Journal of Uncertain Systems
Vol.6, No.2, pp.128-137, 2012
Online at: www.jus.org.uk

Towards More Detailed Value-added Teacher Assessments: How Intervals can Help

Karen Villaverde[1], Olga Kosheleva[2,*]

[1] *Department of Computer Science, New Mexico State University, Las Cruces, New Mexico 88003, USA*

[2] *Department of Teacher Education, University of Texas at El Paso, El Paso, Texas 79968, USA*

Received 18 August 2011; Revised 12 April 2012

Abstract

Sometimes, the efficiency of a class is assessed by assessing the amount of knowledge that the students have after taking this class. However, this amount depends not only on the quality of the class, but also on how prepared were the students when they started taking this class. A more adequate assessment should therefore be *value-added*, estimating the added value that the class brought to the students.

In pedagogical practice, there are many value-added assessment models. However, most existing models have two limitations. First, they model the effect of the class as an additive factor independent on the initial knowledge. In reality, the amount of knowledge learned depends on the amount of the initial knowledge. Second, the existing models implicitly assume that the assessment values are known exactly. In reality, we usually only know bounds on the assessment values. Thus, interval techniques provide, in our opinion, a more adequate way of processing these values.

In this paper, we describe how the use of interval techniques can help us overcome both limitations of the existing value-added assessments.

Keywords: interval computations, teacher assessment, valued-added

1 Formulation of the Problem

Assessment is important. In order to improve the efficiency of education, it is important to assess this efficiency, i.e., to describe this efficiency in quantitative terms. This is important on all education levels: from elementary schools to middle and high schools to universities.

Quantitative description is needed because it allows natural comparison of different strategies of teaching and learning – and selection of the best strategy.

Traditional assessment. Sometimes, the efficiency of a class is assessed by assessing the amount of knowledge that the students have after taking this class. For example, we can take the average score of the students on some standardized test – this is actually how the quality of elementary and high school classes is now estimated in the US.

Limitation of the traditional assessment. The main problem with the traditional (outcome-only) assessment is that the class outcome depends

- not only on the quality of the class, but
- also on how prepared were the students when they started taking this class.

The idea of value-added assessment. Since the outcome depends on the initial level of students' knowledge and skills, a more adequate assessment should therefore be *value-added*, estimating the *added value* that the class brought to the students.

There exist several value-added assessment techniques. In pedagogical practice, there are many value-added assessment models; see, e.g., [1, 5, 7] and references therein.

Main idea behind the existing techniques. The main objective of these techniques is to estimate the added value. It therefore seems reasonable to evaluate this added value by subtracting the outcome from the

*Corresponding author. Email: olgak@utep.edu (O. Kosheleva).

input. For example, we can subtract the average grade after the class (on the post-test) on the average grade on similar questions asked before the class (on the pre-test).

This is, of course, the simplest possible approach. The existing techniques take into account additional parameters influencing learning. However, most existing models model the effect of the class as an additive factor independent on the initial knowledge.

Independence on the input is a limitation. In reality, the amount of knowledge learned depends on the amount of the initial knowledge. It is therefore desirable to take this dependence into account in value-added assessment.

Additional limitation. The existing models for value-added assessment implicitly assume that the assessment values are exactly known. In reality, many assessment values come from grading, and are therefore somewhat subjective. The only thing that we can conclude from these values is that the actual (unknown) knowledge quality is somewhere within the given bounds. For example, in the US grading system, a grade of A usually means that the actual quality is between 90 and 100, a grade of B means that the actual quality is between 80 and 90, etc.

What we do. In this paper, we describe how the use of interval techniques can help us overcome both limitations of the existing value-added assessments.

Comment. These results were previously presented at two conferences [9, 10].

2 Linear Dependence Instead of Addition: Idea and Examples

Traditional approach to valued-added assessment: reminder. Value-added assessment described how the post-test result y depends on the pre-test result x.

As we have mentioned earlier, in the traditional approach we, in effect, assume that the post-test result y is obtained from the pre-test result x by adding a certain amount a of new knowledge (and new skills): $y \approx x + a$. Here, we say that y is only approximately equal to $x + a$, to take into account measurement errors, random fluctuations, and the effect of factors that we do not take into account in this simple model.

