of the confidence intervals overlap, this does not necessarily imply non-significance of differences between the individual regression approaches' error measures in question. As these were calculated on the same resampled data sets, the measures on each instance were non-independent, and so it is possible for variability evident in the CIs to represent common inter-simulation variance between measures, while for example one was a reliably better performer than another.

1) Systolic and Diastolic BP Data Set: For this data set, we regress the IV *diastolic* BP (regressand) on those of the *systolic* BP (regressor) to estimate the bounds of *diastolic* BP. Figure 8 presents this data set and the associated performance of the regression methods.

Figure 8(b) shows that the PM, LM_w, and LM_c methods produce comparable performance in estimating the regressand bounds. With respect to some of the evaluation metrics ($RMSE^+$ and $RMSE^-$, and MAE^+), the PM produces the best estimation, whereas the LM_w and LM_c methods perform better for MAE^-; all these three methods perform equally best for *MMRE*. Fit for the CRM, CCRM and MinMax methods for these data is also better according to *MMRE*. In this particular case, the CM and CIM methods result in the worst fit.

Along with evaluating model fitness, we also explore the relationship between the uncertainty (and position) of the *systolic* BP and that of the *diastolic* BP in Fig. 9. Note that for space saving, we only present here the *IRG*s using the PM method. Figures 9(a) and (b) show that the center of *systolic* BP is positively correlated with the center of the *diastolic* BP and the range of *systolic* BP exhibits the same trend, i.e., increasing rapidly with the increasing range of the *diastolic* BP. These figures also highlight that the center of *systolic* BP decreases, at a lesser rate, as the range in *diastolic* BP increases, and vice versa, i.e., the range of *systolic* BP decreases as the center of *diastolic* BP increases.

2) Food Snacks Purchase Intention Data Set: With this data set, we first remove outliers and then separately regress the *overall purchase intention* (regressand) on each of the six attributes of snack-bars (i.e., *visual appeal*, *value for money*, *healthiness*, *taste*, *branding*, and *ethics*). Due to page limitations, we present regression results only with respect to two attributes—*ethics* and *taste* (see Figs. 10 and 12). The estimation results in Figs. 10(b) and 12(b) show that the PM, LM_w, LM_c, and MinMax methods produce the best fit with respect to one or more indices. However, bootstrap confidence intervals indicate substantially greater proximity and variability in these estimates than for the previous BP data set. This is especially true for the PM method for the data set representing *taste* and *overall purchase intention*—demonstrating how in cases where transformation is required to maintain mathematical coherence, this can have a large



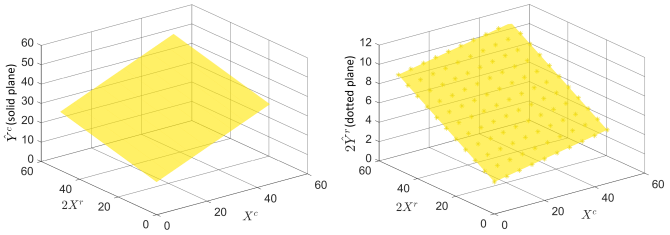
(a) *Ethics* and *overall purchase intention*

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	30.019 [28.708 32.576]	30.240 [27.826 31.679]	26.310 [24.384 28.699]	26.592 [23.929 28.415]	2.941 [2.082 3.792]
MinMax	30.491 [28.451 32.177]	29.839 [27.750 31.569]	26.955 [24.656 28.856]	**26.185** [23.775 28.113]	3.008 [2.141 3.850]
CRM	30.019 [28.802 31.646]	30.240 [28.225 32.155]	26.310 [23.925 28.277]	26.592 [24.127 28.673]	2.961 [2.112 3.819]
CCRM	30.019 [28.802 31.646]	30.240 [28.225 32.155]	26.310 [23.925 28.277]	26.592 [24.127 28.673]	2.961 [2.112 3.819]
CIM	30.624 [28.568 32.342]	29.974 [27.906 31.758]	26.769 [24.439 28.104]	26.510 [24.777 28.556]	3.014 [2.147 3.873]
PM	**29.631** [27.461 32.647]	**29.825** [27.698 32.397]	25.931 [23.389 28.190]	26.203 [23.715 28.406]	2.864 [1.847 3.666]
LM_c	29.636 [27.463 31.283]	29.830 [27.856 31.648]	**25.856** [23.313 27.737]	26.250 [23.771 28.116]	**2.859** [2.028 3.706]
LM_w	29.636 [27.463 31.283]	29.830 [27.856 31.648]	**25.856** [23.313 27.737]	26.250 [23.771 28.116]	**2.859** [2.028 3.706]

Bold = Best; Bootstrapped 95% confidence intervals at 10000 simulations are provided.

(b) Model Performance

Fig. 10: Graphical presentation of *ethics* and *overall purchase intention* [32] and performance of different regression models as to evaluation metrics using this set.



(a) with respect to $\hat{Y}^c$ (b) with respect to $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$)

Fig. 11: *IRG*s showing the relationship between *ethics* and *overall purchase intention* [32] in terms of center and range for the PM method.

detrimental impact on model fit, as discussed earlier.

Beyond the evaluation of model fit, we also use the *IRG*s to explore how the selected attributes of snack-bars influence overall purchase intention, as discussed in [32]. We construct *IRG*s to show the impact of two attributes—*ethics* and *taste* on those of the *overall purchase intention*. Again, to save space, we only show the *IRG*s corresponding to the PM method. Figure 11 shows how the position and uncertainty of *ethics* affect those of *overall purchase intention*. The center of *overall purchase intention*, $\hat{Y}^c$ increases with the center of *ethics*, X^c, as well as its range, $2\hat{X}^r$. Similarly, uncertainty around *overall purchase intention*, $2\hat{Y}^r$ is positively influenced to some extent by both position, X^c and associated uncertainty of the *ethics*, $2X^r$, though much more strongly by the latter. The *IRG*s (shown in Fig. 13) relating to *taste* also suggest implications for both position and width on each aspect of *overall purchase intention*. However, comparing the *IRG*s for *taste* and *ethics*, we observe that the position and uncertainty of two attributes tend to have distinctive impacts on *overall purchase intention*. For example, increasing uncertainty in *ethics* (together with increasing center values) increases position of *overall purchase intention*, whereas this tends to decline for rising uncertainty in *taste* (while also increasing for higher center values).

These results, which are uniquely captured by the IV data, are made visible via the *IRG*s. We emphasise that the regression analysis using only two variables as shown here, while a useful illustration, is limited and a broader analysis would be expected to extract real-world insights. We point the interested reader to [32], where a more detailed analysis—but based on a discrete representation of the data, is conducted.

3) Cyber-security Data Set: Section III-C provided an overview of this data set. Here, we only consider the evade hops as with three attributes these provide a concise basis for presenting the results within the constraints of this paper. For each evade hop, 38 experts gave ratings on following three attributes: *complexity*, *availability of information* and *maturity*. They were also asked to provide an *overall rating* on how difficult it would be *overall* for an attacker to evade a given hop. These ratings were given on a scale from 0 to 100. Table VI summarises the attributes for the evade hops.

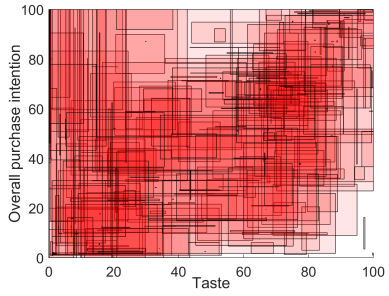
TABLE VI: Attributes for evade hops and associated questions

Attributes	Related Question
Complexity	How complex is the job of providing this kind of defence?
Availability of information	How likely is that there will be publicly available information that could help with evading defence?
Maturity	How mature is this type of technology?
Overall rating	Overall, how difficult would it be for an attacker to do this?

We explore how well the regression methods estimate the IV *overall* difficulty ratings for selected evade components, based upon the three interval-valued attribute ratings, and by comparison with experts' actual overall ratings. For conciseness, we present two evade hops – number 4 and number 22, shown in Figs. 14 and 15 respectively.

Figure 14 (b) shows that the PM method appears to be the best performer for Evade-4 with respect to four evaluation metrics ($RMSE^-$, $RMSE^+$, MAE^+, and $MMRE$) whereas the LM_w method produces the joint best fit on MAE^-. The LM_c method also performs very close to the LM_w method. Again, the CM method provides poor fit to the data. As for the Evade-4 data set, Fig. 15 (b) displays similar results for Evade-22. Particularly, Fig. 15 reveals the PM method as the best approach in terms of most metrics. Further, the LM_w, LM_c, CRM, and CCRM methods come out as suitable measures as to either $RMSE^+$ or MAE^+. We provide all estimated coefficients for both synthetic and real-world data sets in Appendix (see Table VIII and IX). Note here again that bootstrap confidence intervals indicate substantially higher variability in model fit than for the first data set. However, in contrast to the second data set—for which the PM method often showed substantially higher upper confidence bounds than the other methods—the PM method's upper confidence bounds are now often the lowest, or nearly lowest. This may indicate relatively few cases in which transformation was required to maintain mathematical coherence for the PM method with this data set, avoiding the associated model fitness penalties.