Traditional approach to valued-added assessment: graphical description. To make our text easier to understand, we will try to graphically illustrate all the dependencies. In these graphical explanation, we will assume that both the pre-test assessment x and the post-test assessment y take values from the interval $[0, 1]$, with 0 corresponding to the complete lack of knowledge and 1 to perfect knowledge.

Of course, if a student's pre-test knowledge is perfect or almost perfect, there is no sense for this student to take a class. Thus, we will assume that the pre-test value x only goes until a certain threshold t.

In these terms, the additive approximate dependence can be graphically represented as follows; see Fig.1.

The actual dependence is non-additive. As we have mentioned, the actual dependence of the post-test value y on the pre-test value x is more complex, because the difference $y - x$ changes with x. To describe this dependence, we therefore need to use more general formulas than $y = x + a$.

First Approximation: Linear Dependence. The natural next approximation is to use the general linear dependence of the post-test value y on the pre-test value x: $y \approx m \cdot x + a$.

How to access the efficiency of the class under the new assessment model. With existing value-added assessment models, accessing the efficiency of a class appears straightforward: the higher the new knowledge amount a added, the better. For these models, the comparison of different teaching strategies is straightforward: we find the amount a corresponding to different strategies, and we select the strategy for which this amount a is the largest.

With the new teaching assessment models proposed in this paper, accessing the efficiency of a class is a little bit more complex, since there are more parameters now. Specifically, the resulting efficiency of different teaching strategies depends not only on the strategy itself, but also on the prior knowledge of the class. For example, for two linear functions $f_1(x) = m_1 \cdot x + a_1$ and $f_2(x) = m_2 \cdot x + a_2$ corresponding to two different teaching strategies, we may have $f_1(x_1) < f_2(x_1)$ for some x_1 and $f_1(x_2) > f_2(x_2)$ for some $x_2 > x_1$. In this case,

- for weaker students, with prior knowledge x_1, the second strategy is better;

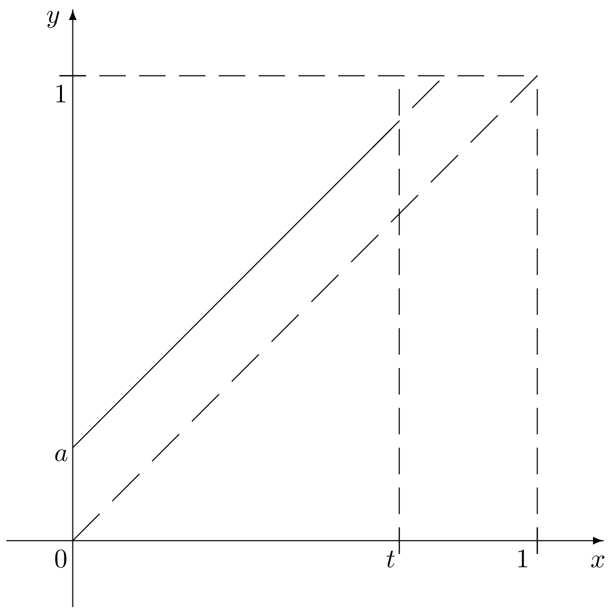

Figure 1: Additive dependence

- for stronger students, with prior knowledge $x_2 > x_1$, the first strategy is better.

Thus, the new model provides a more nuanced – and hence, more realistic – comparison between different teaching strategies.

In general, once we know the pre-test values $x_1, \ldots, x_n$ of different students of the class, we can use the known functions $f_1(x)$, $f_2(x)$, $\ldots$, describing different teaching strategies, and predict the post-test values $y_{1,j} = f_j(x_1), y_{2,j} = f_j(x_2), \ldots, y_{j,n} = f_j(x_n)$ for each strategy j. Then, for each teaching strategy j, we evaluate the value of our objective function – e.g., the mean post-test grade or a more sophisticated function (see examples below) – and select the strategy for which this value is the largest.

Linear dependence: examples. To better understand possible types of a linear dependence, let us describe and illustrate several examples of such a dependence.