Across the experiments, confidence bounds for the PM and the LM methods are overall fairly similar. For the BP data,



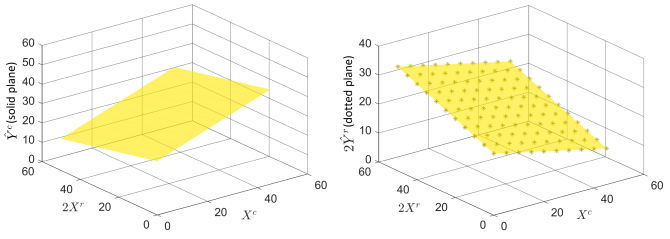
(a) *Taste* and *overall purchase intention*

Methods	$RMSE^{-}$	$RMSE^{+}$	MAE^{-}	MAE^{+}	$MMRE$
CM	24.403 [22.653 26.043]	30.492 [28.080 32.137]	20.756 [18.979 22.375]	24.797 [22.341 26.743]	3.035 [2.399 3.684]
MinMax	23.675 [21.965 25.223]	29.858 [27.610 31.377]	**19.800** [18.078 21.404]	25.110 [22.648 26.913]	**2.697** [2.118 3.277]
CRM	24.031 [22.484 25.659]	29.951 [27.530 31.313]	20.552 [18.458 21.864]	24.484 [22.349 26.677]	2.820 [2.180 3.350]
CCRM	24.031 [22.571 25.754]	29.951 [27.631 31.478]	20.552 [18.894 22.283]	24.484 [22.048 26.303]	2.820 [2.259 3.439]
CIM	24.474 [22.729 26.096]	30.457 [28.047 32.109]	20.856 [19.086 22.446]	24.864 [22.412 26.791]	3.066 [2.428 3.713]
PM	**23.645** [21.973 51.001]	**29.669** [27.545 65.130]	19.806 [18.151 41.454]	24.823 [22.466 56.996]	2.708 [0.898 3.253]
LM_c	23.830 [22.088 25.393]	29.817 [28.279 31.963]	20.387 [18.669 21.941]	**24.338** [22.564 26.673]	2.821 [2.279 3.470]
LM_w	23.830 [22.088 25.393]	29.817 [28.244 31.980]	20.387 [18.669 21.941]	**24.338** [22.516 26.637]	2.821 [2.236 3.436]

Bold = Best; Bootstrapped 95% confidence intervals at 10000 simulations are provided.

(b) Model Performance

Fig. 12: Graphical presentation of *taste* and *overall purchase intention* [32] and performance of different regression models as to evaluation metrics using this set.



(a) with respect to $\hat{Y}^c$ (b) with respect to $\hat{Y}^w$ ($\simeq 2\hat{Y}^r$)

Fig. 13: *IRG*s showing the relationship between *taste* and *overall purchase intention* [32] in terms of center and range for the PM method.

Fig. 8, errors in model fit are fairly small in size due to densely placed intervals. Similarly, for the cyber-security data sets, Figs. 14 and 15, the PM and the LM methods show similar variability in confidence bounds, but the errors are larger as the intervals are relatively scattered, invariably impacting the feasibility of a linear model fit. Similarly, for the food snacks data with respect to *ethics*, Fig. 10, the methods show similar variability in CI bounds, but bigger errors in terms of size—similar to the cyber-security data set.

The only exception where the PM method shows wider variability in CI bounds than other approaches is for the *taste* vs *overall purchase intention* data set, Fig. 12. This higher variability in CI bounds may be due to greater incidence of very wide and scattered intervals within the data set, resulting in increased application of the box-cox transformation and poorer fit across a larger subset of samples/simulations.

Figure 20 in Appendix shows the total number of times across the 10000 simulations the regression approaches (CCRM, PM, LM_c, and LM_w) apply the positivity restrictions/regressand bound transformations to maintain coherence for the real-world data sets. Figure 20 reveals that for the *taste* vs *overall purchase intention* data set, for a significant proportion (22.85% of total 10000 simulations), positivity restrictions/transformation of bounds are applied by the PM and LM_w methods, compared to other sets. Exactly why the PM method is more strongly impacted than the LM_w method for this case requires further research which we will address in a future publication.

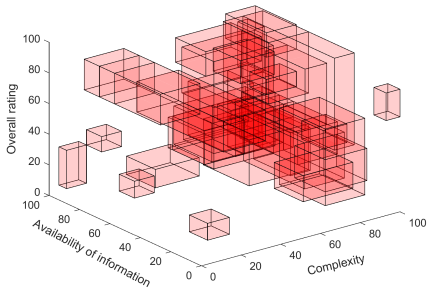
V. Discussion on Suitability of Existing Regression Approaches

This section discusses the suitability of different regression techniques with respect to interval data sets with different properties (e.g., skinny, puffy or mixed), and based upon the results of the analyses conducted above. The experiments consistently show that the PM and LM_w (and often LM_c) generally produce the best model fit—with reliably good performance for skinny/puffy or mixed sets of intervals, when these are disjoint or overlapping. With increasing overlap among intervals, estimation errors also decrease for the CRM, CCRM and MinMax methods, but not generally so far as to equal the performance of either PM or LM methods. In all cases, both MinMax and CRM require checking as to whether they maintain consistency of interval bounds. The real-world data experiments largely substantiate the synthetic experiments in terms of relative model performance.

Results reveal that the LM_w method produces better fitness than its permanently constrained sibling (LM_c) in most cases, as the weaker constraining approach lets model parameters vary freely unless mathematical coherence is being compromised. Both approaches generate the same outcome when the coherence of the bounds does break down and LM_w effectively reverts to LM_c. Similarly, the PM approach may produce much poorer fit in cases where it requires the Box-Cox transformation [21], [25] of the regressand to restore coherence of the bounds. We provide such an example in Appendix (see Fig. 17), and we also hypothesise that this is the cause of the high upper 95% bootstrap confidence bounds in relation to this model's performance on our second real-world data set. Generally, both CM and CIM methods yield poor model-fit.

Having said that, the PM and LM models (particularly its weakly-constrained form, LM_w) were overall the most widely suitable regression approaches. Nevertheless, as we discuss throughout, better performance can be found for data sets following specific properties. To help researchers choose the most suitable method for their data, Fig. 16 provides a flowchart to summarise the findings presented. The figure provides a recommendation on suitability of regression approaches as to different features of the intervals in given data sets.

In addition to comparing model fitness among regression approaches, we have also visualized the relationship between IV variables with respect to their key features (i.e., center and range) by introducing the *IRG*. The *IRG* provides a powerful visual tool, akin to the 'regression line' of traditional numeric regression, offering rapid, intuitive insight into the data. While the *IRG* can be generated for all regression models, irrespective of the regression method, we generally favour the PM method as consistently producing strong model fit.



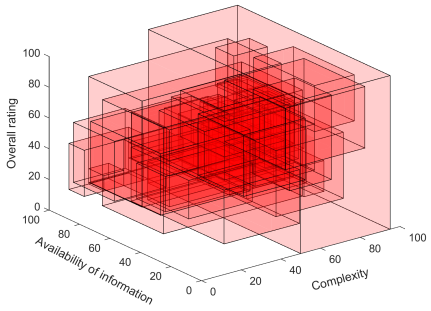
(a) Evade-4

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	24.071 [21.184 29.976]	25.124 [22.109 31.672]	20.745 [17.909 26.513]	20.478 [16.180 25.812]	2.90 [1.243 4.683]
MinMax	21.737 [16.614 24.375]	22.856 [18.159 25.004]	17.948 [13.028 21.189]	19.946 [14.546 22.468]	2.228 [0.965 3.372]
CRM	22.269 [16.612 24.434]	21.562 [17.975 24.792]	18.631 [13.177 21.335]	18.641 [14.401 22.139]	2.296 [0.949 3.294]
CCRM	22.164 [16.670 24.470]	21.756 [17.986 24.776]	18.385 [13.239 21.356]	18.820 [14.453 22.149]	2.271 [0.950 3.279]
CIM	22.593 [17.575 25.378]	22.582 [18.009 25.128]	19.297 [14.019 22.499]	19.158 [14.037 21.830]	2.570 [1.067 3.918]
PM	**21.681** [15.393 24.117]	**20.984** [15.188 22.946]	17.852 [11.628 20.831]	**17.541** [11.630 20.237]	**2.188** [0.858 3.128]
LM_c	21.714 [15.467 23.811]	21.194 [15.738 23.708]	**17.807** [11.759 20.374]	17.909 [12.08 20.542]	2.218 [0.869 3.187]
LM_w	21.714 [15.467 23.811]	21.019 [21.285 32.519]	**17.807** [11.759 20.374]	17.671 [17.580 27.886]	2.218 [0.977 3.303]

Bold = Best; Bootstrapped 95% confidence intervals at 10000 simulations are provided.

(b) Model Performance

Fig. 14: Graphical presentation of Evade-4 data set [34] and performance of different regression models as to evaluation metrics using this data set.



(a) Evade-22

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	27.132 [23.396 39.721]	28.312 [21.275 39.518]	22.759 [20.573 34.154]	23.112 [16.451 32.551]	4.358 [2.406 6.785]
MinMax	14.449 [10.331 16.469]	17.972 [12.832 20.861]	12.051 [8.264 13.907]	14.591 [10.266 17.171]	2.106 [1.035 2.765]
CRM	12.491 [10.186 16.526]	15.399 [12.333 18.632]	10.194 [7.794 13.490]	12.703 [9.816 16.059]	1.492 [0.943 2.381]
CCRM	12.487 [10.305 16.552]	15.403 [12.540 18.743]	10.186 [8.0 13.516]	12.714 [9.994 16.195]	1.485 [0.886 2.357]
CIM	15.636 [11.698 17.887]	18.706 [13.981 21.062]	13.405 [9.767 15.599]	16.011 [11.577 17.740]	2.571 [1.315 3.530]
PM	**12.333** [8.158 14.186]	**15.220** [9.978 16.970]	**9.862** [6.354 11.816]	12.756 [7.610 14.280]	**1.352** [0.647 1.807]
LM_c	12.373 [8.318 14.117]	15.255 [10.667 42.247]	9.903 [6.503 11.798]	**12.652** [8.284 37.102]	1.393 [0.726 1.989]
LM_w	12.373 [8.318 14.117]	15.253 [24.949 50.146]	9.903 [6.503 11.798]	**12.652** [8.284 44.468]	1.392 [0.876 2.115]

Bold = Best; Bootstrapped 95% confidence intervals at 10000 simulations are provided.

(b) Model Performance

Fig. 15: Graphical presentation of Evade-22 data set [34] and performance of different regression models as to evaluation metrics for this data set.

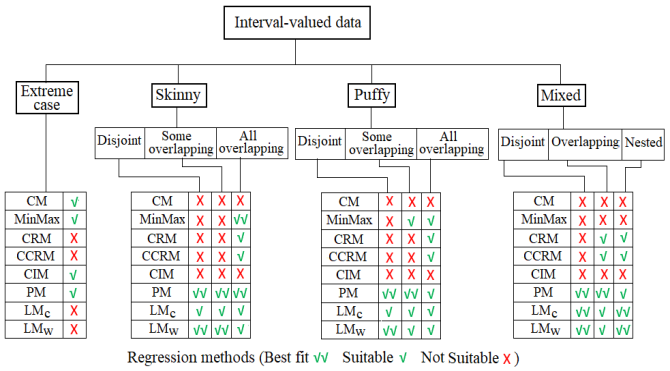


Fig. 16: A flowchart presenting recommendations on linear regression methods with respect to different IV data sets.

Lastly, as for computational efficiency, the LM_c generally requires more execution time than LM_w as it generically restricts parameters. Similarly, the PM approach involves extra execution time when it applies the Box-Cox transformation to restore consistency, in other words, its execution time is dependent on the data set. Overall, in this paper we do not explore execution time in more detail, instead focusing on performance as this is the more common driver of data analysis choices in the domains considered.

VI. Conclusions

Inferential statistics and in particular (linear) regression represent a key building block of AI. For numeric variables, linear regression does not only afford a model for estimation, it provides powerful insight into the relationships between variables. Indeed, the regression-line or graph is perhaps the most widely understood visualisation in data analysis.

While well understood for the numeric case, linear regression for 'fully' IV data sets, i.e. data sets where both regressor and regressand are IV, remain challenging and an active research topic. At the same time, the potential for insights from such data sets is extremely promising, not only for regression, but for AI in general.

The complexity of IV data in terms of how intervals are distributed, how they vary in size, how they overlap–or not, makes it notoriously challenging to establish universal superiority of a single technique. Acknowledging this, in this paper, we have carried out an in-depth review and analysis of linear regression models, exploring their behaviour with respect to systematically designed synthetic and real-world, fully IV data sets, with a view to support researchers and practitioners in the choice of techniques suitable to their problem and data.

Building on existing literature, we have articulated key features of IV data, such as whether intervals are *puffy, skinny, scattered, densely placed* or *overlapping*. For all methods, we have investigated how they do, or do not, maintain mathematical coherence, and the impact of ensuring coherence on model-fitness. As part of this, we also extended current methods, proposing two variants of the LM method [20] to enhance its general suitability for real-world applications by ensuring mathematical coherence of bounds.

Going beyond the review of the state of the art and refinement of regression approaches, the paper discusses the potential of IV data in regression and AI more broadly, in

particular, highlighting the value of mapping not only variable *position* (as in the numeric case), but also uncertainty/range—for example to infer how uncertainty about a regressor, such as the ethical origin of a snack-food product's ingredients, impacts uncertainty—and/or position of a regressand, such as a consumer's purchase-intention.

Acknowledging the powerful aspect of visualization in terms of the popularity of traditional linear regression, we have introduced the Interval Regression Graph (*IRG*), facilitating the interpretation of IV regression models, and by extension, IV data. *IRG*s succinctly visualize and articulate the relationship between IV regressor and regressand as to both position (center) and uncertainty (range). As *IRG*s use all three visualizable dimensions, they are limited to the visualization of the relationship between two variables at a time. Naturally, this can be applied to multivariate cases by presenting the relationship between two variables while keeping other variables constant. An alternative to this approach could be to leverage techniques such as Principal Component Analysis (PCA) for intervals [41], and visualizing the relationships between resulting components. However, here, the intuitive interpretation of said intervals will be complex, potentially limiting one of the primary assets of leveraging intervals in real-world settings as discussed in this paper. We will explore these aspects in future work.

Finally, as alluded to above, we have leveraged the empirical results to provide recommendations on which regression models are best suited to which type of IV data sets, based on the data's features, such as whether intervals strongly overlap, are narrow or wide, etc. This is supported by complementary insights, such as generally better model prediction observed for IV sets with narrow range or condensed interval placement. Further, in this paper, we have provided open-source software which–for the first time–offers access to the suite of IV regression approaches for facilitating adoption and replication.

In summary, analyses have shown that the PM [21] and the LM [20] methods (with its weaker set of restrictions, i.e., LM_w) are the most suitable regression approaches in many cases including where intervals are skinny, puffy or a mix thereof. The constrained variant of the LM model [20], LM_c also performs equally well in many cases. However, when intervals are densely placed or some overlapping/mixed, other existing methods, such as CRM [17], CCRM [18], and MinMax [16] may turn into suitable estimation approaches, offering both relatively simple implementation and efficient computation. In such instances, it is desirable to verify if the CRM [17] and MinMax [16] methods maintain consistency of the bounds. Certainly, maintaining this coherence automatically, such as by the PM, LM, and CCRM methods, tends to require more execution time. The latter may be relevant as the number of variables increases.

With increasing popularity of IV data, being able to analyse and derive the important insights available from these data using AI becomes an increasingly important topic. The substantial body of work advancing IV regression over the last ten years, and advances such as those put forward in this paper, provide the foundation for leveraging intervals more generally in AI. Substantial work remains, from further refining our capacity for communicating resulting insights to advancing our modelling techniques, such as to non-linear approaches.

Beyond advancing techniques, in the future, we aim to explore additional real-world cases to demonstrate and advance the understanding of the value of using intervals, rather than 'crisp', numeric data. Specifically, we will revisit the nature of the information captured by intervals, whether disjunctive (such as in confidence intervals), or conjunctive (true ranges, such as the real numbers between 2 and 6) and explore how different analysis techniques can, and whether they should, deal with such data differently.

ACKNOWLEDGMENT

This work was supported by the UK EPSRC's Leveraging the Multi-Stakeholder Nature of Cyber Security EP/P011918/1 and Horizon: Trusted Data-Driven Products EP/T022493/1 grants.

REFERENCES

[1] L.-C. Lin, H.-L. Chien, and S. Lee, "Symbolic interval-valued data analysis for time series based on auto-interval-regressive models," *Statistical Methods & Applications*, pp. 1–21, 2020.

[2] Y. Sun and D. Ralescu, "A normal hierarchical model and minimum contrast estimation for random intervals," *Annals of the Institute of Statistical Mathematics*, vol. 67, no. 2, pp. 313–333, 2015.

[3] F. Liu and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1503–1521, 2008.

[4] C. Wagner, S. Miller, J. M. Garibaldi, D. T. Anderson, and T. C. Havens, "From interval-valued data to general type-2 fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 2, pp. 248–269, 2015.

[5] Z. Ellerby, C. Wagner, and S. B. Broomell, "Capturing richer information: On establishing the validity of an interval-valued survey response mode," *Behavior Research Methods*, pp. 1–23, 2021.

[6] S. Kabir, C. Wagner, T. C. Havens, and D. T. Anderson, "A similarity measure based on bidirectional subsethood for intervals," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 11, pp. 2890–2904, 2020.

[7] P. Brito, "Modelling and analysing interval data," in *Advances in data analysis*. Springer, 2007, pp. 197–208.

[8] C.-C. Yeh, "Tree-based regression for interval-valued data," Master of Statistics, Utah State University, USA, 2017.

[9] G. González-Rivera and W. Lin, "Constrained regression for interval-valued data," *Journal of Business & Economic Statistics*, vol. 31, no. 4, pp. 473–490, 2013.

[10] G. González-Rodríguez, Á. Blanco, N. Corral, and A. Colubi, "Least squares estimation of linear regression models for convex compact random sets," *Advances in Data Analysis and Classification*, vol. 1, no. 1, pp. 67–81, 2007.

[11] A. Blanco-Fernández, N. Corral, and G. González-Rodríguez, "Estimation of a flexible simple linear model for interval data based on set arithmetic," *Computational Statistics & Data Analysis*, vol. 55, no. 9, pp. 2568–2578, 2011.

[12] R. Boukezzoula, S. Galichet, and A. Bisserier, "A midpoint–radius approach to regression with interval data," *International Journal of Approximate Reasoning*, vol. 52, no. 9, pp. 1257–1271, 2011.