Ideal case: perfect learning. In the ideal case, no matter what the original knowledge is, the resulting knowledge is perfect, $y \equiv 1$. The resulting constant function is a particular case of the general linear dependence, with $a = 1$ and $m = 0$; see Fig.2.

Example from middle schools and high schools. In many middle schools and high schools, one of the main objectives is to minimize the failure rate. Schools with high failure rate get penalized – and even disbanded.

Failure is most probable for students who start with the low starting knowledge, i.e., in our terms, with small values of x. Thus, to avoid failure, we must concentrate on the students with low x.

Since the amount of resources is limited, this means that only a few efforts are allocated to students with the originally higher level of x. As a result, the knowledge of students with a low level of x increases drastically, while the knowledge of students with high original knowledge level x does not increase that much (compared to the students with low original knowledge).

This behavior corresponds to a linear dependence in which $m < 1$. A graphical illustration of such a dependence is given in Fig.3.

Typical top school strategy. In selective top schools, the emphasis is often on the top students. Due to the limited resources, this means that the knowledge of the top students, with $x \approx t$, increases drastically, practically to perfect knowledge, while the knowledge of the bottom students does not increase that much (compared to the students with high original knowledge).

In terms of a linear dependence, this means that we take $m > 1$. A graphical illustration of such a dependence is given in Fig.4.