[13] A. Blanco-Fernández, A. Colubi, and M. GarcíA-BáRzana, "A set arithmetic-based linear regression model for modelling interval-valued responses through real-valued variables," *Information Sciences*, vol. 247, pp. 109–122, 2013.

[14] M. García-Bárzana, A. B. Ramos-Guajardo, A. Colubi, and E. J. Kontoghiorghes, "Multiple linear regression models for random intervals: a set arithmetic approach," *Computational Statistics*, vol. 35, no. 2, pp. 755–773, 2020.

[15] L. Billard and E. Diday, "Regression analysis for interval-valued data," in *Data Analysis, Classification, and Related Methods*. Springer, 2000, pp. 369–374.

[16] ——, "Symbolic regression analysis," in *Classification, Clustering, and Data Analysis*. Springer, 2002, pp. 281–288.

[17] E. d. A. L. Neto and F. d. A. de Carvalho, "Centre and range method for fitting a linear regression model to symbolic interval data," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1500–1515, 2008.

[18] ——, "Constrained linear regression models for symbolic interval-valued variables," *Computational Statistics & Data Analysis*, vol. 54, no. 2, pp. 333–347, 2010.
[19] H. Wang, R. Guan, and J. Wu, "Linear regression of interval-valued data based on complete information in hypercubes," *Journal of Systems Science and Systems Engineering*, vol. 21, no. 4, pp. 422–442, 2012.
[20] Y. Sun and D. Ralescu, "A linear model for interval-valued data," *arXiv preprint arXiv:1506.03541*, 2015.
[21] L. C. Souza, R. M. Souza, G. J. Amaral, and T. M. Silva Filho, "A parametrized approach for linear regression of interval data," *Knowledge-Based Systems*, vol. 131, pp. 149–159, 2017.
[22] P. Hao and J. Guo, "Constrained center and range joint model for interval-valued symbolic data regression," *Computational Statistics & Data Analysis*, vol. 116, pp. 106–138, 2017.
[23] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. Prentice-Hall, Newyork, 1974.
[24] R. E. Moore, *Interval analysis*. Prentice-Hall Englewood Cliffs, 1966, vol. 4.
[25] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–243, 1964.
[26] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, "Experimental uncertainty estimation and statistics for data having interval uncertainty," *Sandia National Laboratories, Report SAND2007-0939*, vol. 162, 2007.
[27] I. Couso and D. Dubois, "Statistical reasoning with set-valued information: Ontic vs. epistemic views," *International Journal of Approximate Reasoning*, vol. 55, no. 7, pp. 1502–1518, 2014.
[28] R. R. Picard and R. D. Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 575–583, 1984.
[29] G. Beliakov and S. James, "A penalty-based aggregation operator for non-convex intervals," *Knowledge-Based Systems*, vol. 70, pp. 335–344, 2014.
[30] N. R. Draper and H. Smith, *Applied regression analysis*. John Wiley & Sons, 1998, vol. 326.
[31] S. Kabir and C. Wagner, "Visualization of interval regression for facilitating data and model insight," in *IEEE International Conference on Fuzzy Systems*. IEEE, 2022, pp. 1–7.
[32] Z. Ellerby, O. Miles, J. McCulloch, and C. Wagner, "Insights from interval-valued ratings of consumer products—a decsys appraisal," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2020, pp. 1–8.
[33] Z. Ellerby, J. McCulloch, J. Young, and C. Wagner, "Decsys–discrete and ellipse-based response capture system," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2019, pp. 1–6.
[34] S. Miller, C. Wagner, U. Aickelin, and J. M. Garibaldi, "Modelling cyber-security experts' decision making processes using aggregation operators," *Computers & Security*, vol. 62, pp. 229–245, 2016.
[35] Z. Ellerby, J. McCulloch, M. Wilson, and C. Wagner, "Exploring how component factors and their uncertainty affect judgements of risk in cyber-security," in *International Conference on Critical Information Infrastructures Security*. Springer, 2019, pp. 31–42.
[36] D. İçen and H. Demirhan, "Error measures for fuzzy linear regression: Monte carlo simulation approach," *Applied Soft Computing*, vol. 46, pp. 104–114, 2016.
[37] R. A. Fagundes, R. M. de Souza, and Y. M. Soares, "Quantile regression of interval-valued data," in *23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 2586–2591.
[38] S. Kabir and C. Wagner, "Interval-valued regression-sensitivity to data set features," in *IEEE International Conference on Fuzzy Systems*. IEEE, 2021, pp. 1–7.
[39] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.
[40] T. C. Hesterberg, "What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum," *The American Statistician*, vol. 69, no. 4, pp. 371–386, 2015.
[41] F. Gioia and C. N. Lauro, "Principal component analysis on interval data," *Computational Statistics*, vol. 21, no. 2, pp. 343–363, 2006.

APPENDIX

TABLE VII: Synthetic IV data sets used for sensitivity analysis (Set-1 to Set-11)

(a) Set-1

Entry	X^-	X^+	Y^-	Y^+
1	0.6	0.95	1.0	2.0
2	1.4	1.92	1.01	3.01
3	2.4	3.13	1.02	4.02
4	3.6	4.45	1.03	5.03
5	5.0	5.95	1.04	6.04
	μ_{X^c}=2.94	μ_{X^w}=0.68	μ_{Y^c}=2.52	μ_{Y^w}=3.0
	σ_{X^c}=1.87	σ_{X^w}=0.24	σ_{Y^c}=0.81	σ_{Y^w}=1.58

(b) Set-2

Entry	X^-	X^+	Y^-	Y^+
1	0.6	0.65	1.0	2.0
2	1.4	1.44	1.01	3.01
3	2.4	2.46	1.02	4.02
4	3.6	3.61	1.03	5.03
5	5.0	5.03	1.04	6.04
	μ_{X^c}=2.62	μ_{X^w}=0.04	μ_{Y^c}=2.52	μ_{Y^w}=3.0
	σ_{X^c}=1.74	σ_{X^w}=0.02	σ_{Y^c}=0.81	σ_{Y^w}=1.58

(c) Set-3

Entry	X^-	X^+	Y^-	Y^+
1	4.4	4.7	3.8	4.2
2	3.4	3.8	5.8	6.3
3	4.8	5.3	7.5	7.8
4	6.4	6.7	5.4	5.8
5	8.1	8.5	6.8	7.3
	μ_{X^c}=5.61	μ_{X^w}=0.38	μ_{Y^c}=6.07	μ_{Y^w}=0.42
	σ_{X^c}=1.84	σ_{X^w}=0.08	σ_{Y^c}=1.41	σ_{Y^w}=0.08

(d) Set-4

Entry	X^-	X^+	Y^-	Y^+
1	3.7	4.0	5.5	5.9
2	3.4	3.8	5.8	6.3
3	4.8	5.3	7.5	7.8
4	7.9	8.2	6.5	6.9
5	8.1	8.5	6.8	7.3
	μ_{X^c}=5.77	μ_{X^w}=0.38	μ_{Y^c}=6.63	μ_{Y^w}=0.42
	σ_{X^c}=2.26	σ_{X^w}=0.08	σ_{Y^c}=0.78	σ_{Y^w}=0.08

(e) Set-5

Entry	X^-	X^+	Y^-	Y^+
1	5.4	5.7	6.4	6.8
2	5.6	6.0	6.1	6.6
3	5.3	5.8	6.6	6.9
4	5.5	5.8	6.6	7.0
5	5.7	6.1	6.3	6.8
	μ_{X^c}=5.69	μ_{X^w}=0.38	μ_{Y^c}=6.61	μ_{Y^w}=0.42
	σ_{X^c}=0.16	σ_{X^w}=0.08	σ_{Y^c}=0.18	σ_{Y^w}=0.08

(f) Set-6

Entry	X^-	X^+	Y^-	Y^+
1	1.2	2.4	4.8	6.8
2	2.6	4.2	3.6	5.8
3	3.8	5.6	7.6	9.5
4	5.0	7.8	2.6	4.5
5	7.6	9.4	5.0	7.6
	μ_{X^c}=4.96	μ_{X^w}=1.84	μ_{Y^c}=5.78	μ_{Y^w}=2.12
	σ_{X^c}=2.60	σ_{X^w}=0.59	σ_{Y^c}=1.88	σ_{Y^w}=0.29

(g) Set-7

Entry	X^-	X^+	Y^-	Y^+
1	1.2	2.4	4.8	6.8
2	2.0	3.6	3.3	5.5
3	3.8	5.6	7.6	9.5
4	5.8	8.6	4.2	6.1
5	7.4	9.2	4.8	7.4
	μ_{X^c}=4.96	μ_{X^w}=1.84	μ_{Y^c}=6.00	μ_{Y^w}=2.12
	σ_{X^c}=2.78	σ_{X^w}=0.59	σ_{Y^c}=1.57	σ_{Y^w}=0.29

(h) Set-8

Entry	X^-	X^+	Y^-	Y^+
1	5.3	6.5	6.2	8.2
2	5.5	7.1	5.2	7.4
3	4.9	6.7	6.5	8.4
4	5.1	7.9	5.7	7.6
5	6.2	8.0	5.9	8.5
	μ_{X^c}=6.32	μ_{X^w}=1.84	μ_{Y^c}=6.96	μ_{Y^w}=2.12
	σ_{X^c}=0.52	σ_{X^w}=0.59	σ_{Y^c}=0.47	σ_{Y^w}=0.29