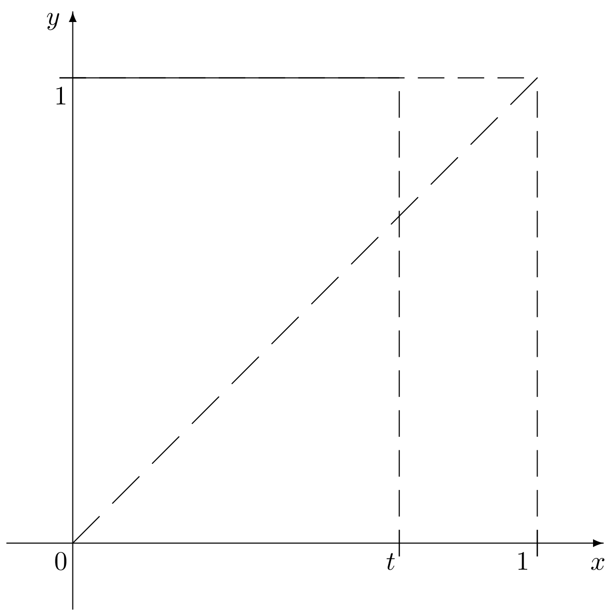

Figure 2: Ideal case: perfect learning

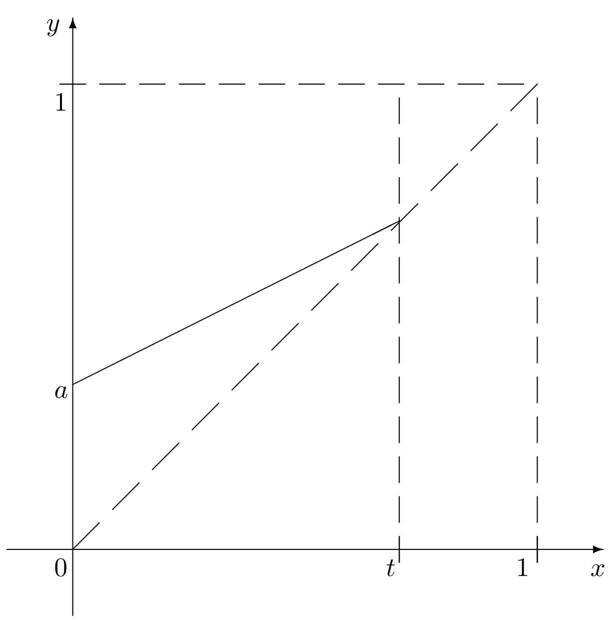

Figure 3: Case of $m < 1$

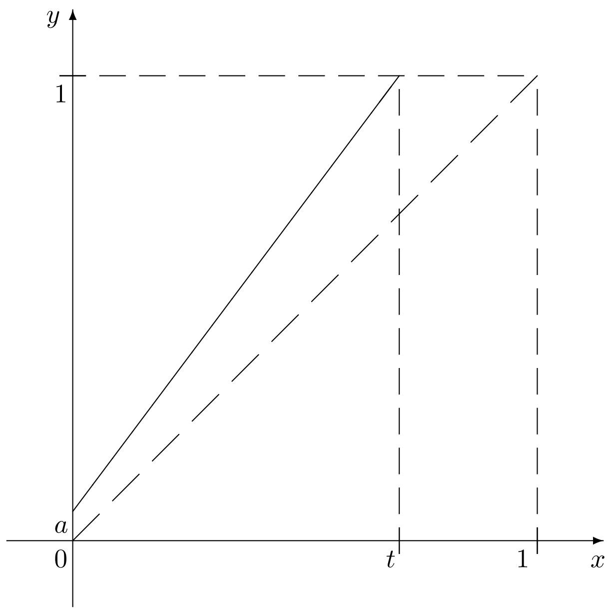

Figure 4: Case of $m > 1$

3 How to Determine the Coefficients m and a: Case of Exactly Known x_i

Problem: reminder. In the linear model, to quantitatively describe the success of the learning process, we must determine the parameters m and a of the corresponding linear dependence $y \approx m \cdot x + a$.

Let us start with a simple case when we know the exact values of the pre-test grades $x_1, \ldots, x_n$ and the exact values of the post-test grades $y_1, \ldots, y_n$. In this case, the problem is to find the values m and a for which $y_i \approx m \cdot x_i + a$ for all $i = 1, 2, \ldots, n$.

Natural idea: use least squares method. We would like to make all n differences $e_i \stackrel{\text{def}}{=} y_i - (m \cdot x_i + a)$ close to 0. In other words, we want the vector $e = (e_1, \ldots, e_n)$ to be as close to the 0 vector $0 = (0, \ldots, 0)$ as possible.

The Euclidean distance between the vectors e and 0 is equal to $\sqrt{\sum\limits_{i=1}^{n} e_i^2}$. Thus, the vector is the closest to 0 if this distance (or, equivalently, its square) is the smallest possible: $\sum\limits_{i=1}^{n} e_i^2 \to \min$, or, equivalently, $\sum\limits_{i=1}^{n} (y_i - (m \cdot x_i + a))^2 \to \min_{m,a}$. The resulting Least Squares approach is a standard approach in statistics; see, e.g., [8], because it is the optimal approach when the error are independent and normally distributed.

There exist good algorithms for solving the Least Squares problem. In our case, we can simply differentiate the minimized expression with respect to the unknowns m and a and equate the resulting derivatives to 0. As a result, we get explicit formulas for m and a: $m = C(x,y)/V(x)$, where

$$C(x,y) \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - E(x)) \cdot (y_i - E(y)); \quad V(x) \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - E(x))^2;$$

$$E(x) \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \quad E(y) \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} y_i; \quad a = E(y) - m \cdot E(x).$$

4 Case of Interval Uncertainty: Analysis

Case of interval uncertainty: description. In the previous section, we assumed that we know the numerical grade on the exam represents an exact measure of the student knowledge. In practice, however, the number grades are reasonably subjective.

Usually, instructors allocate certain number of points to different questions and problems on the test and to different aspects of the same question or problem. As a result, when the answer to each of the problems or questions is either absolutely correct or absolutely wrong (or missing), the resulting grade is uniquely determined. The subjectivity comes when the answer is partly correct, and we need to decide how much partial credit this answer deserves. Some such situations can be described from the very beginning, but often, it is not practically possible to foresee all possible mistakes and thus, to decide how much partial credit the student deserves.

Often, when two instructors co-teach a class or teach two different sections of the same class, their grades for similar mistakes can slightly differ – because of the slightly different allocation of partial credit. Even the same instructor, when grading two different student papers with similar mistakes, can sometimes assign two slightly different numerical grades.

As a result of this subjectivity, the numerical grade given to the test is not an exact measure of the student knowledge – because other instructors may assign a slightly different number grade to the same test results.