(i) Set-9

Entry	X^-	X^+	Y^-	Y^+
1	2.8	5.4	6.4	9.2
2	3.9	5.1	4.4	5.8
3	1.1	3.2	2.4	4.6
4	5.3	5.6	1.8	2.3
5	7.7	9.3	3.2	5.0
	μ_{X^c}=4.94	μ_{X^w}=1.56	μ_{Y^c}=4.51	μ_{Y^w}=1.74
	σ_{X^c}=2.32	σ_{X^w}=0.88	σ_{Y^c}=2.15	σ_{Y^w}=0.86

(j) Set-10

Entry	X^-	X^+	Y^-	Y^+
1	4.8	6.4	5.4	7.2
2	4.8	5.1	4.0	4.5
3	4.2	5.4	4.4	5.8
4	5.0	7.6	3.6	6.4
5	6.1	8.2	4.8	7.0
	μ_{X^c}=5.76	μ_{X^w}=1.56	μ_{Y^c}=5.31	μ_{Y^w}=1.74
	σ_{X^c}=0.98	σ_{X^w}=0.88	σ_{Y^c}=0.80	σ_{Y^w}=0.86

(k) Set-11

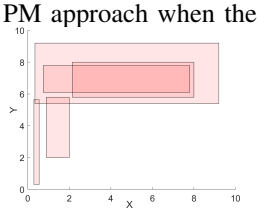
Entry	X^-	X^+	Y^-	Y^+
1	6.1	6.4	4.7	5.2
2	5.6	6.8	4.3	5.7
3	5.4	7.0	4.1	5.9
4	5.2	7.3	3.9	6.1
5	5.0	7.6	3.6	6.4
	μ_{X^c}=6.24	μ_{X^w}=1.56	μ_{Y^c}=4.99	μ_{Y^w}=1.74
	σ_{X^c}=0.04	σ_{X^w}=0.88	σ_{Y^c}=0.02	σ_{Y^w}=0.86

Example-1: Relatively low fit of the PM approach when the Box-Cox transformation is applied.

(a) Set-12

Entry	X^-	X^+	Y^-	Y^+
1	0.5	0.6	5.5	5.8
2	0.2	0.5	0.8	10
3	0.8	1.0	4.8	6.8
4	0.6	0.9	2.2	10
5	2.0	2.3	1.8	9.8
	μ_{X^c}=0.94	μ_{X^w}=0.24	μ_{Y^c}=5.75	μ_{Y^w}=5.46
	σ_{X^c}=0.71	σ_{X^w}=0.09	σ_{Y^c}=0.25	σ_{Y^w}=4.02

(b) 2D visualization



(c) Model Performance

Methods	$RMSE^-$	$RMSE^+$	MAE^-	MAE^+	$MMRE$
CM	3.271	3.251	2.716	2.716	1.159
MinMax	1.797	1.714	1.654	1.558	0.573
CRM	0.645	0.441	0.567	0.372	0.187
CCRM	0.995	0.899	0.859	0.783	0.280
CIM	3.271	3.251	2.716	2.716	1.159
PM	2.636	7.594	2.037	7.375	0.715
LM_c	0.646	1.603	0.565	1.380	0.261
LM_w	0.646	1.603	0.565	1.380	0.261

Fig. 17: Comparatively low fit of the PM approach.

This article has been accepted for publication in IEEE Transactions on Artificial Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAI.2023.3234930

TABLE VIII: Estimated coefficients of different regression methods for synthetic data sets, Set-1 to set-11

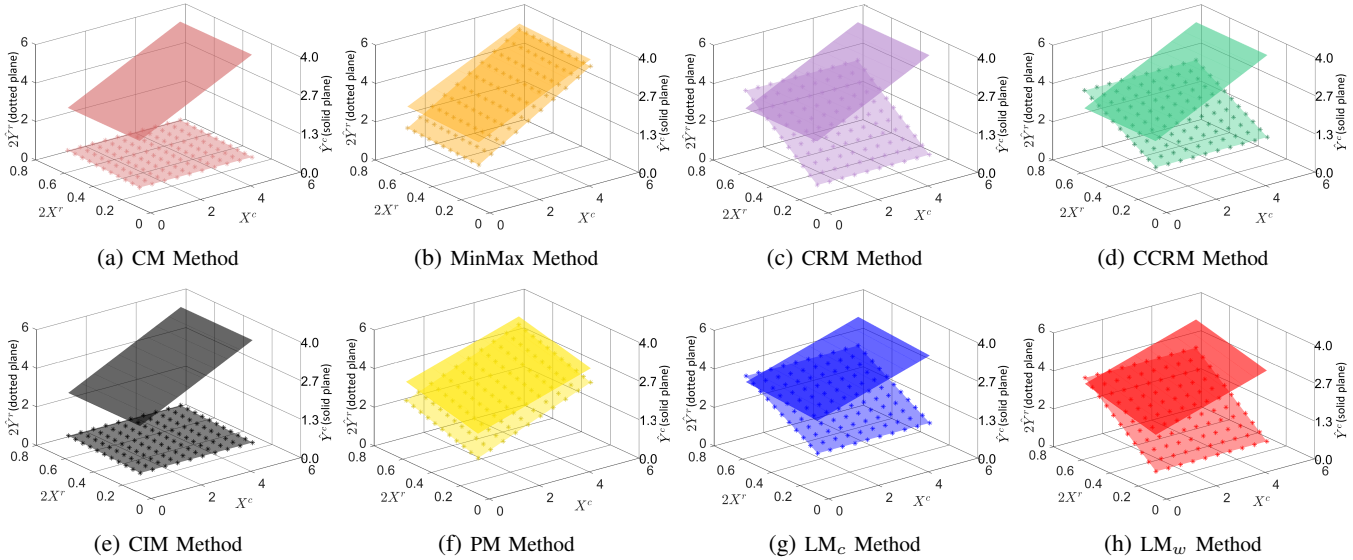
Methods	Synthetic data sets					
	Set-1	Set-2	Set-3	Set-4	Set-5	Set-6
CM	$\hat{\beta}_0^c = 1.256, \hat{\beta}_1^c = 0.43$	$\hat{\beta}_0^c = 1.316, \hat{\beta}_1^c = 0.46$	$\hat{\beta}_0^c = 4.608, \hat{\beta}_1^c = 0.261$	$\hat{\beta}_0^c = 5.637, \hat{\beta}_1^c = 0.172$	$\hat{\beta}_0^c = 10.54, \hat{\beta}_1^c = -0.691$	$\hat{\beta}_0^c = 5.866, \hat{\beta}_1^c = -0.017$
MinMax	$\hat{\beta}_0^- = 0.997, \hat{\beta}_1^- = 0.009$ $\hat{\beta}_0^+ = 1.393, \hat{\beta}_1^+ = 0.801$	$\hat{\beta}_0^- = 0.997, \hat{\beta}_1^- = 0.009$ $\hat{\beta}_0^+ = 1.611, \hat{\beta}_1^+ = 0.913$	$\hat{\beta}_0^- = 4.56, \hat{\beta}_1^- = 0.24$ $\hat{\beta}_0^+ = 4.65, \hat{\beta}_1^+ = 0.281$	$\hat{\beta}_0^- = 5.505, \hat{\beta}_1^- = 0.164$ $\hat{\beta}_0^+ = 5.766, \hat{\beta}_1^+ = 0.18$	$\hat{\beta}_0^- = 11.35, \hat{\beta}_1^- = -0.90$ $\hat{\beta}_0^+ = 9.433, \hat{\beta}_1^+ = -0.444$	$\hat{\beta}_0^- = 4.778, \hat{\beta}_1^- = -0.015$ $\hat{\beta}_0^+ = 6.967, \hat{\beta}_1^+ = -0.022$
CRM	$\hat{\beta}_0^c = 1.256, \hat{\beta}_1^c = 0.43$ $\hat{\beta}_0^r = -0.678, \hat{\beta}_1^r = 6.407$	$\hat{\beta}_0^c = 1.316, \hat{\beta}_1^c = 0.46$ $\hat{\beta}_0^r = 2.398, \hat{\beta}_1^r = -47.297$	$\hat{\beta}_0^c = 4.608, \hat{\beta}_1^c = 0.261$ $\hat{\beta}_0^r = 0.264, \hat{\beta}_1^r = -0.286$	$\hat{\beta}_0^c = 5.636, \hat{\beta}_1^c = 0.172$ $\hat{\beta}_0^r = 0.264, \hat{\beta}_1^r = -0.286$	$\hat{\beta}_0^c = 10.54, \hat{\beta}_1^c = -0.691$ $\hat{\beta}_0^r = 0.264, \hat{\beta}_1^r = -0.286$	$\hat{\beta}_0^c = 5.867, \hat{\beta}_1^c = -0.017$ $\hat{\beta}_0^r = 1.168, \hat{\beta}_1^r = -0.117$
CCRM	$\hat{\beta}_0^c = 1.256, \hat{\beta}_1^c = 0.43$ $\hat{\beta}_0^r = 0.0, \hat{\beta}_1^r = 4.60$	$\hat{\beta}_0^c = 1.316, \hat{\beta}_1^c = 0.46$ $\hat{\beta}_0^r = 1.5, \hat{\beta}_1^r = 0.0$	$\hat{\beta}_0^c = 4.608, \hat{\beta}_1^c = 0.261$ $\hat{\beta}_0^r = 0.21, \hat{\beta}_1^r = 0.0$	$\hat{\beta}_0^c = 5.636, \hat{\beta}_1^c = 0.172$ $\hat{\beta}_0^r = 0.21, \hat{\beta}_1^r = 0.0$	$\hat{\beta}_0^c = 10.54, \hat{\beta}_1^c = -0.691$ $\hat{\beta}_0^r = 0.21, \hat{\beta}_1^r = 0.0$	$\hat{\beta}_0^c = 5.866, \hat{\beta}_1^c = -0.017$ $\hat{\beta}_0^r = 1.06, \hat{\beta}_1^r = 0.0$
CIM	$\hat{\beta}_0 = 1.275, \hat{\beta}_1 = 0.423$	$\hat{\beta}_0 = 1.316, \hat{\beta}_1 = 0.46$	$\hat{\beta}_0 = 4.615, \hat{\beta}_1 = 0.26$	$\hat{\beta}_0 = 5.639, \hat{\beta}_1 = 0.172$	$\hat{\beta}_0 = 9.0, \hat{\beta}_1 = -0.42$	$\hat{\beta}_0 = 5.861, \hat{\beta}_1 = -0.016$
PM	$\hat{\beta}_0^- = 0.987, \hat{\alpha}_1^- = -0.023$ $\hat{\omega}_1^- = 0.028, \hat{\beta}_0^+ = 0.737$ $\hat{\alpha}_1^+ = -2.295, \hat{\omega}_1^+ = 2.82$	$\hat{\beta}_0^- = 0.997, \hat{\alpha}_1^- = 0.008$ $\hat{\omega}_1^- = 0.001, \hat{\beta}_0^+ = 1.655$ $\hat{\alpha}_1^+ = 0.806, \hat{\omega}_1^+ = 0.101$	$\hat{\beta}_0^- = -1.683, \hat{\alpha}_1^- = -15.185$ $\hat{\omega}_1^- = 15.49, \hat{\beta}_0^+ = -1.206$ $\hat{\alpha}_1^+ = -14.908, \hat{\omega}_1^+ = 15.222$	$\hat{\beta}_0^- = 2.295, \hat{\alpha}_1^- = -7.562$ $\hat{\omega}_1^- = 7.772, \hat{\beta}_0^+ = 2.796$ $\hat{\alpha}_1^+ = -7.289, \hat{\omega}_1^+ = 7.503$	$\hat{\beta}_0^- = 11.66, \hat{\alpha}_1^- = -0.60$ $\hat{\omega}_1^- = -0.333, \hat{\beta}_0^+ = 9.433$ $\hat{\alpha}_1^+ = 0.0, \hat{\omega}_1^+ = -0.444$	$\hat{\beta}_0^- = 7.26, \hat{\alpha}_1^- = 2.069$ $\hat{\omega}_1^- = -1.854, \hat{\beta}_0^+ = 9.58$ $\hat{\alpha}_1^+ = 2.565, \hat{\omega}_1^+ = -2.228$
LM$_c$	$\hat{\eta} = 1.541, \hat{\alpha}_1 = 2.045$ $\hat{\beta}_1 = -1.779, \hat{\theta} = 0.0$ $\hat{\gamma}_1 = 4.60$	$\hat{\eta} = -1.073, \hat{\alpha}_1 = -23.241$ $\hat{\beta}_1 = 23.7, \hat{\theta} = 3.0$ $\hat{\gamma}_1 = 0.0$	$\hat{\eta} = -1.709, \hat{\alpha}_1 = -15.189$ $\hat{\beta}_1 = 15.499, \hat{\theta} = 0.42$ $\hat{\gamma}_1 = 0.0$	$\hat{\eta} = 2.282, \hat{\alpha}_1 = -7.568$ $\hat{\beta}_1 = 7.78, \hat{\theta} = 0.42$ $\hat{\gamma}_1 = 0.0$	$\hat{\eta} = 10.282, \hat{\alpha}_1 = -0.443$ $\hat{\beta}_1 = -0.246, \hat{\theta} = 0.42$ $\hat{\gamma}_1 = 0.0$	$\hat{\eta} = 7.252, \hat{\alpha}_1 = 2.258$ $\hat{\beta}_1 = -1.982, \hat{\theta} = 2.12$ $\hat{\gamma}_1 = 0.0$
LM$_w$	$\hat{\eta} = 1.541, \hat{\alpha}_1 = 2.045$ $\hat{\beta}_1 = -1.779, \hat{\theta} = -1.357$ $\hat{\gamma}_1 = 6.407$	$\hat{\eta} = -1.073, \hat{\alpha}_1 = -23.241$ $\hat{\beta}_1 = 23.7, \hat{\theta} = 4.797$ $\hat{\gamma}_1 = -47.297$	$\hat{\eta} = -1.709, \hat{\alpha}_1 = -15.189$ $\hat{\beta}_1 = 15.499, \hat{\theta} = 0.529$ $\hat{\gamma}_1 = -0.286$	$\hat{\eta} = 2.282, \hat{\alpha}_1 = -7.568$ $\hat{\beta}_1 = 7.78, \hat{\theta} = 0.529$ $\hat{\gamma}_1 = -0.286$	$\hat{\eta} = 10.282, \hat{\alpha}_1 = -0.443$ $\hat{\beta}_1 = -0.246, \hat{\theta} = 0.529$ $\hat{\gamma}_1 = -0.286$	$\hat{\eta} = 7.252, \hat{\alpha}_1 = 2.258$ $\hat{\beta}_1 = -1.982, \hat{\theta} = 2.337$ $\hat{\gamma}_1 = -0.118$

Methods	Synthetic Data Sets				
	Set-7	set-8	Set-9	Set-10	Set-11
CM	$\hat{\beta}_0^c = 5.702, \hat{\beta}_1^c = 0.06$	$\hat{\beta}_0^c = 8.186, \hat{\beta}_1^c = -0.194$	$\hat{\beta}_0^c = 5.177, \hat{\beta}_1^c = -0.135$	$\hat{\beta}_0^c = 2.97, \hat{\beta}_1^c = 0.406$	$\hat{\beta}_0^c = 5.436, \hat{\beta}_1^c = -0.071$
MinMax	$\hat{\beta}_0^- = 4.738, \hat{\beta}_1^- = 0.049$ $\hat{\beta}_0^+ = 6.683, \hat{\beta}_1^+ = 0.064$	$\hat{\beta}_0^- = 7.736, \hat{\beta}_1^- = -0.34$ $\hat{\beta}_0^+ = 8.654, \hat{\beta}_1^+ = -0.088$	$\hat{\beta}_0^- = 4.272, \hat{\beta}_1^- = -0.152$ $\hat{\beta}_0^+ = 5.528, \hat{\beta}_1^+ = -0.026$	$\hat{\beta}_0^- = 3.603, \hat{\beta}_1^- = 0.168$ $\hat{\beta}_0^+ = 2.309, \hat{\beta}_1^+ = 0.592$	$\hat{\beta}_0^- = -1.202, \hat{\beta}_1^- = 0.975$ $\hat{\beta}_0^+ = -0.961, \hat{\beta}_1^+ = 0.972$
CRM	$\hat{\beta}_0^c = 5.702, \hat{\beta}_1^c = 0.06$ $\hat{\beta}_0^r = 1.168, \hat{\beta}_1^r = -0.118$	$\hat{\beta}_0^c = 8.186, \hat{\beta}_1^c = -0.194$ $\hat{\beta}_0^r = 1.168, \hat{\beta}_1^r = -0.118$	$\hat{\beta}_0^c = 5.177, \hat{\beta}_1^c = -0.135$ $\hat{\beta}_0^r = 0.104, \hat{\beta}_1^r = 0.983$	$\hat{\beta}_0^c = 2.97, \hat{\beta}_1^c = 0.406$ $\hat{\beta}_0^r = 0.104, \hat{\beta}_1^r = 0.983$	$\hat{\beta}_0^c = 5.436, \hat{\beta}_1^c = -0.071$ $\hat{\beta}_0^r = 0.104, \hat{\beta}_1^r = 0.983$
CCRM	$\hat{\beta}_0^c = 5.702, \hat{\beta}_1^c = 0.06$ $\hat{\beta}_0^r = 1.06, \hat{\beta}_1^r = 0.0$	$\hat{\beta}_0^c = 8.186, \hat{\beta}_1^c = -0.194$ $\hat{\beta}_0^r = 1.06, \hat{\beta}_1^r = 0.0$	$\hat{\beta}_0^c = 5.177, \hat{\beta}_1^c = -0.135$ $\hat{\beta}_0^r = 0.104, \hat{\beta}_1^r = 0.983$	$\hat{\beta}_0^c = 2.97, \hat{\beta}_1^c = 0.406$ $\hat{\beta}_0^r = 0.104, \hat{\beta}_1^r = 0.983$	$\hat{\beta}_0^c = 5.436, \hat{\beta}_1^c = -0.071$ $\hat{\beta}_0^r = 0.104, \hat{\beta}_1^r = 0.983$
CIM	$\hat{\beta}_0 = 5.716, \hat{\beta}_1 = 0.057$	$\hat{\beta}_0 = 7.47, \hat{\beta}_1 = -0.081$	$\hat{\beta}_0 = 5.14, \hat{\beta}_1 = -0.128$	$\hat{\beta}_0 = 3.554, \hat{\beta}_1 = 0.305$	$\hat{\beta}_0 = 4.992, \hat{\beta}_1 = 0.001$
PM	$\hat{\beta}_0^+ = 5.526, \hat{\alpha}_1^- = 0.769$ $\hat{\omega}_1^- = -0.628, \hat{\beta}_0^+ = 7.992$ $\hat{\alpha}_1^+ = 1.342, \hat{\omega}_1^+ = -1.081$	$\hat{\beta}_0^+ = 8.54, \hat{\alpha}_1^- = -0.157$ $\hat{\omega}_1^- = -0.248, \hat{\beta}_0^+ = 7.561$ $\hat{\alpha}_1^+ = 0.431, \hat{\omega}_1^+ = -0.258$	$\hat{\beta}_0^+ = 0.842, \hat{\alpha}_1^- = -1.408$ $\hat{\omega}_1^- = 1.513, \hat{\beta}_0^+ = 0.973$ $\hat{\alpha}_1^+ = -2.395, \hat{\omega}_1^+ = 2.512$	$\hat{\beta}_0^+ = 3.458, \hat{\alpha}_1^- = 0.338$ $\hat{\omega}_1^- = -0.107, \hat{\beta}_0^+ = 3.934$ $\hat{\alpha}_1^+ = -0.728, \hat{\omega}_1^+ = 0.897$	$\hat{\beta}_0^+ = 7.086, \hat{\alpha}_1^- = 0.283$ $\hat{\omega}_1^- = -0.643, \hat{\beta}_0^+ = 7.082$ $\hat{\alpha}_1^+ = -0.682, \hat{\omega}_1^+ = 0.356$
LM$_c$	$\hat{\eta} = 5.591, \hat{\alpha}_1 = 0.997$ $\hat{\beta}_1 = -0.796, \hat{\theta} = 2.12$ $\hat{\gamma}_1 = 0.0$	$\hat{\eta} = 6.882, \hat{\alpha}_1 = 0.078$ $\hat{\beta}_1 = -0.194, \hat{\theta} = 2.12$ $\hat{\gamma}_1 = 0.0$	$\hat{\eta} = 0.804, \hat{\alpha}_1 = -1.408$ $\hat{\beta}_1 = 1.521, \hat{\theta} = 0.207$ $\hat{\gamma}_1 = 0.983$	$\hat{\eta} = 3.593, \hat{\alpha}_1 = 0.297$ $\hat{\beta}_1 = -0.096, \hat{\theta} = 0.207$ $\hat{\gamma}_1 = 0.983$	$\hat{\eta} = 6.98, \hat{\alpha}_1 = 0.292$ $\hat{\beta}_1 = -0.634, \hat{\theta} = 0.207$ $\hat{\gamma}_1 = 0.983$
LM$_w$	$\hat{\eta} = 5.591, \hat{\alpha}_1 = 0.997$ $\hat{\beta}_1 = -0.796, \hat{\theta} = 2.337$ $\hat{\gamma}_1 = -0.118$	$\hat{\eta} = 6.882, \hat{\alpha}_1 = 0.078$ $\hat{\beta}_1 = -0.194, \hat{\theta} = 2.337$ $\hat{\gamma}_1 = -0.118$	$\hat{\eta} = 0.804, \hat{\alpha}_1 = -1.408$ $\hat{\beta}_1 = 1.521, \hat{\theta} = 0.207$ $\hat{\gamma}_1 = 0.983$	$\hat{\eta} = 3.593, \hat{\alpha}_1 = 0.297$ $\hat{\beta}_1 = -0.096, \hat{\theta} = 0.207$ $\hat{\gamma}_1 = 0.983$	$\hat{\eta} = 6.98, \hat{\alpha}_1 = 0.292$ $\hat{\beta}_1 = -0.635, \hat{\theta} = 0.207$ $\hat{\gamma}_1 = 0.983$