This subjectivity is well understood by instructors. This is one of the reasons why student transcripts usually list not the exact overall number grades, but rather the letter grades.

For example, usually, a letter grade A is assigned to all the numerical grades from 90 to 100, and a letter grade B is assigned to all numerical grades between 80 and 89. This assignment is in good accordance with the fact that

- while the difference between, say, 85 and 95 is meaningful and most probably not subjective, and a student with a grade of 95 has a higher knowledge level than a student with a grade of 85,
- the difference between, say 92 and 93 can be caused by the subjective reasons – and thus, a student with a grade of 93 does not necessarily know the material better than a student with a grade of 92.

The traditional letter grades may provide too crude a picture. In many cases, the distinction between, say, low 90s and high 90s also makes sense. To emphasize such a difference, some schools, in addition to usual letter grades, also use signed letter-type grades like A− or B+. Letter grades from the resulting set correspond to intervals which are narrower than the width-10 intervals describing the usual letter grades.

Because the distinction within each interval may be caused by the subjectivity of an individual instructor grading, it makes sense, when describing how well the students learned, to use not the original numerical grades x_i, but rather the corresponding letter grades – i.e., in other words, the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ that describe possible values of the student knowledge.

Comment. Education is, of course, not the only area where intervals appear. Intervals appear in many measurement situations where we only know the upper bound Δ_i on the measurement inaccuracy $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$, i.e., on the difference between the measurement result $\widetilde{x}_i$ and the actual (unknown) value x_i of the measured quantity: $|\Delta x_i| \le \Delta_i$. In such cases, the only information that we have about the desired value x_i is that this value belongs to the interval $\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

There exist many algorithms for processing such interval uncertainty; see, e.g., [2, 3, 4, 6].

How to describe dependence between x and y under interval uncertainty. We want to describe a dependence between the pre-test grade x and the post-test grade y.

In the crisp case, we have an exact grade x and we want to predict the exact grade y. In the ideal case, to every value x, we would like to assign the corresponding value y. In mathematical terms, this means that we would like to have a *function* $y = f(x)$ that maps numbers (= pre-test grades) into numbers (= post-test grades).

In the interval case, we start with an interval pre-test grade $\mathbf{x}$, and we would like to predict the interval post-test grade $\mathbf{y}$. Thus, to every interval value $\mathbf{x}$, we would like to assign the corresponding interval value $\mathbf{y}$. In mathematical terms, this means that we would like to have a *function* $\mathbf{y} = f(\mathbf{x})$ that maps intervals (= pre-test grades) into intervals (= post-test grades).

Number-to-number case. To analyze which interval-to-interval functions f can represent the map from pre-test to post-test intervals, let us first consider a simplified situation, in which,

- for each student, we know the pre-test and post-test grades which exactly describe the student's knowledge, and
- the student's post-test grade is uniquely determined by his or her pre-test grade.

In this simplified situation, due to uniqueness, the dependence between the student's pre-test grade x and his or her post-test grade y can be described by a number-to-number function $f(x)$.

In principle, an arbitrary mathematical mapping from real numbers to real numbers can occur in real learning. One might argue that we probably should require that $f(x) \geq x$, since the knowledge at the end cannot be smaller than the starting knowledge.

In reality, however, even this requirement is not necessary: people forget, so it is quite possible that without repetitions, some students will score much worse on a post-test than on the pre-test.

For intervals, there are additional restrictions on interval-to-interval functions: example. In our more realistic description, we do not know the exact value of the characteristic describing the student's pre-class and post-class knowledge; instead, for each student, we only know:

- the pre-test interval grade $\mathbf{x}$ that describes the possible values of the student's pre-class knowledge, and
- the post-test interval grade $\mathbf{y}$ that describes the possible values of the student's post-class knowledge.

Within this description, an interval-to-interval function $f(\mathbf{x})$ describes the set of all possible post-test grades for all the students who pre-test grades are within the interval $\mathbf{x}$.

Let us show that, in contrast to a number-to-number case where every mathematical number-to-number function could be potentially interpreted as a pre-test-to-post-test function f, in the more realistic interval-to-interval case, not all mathematically possible interval-to-interval functions can be thus interpreted: only interval-to-interval functions that satisfy a certain restriction can be interpreted as pre-test-to-post-test functions.