TABLE IX: Estimated coefficients of different regression methods for real-world IV data sets

Methods	Systolic-Diastolic Blood Pressure data set [13]	Cyber-Security Vulnerability Assessment Data Set [34]	
		Evade-4	Evade-22
CM	$\hat{\beta}_0^c = 1.886, \hat{\beta}_1^c = 0.438$	$\hat{\beta}_0^c = 25.979, \hat{\beta}_1^c = 0.029, \hat{\beta}_2^c = -0.289, \hat{\beta}_3^c = 0.448$	$\hat{\beta}_0^c = 60.214, \hat{\beta}_1^c = 0.052, \hat{\beta}_2^c = -0.526, \hat{\beta}_3^c = 0.146$
MinMax	$\hat{\beta}_0^- = -0.769, \hat{\beta}_1^- = 0.591$ $\hat{\beta}_0^+ = 5.301, \hat{\beta}_1^+ = 0.303$	$\hat{\beta}_0^- = -19.068, \hat{\beta}_1^- = -0.020, \hat{\beta}_2^- = -0.253, \hat{\beta}_3^- = 0.412$ $\hat{\beta}_0^+ = 26.608, \hat{\beta}_1^+ = 0.078, \hat{\beta}_2^+ = -0.243, \hat{\beta}_3^+ = 0.472$	$\hat{\beta}_0^- = -19.06, \hat{\beta}_1^- = 0.089, \hat{\beta}_2^- = -0.212, \hat{\beta}_3^- = 0.141$ $\hat{\beta}_0^+ = 73.163, \hat{\beta}_1^+ = 0.174, \hat{\beta}_2^+ = -0.657, \hat{\beta}_3^+ = 0.279$
CRM	$\hat{\beta}_0^c = 1.886, \hat{\beta}_1^c = 0.438$ $\hat{\beta}_0^r = 1.621, \hat{\beta}_1^r = -0.246$	$\hat{\beta}_0^c = 25.979, \hat{\beta}_1^c = 0.029, \hat{\beta}_2^c = -0.289, \hat{\beta}_3^c = 0.448$ $\hat{\beta}_0^r = 3.511, \hat{\beta}_1^r = -0.366, \hat{\beta}_2^r = 0.414, \hat{\beta}_3^r = 0.695$	$\hat{\beta}_0^c = 60.214, \hat{\beta}_1^c = .0523, \hat{\beta}_2^c = -0.526, \hat{\beta}_3^c = 0.146$ $\hat{\beta}_0^r = -0.289, \hat{\beta}_1^r = 0.445, \hat{\beta}_2^r = 0.433, \hat{\beta}_3^r = 0.341$
CCRM	$\hat{\beta}_0^c = 1.886, \hat{\beta}_1^c = 0.438$ $\hat{\beta}_0^r = 2.056, \hat{\beta}_1^r = 0.0$	$\hat{\beta}_0^c = 25.979, \hat{\beta}_1^c = 0.029, \hat{\beta}_2^c = -0.289, \hat{\beta}_3^c = 0.448$ $\hat{\beta}_0^r = 2.056, \hat{\beta}_1^r = 0.0, \hat{\beta}_2^r = 0.341, \hat{\beta}_3^r = 0.608$	$\hat{\beta}_0^c = 60.214, \hat{\beta}_1^c = 0.0523, \hat{\beta}_2^c = -0.526, \hat{\beta}_3^c = 0.146$ $\hat{\beta}_0^r = 0.0, \hat{\beta}_1^r = 0.437, \hat{\beta}_2^r = 0.432, \hat{\beta}_3^r = 0.332$
CIM	$\hat{\beta}_0 = 5.619, \hat{\beta}_1 = 0.185$	$\hat{\beta}_0 = 27.589, \hat{\beta}_1 = 0.025, \hat{\beta}_2 = -0.245, \hat{\beta}_3 = 0.388$	$\hat{\beta}_0 = 56.199, \hat{\beta}_1 = 0.031, \hat{\beta}_2 = -0.416, \hat{\beta}_3 = 0.106$
PM	$\hat{\beta}_0^- = -0.096, \hat{\alpha}_1^- = 0.664, \hat{\omega}_1^- = -0.082$ $\hat{\beta}_0^+ = 4.380, \hat{\alpha}_1^+ = 0.324, \hat{\omega}_1^+ = 0.154$	$\hat{\beta}_0^- = 22.376, \hat{\alpha}_1^- = 0.147, \hat{\omega}_1^- = -0.178$ $\hat{\alpha}_2^- = -0.331, \hat{\omega}_2^- = 0.089, \hat{\alpha}_3^- = 0.472, \hat{\omega}_3^- = -0.074$ $\hat{\beta}_0^+ = 19.204, \hat{\alpha}_1^+ = 0.534, \hat{\omega}_1^+ = -0.495$ $\hat{\alpha}_2^+ = -0.737, \hat{\omega}_2^+ = -0.491, \hat{\alpha}_3^+ = -0.183, \hat{\omega}_3^+ = 0.657$	$\hat{\beta}_0^- = 63.771, \hat{\alpha}_1^- = 0.128, \hat{\omega}_1^- = -0.081$ $\hat{\alpha}_2^- = 0.077, \hat{\omega}_2^- = -0.620, \hat{\alpha}_3^- = 0.179, \hat{\omega}_3^- = -0.075$ $\hat{\beta}_0^+ = 53.141, \hat{\alpha}_1^+ = -0.311, \hat{\omega}_1^+ = 0.439$ $\hat{\alpha}_2^+ = -0.342, \hat{\omega}_2^+ = -0.213, \hat{\alpha}_3^+ = -0.151, \hat{\omega}_3^+ = 0.328$
LM$_c$	$\hat{\eta} = 0.522, \hat{\alpha}_1 = 0.617, \hat{\beta}_1 = -0.087$ $\hat{\theta} = 3.241, \hat{\gamma}_1 = 0.246$	$\hat{\eta} = 7.279, \hat{\alpha}_1 = 0.158, \hat{\beta}_1 = -0.153$ $\hat{\alpha}_2 = -0.327, \hat{\beta}_2 = 0.083, \hat{\alpha}_3 = 0.492, \hat{\beta}_3 = -0.056$ $\hat{\theta} = 4.112, \hat{\gamma}_1 = 0.0, \hat{\gamma}_2 = 0.3414, \hat{\gamma}_3 = 0.608$	$\hat{\eta} = 58.746, \hat{\alpha}_1 = 0.131, \hat{\beta}_1 = -0.044$ $\hat{\alpha}_2 = 0.085, \hat{\beta}_2 = -0.633, \hat{\alpha}_3 = 0.184, \hat{\beta}_3 = -0.044$ $\hat{\theta} = 0.0, \hat{\gamma}_1 = 0.437, \hat{\gamma}_2 = 0.432, \hat{\gamma}_3 = 0.332$
LM$_w$	$\hat{\eta} = 0.522, \hat{\alpha}_1 = 0.617, \hat{\beta}_1 = -0.087$ $\hat{\theta} = 3.241, \hat{\gamma}_1 = 0.246$	$\hat{\eta} = 17.279, \hat{\alpha}_1 = 0.158, \hat{\beta}_1 = -0.153$ $\hat{\alpha}_2 = -0.327, \hat{\beta}_2 = 0.083, \hat{\alpha}_3 = 0.492, \hat{\beta}_3 = -0.056$ $\hat{\theta} = 7.022, \hat{\gamma}_1 = -0.366, \hat{\gamma}_2 = 0.414, \hat{\gamma}_3 = 0.696$	$\hat{\eta} = 58.746, \hat{\alpha}_1 = 0.131, \hat{\beta}_1 = -0.044$ $\hat{\alpha}_2 = 0.085, \hat{\beta}_2 = -0.632, \hat{\alpha}_3 = 0.184, \hat{\beta}_3 = -0.044$ $\hat{\theta} = -0.579, \hat{\gamma}_1 = 0.445, \hat{\gamma}_2 = 0.433, \hat{\gamma}_3 = 0.341$

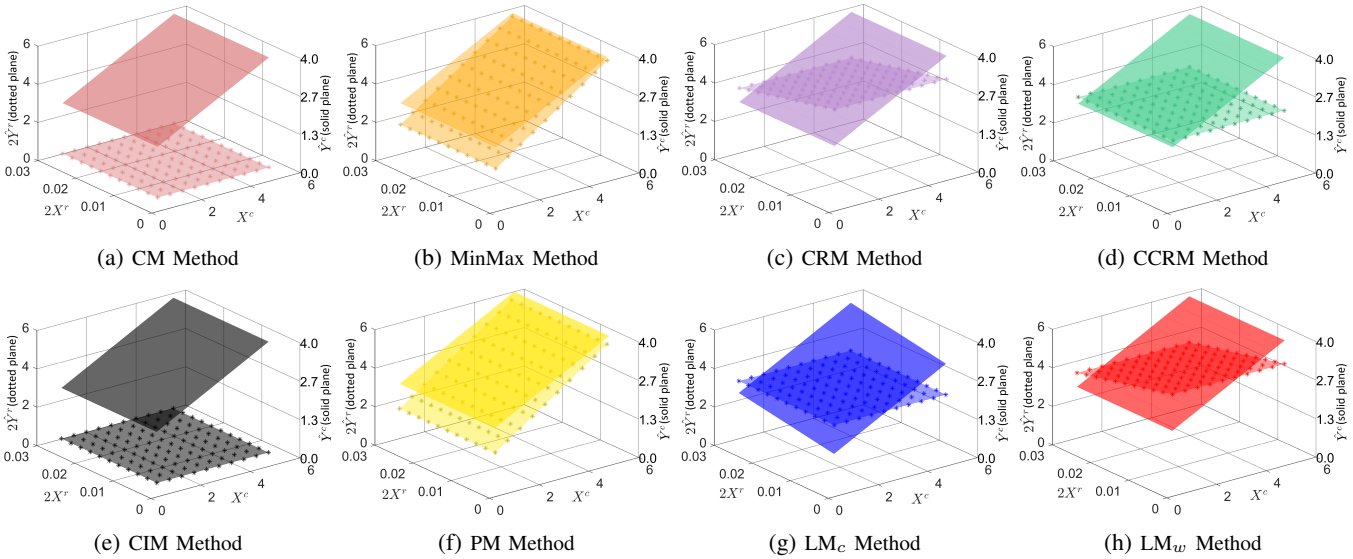
Methods	Food Snacks Purchase Intention Data Set [32]	
	Ethics and Overall purchase intention	Taste and Overall purchase intention
CM	$\hat{\beta}_0^c = 23.9, \hat{\beta}_1^c = 0.497$	$\hat{\beta}_0^c = 25.395, \hat{\beta}_1^c = 0.462$
MinMax	$\hat{\beta}_0^- = 32.164, \hat{\beta}_1^- = 0.3283, \hat{\beta}_0^+ = 18.696, \hat{\beta}_1^+ = 0.756$	$\hat{\beta}_0^- = 15.877, \hat{\beta}_1^- = 0.57, \hat{\beta}_0^+ = 36.864, \hat{\beta}_1^+ = 0.342$
CRM	$\hat{\beta}_0^c = 23.9, \hat{\beta}_1^c = 0.497, \hat{\beta}_0^r = 3.165, \hat{\beta}_1^r = 0.12$	$\hat{\beta}_0^c = 25.395, \hat{\beta}_1^c = 0.462, \hat{\beta}_0^r = 5.989, \hat{\beta}_1^r = 0.274$
CCRM	$\hat{\beta}_0^c = 23.9, \hat{\beta}_1^c = 0.497, \hat{\beta}_0^r = 3.165, \hat{\beta}_1^r = 0.12$	$\hat{\beta}_0^c = 25.395, \hat{\beta}_1^c = 0.462, \hat{\beta}_0^r = 5.989, \hat{\beta}_1^r = 0.274$
CIM	$\hat{\beta}_0 = 26.927, \hat{\beta}_1 = 0.444$	$\hat{\beta}_0 = 26.005, \hat{\beta}_1 = 0.45$
PM	$\hat{\beta}_0^- = 15.269, \hat{\alpha}_1^- = 0.055, \hat{\omega}_1^- = 0.461$ $\hat{\beta}_0^+ = 18.582, \hat{\alpha}_1^+ = -0.051, \hat{\omega}_1^+ = 0.616$	$\hat{\beta}_0^- = 17.245, \hat{\alpha}_1^- = 0.656, \hat{\omega}_1^- = -0.092$ $\hat{\beta}_0^+ = 38.472, \hat{\alpha}_1^+ = 0.264, \hat{\omega}_1^+ = 0.105$
LM$_c$	$\hat{\eta} = 13.761, \hat{\alpha}_1 = 0.062, \hat{\beta}_1 = 0.478$ $\hat{\theta} = 6.329, \hat{\gamma}_1 = 0.12$	$\hat{\eta} = 21.869, \hat{\alpha}_1 = 0.597, \hat{\beta}_1 = -0.131$ $\hat{\theta} = 11.98, \hat{\gamma}_1 = 0.274$
LM$_w$	$\hat{\eta} = 13.761, \hat{\alpha}_1 = 0.062, \hat{\beta}_1 = 0.478$ $\hat{\theta} = 6.329, \hat{\gamma}_1 = 0.12$	$\hat{\eta} = 21.869, \hat{\alpha}_1 = 0.597, \hat{\beta}_1 = -0.131$ $\hat{\theta} = 11.98, \hat{\gamma}_1 = 0.274$



(a) CM Method (b) MinMax Method (c) CRM Method (d) CCRM Method

(e) CIM Method (f) PM Method (g) LM$_c$ Method (h) LM$_w$ Method

Fig. 18: *IRG*s showing the relationship between regressor and regressand in terms of their center and range using different linear regression models for Set-1 (Fig. 1(a)).



(a) CM Method (b) MinMax Method (c) CRM Method (d) CCRM Method

(e) CIM Method (f) PM Method (g) LM$_c$ Method (h) LM$_w$ Method

Fig. 19: *IRG*s showing the relationship between regressor and regressand in terms of their center and range using different linear regression models for Set-2 (Fig. 2(a)).

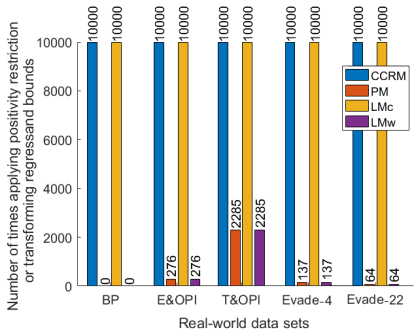


Fig. 20: Number of times in 10000 simulations the regression approaches apply positivity restriction/transformation regressand bounds for real-world cases. BP=blood pressure data set, E&OPI=*ethics* vs *overall purchase intention* data set, T&OPI=*taste* vs *overall purchase intention* data set.