To explain this restriction, let us start with a simple example. Suppose that we know that

- when the pre-test grades are from the interval $\mathbf{x}_1 = [80, 90]$, then the post-test grade is from the interval $\mathbf{y}_1 = f(\mathbf{x}_1) = [85, 95]$; and
- when the pre-test grades are from the interval $\mathbf{x}_2 = [90, 100]$, then the post-test grade is from the interval $\mathbf{y}_2 = f(\mathbf{x}_2) = [92, 100]$.

What if now we have a student whose pre-test grade is between 80 and 100, i.e., for whom $\mathbf{x} = [80, 100]$.

In general, a mathematically defined interval-to-interval function can have any value of $f(\mathbf{x}) = f([80, 100])$, a value which is not necessarily related to the values $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$. For example, from the purely mathematical viewpoint, we can have $f([80, 100]) = [50, 100]$.

However, in our case, the value $f(\mathbf{x})$ has a meaning – it is the set of all possible post-test grades of all the students whose pre-test grades are in the interval $\mathbf{x}$. Let us show that this meaning imposes a restriction on the possible interval-to-interval functions. Indeed, the interval $\mathbf{x} = [80, 100]$ is a union of the previous two intervals $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2$, meaning that the student who has the actual pre-test grade in the interval $\mathbf{x}$ either has the actual grade between 80 and 90, or between 90 and 100.

- In the first case, when $x \in \mathbf{x}_1$, we expect that the final grade y is in the corresponding interval $\mathbf{y}_1$.
- In the second case, when $x \in \mathbf{x}_2$, we expect that the final grade y is in the corresponding interval $\mathbf{y}_2$.

Thus, we can conclude that y belongs either to the interval $\mathbf{y}_1$ or to the interval $\mathbf{y}_2$, i.e., that it belongs to the union

$$\mathbf{y}_1 \cup \mathbf{y}_2 = f(\mathbf{x}_1) \cup f(\mathbf{x}_2) = [85, 95] \cup [92, 100] = [85, 100]$$

of these two intervals. Thus, for the pre-test interval $\mathbf{x} = \mathbf{x}_1 \cup \mathbf{x}_2$, the set $f(\mathbf{x})$ of all possible values of post-test grades should be equal to

$$f(\mathbf{x}) = f(\mathbf{x}_1) \cup f(\mathbf{x}_2) = [85, 100].$$

For intervals, there are additional restrictions on interval-to-interval functions: general formulas. In general, the pre-test-to-post-test mapping f from intervals to intervals must satisfy the following property: $f(\mathbf{x}_1 \cup \mathbf{x}_2) = f(\mathbf{x}_1) \cup f(\mathbf{x}_2)$. Similar argument leads us to the conclusion that $f(\mathbf{x}) = \bigcup\limits_{x \in \mathbf{x}} f([x,x])$.

Towards a description of all interval-to-interval functions that satisfy the above property: analysis. According to the above formula, to describe a pre-test-to-post-test interval-to-interval function $f(\mathbf{x})$, it is sufficient to describe a numbers-to-intervals function $f([x,x])$ corresponding to degenerate intervals of the type $[x,x]$. For each such degenerate value, let us denote the lower endpoint of the interval $f([x,x])$ by $\underline{f}(x)$, and its upper endpoint by $\overline{f}(x)$.

In these terms, the interval $f([x,x])$ corresponding to a degenerate interval $[x,x]$ has the form $f([x,x]) = \left[\underline{f}(x), \overline{f}(x)\right]$. Thus, we have $f([\underline{x}, \overline{x}]) = \bigcup\limits_{x \in [\underline{x}, \overline{x}]} \left[\underline{f}(x), \overline{f}(x)\right]$.

When the take the union of intervals, we thus take the minimum of their lower endpoints, and the maximum of their upper endpoints. Thus, the union $f([\underline{x}, \overline{x}])$ has the form $f([\underline{x}, \overline{x}]) = [\underline{y}, \overline{y}]$, where $\underline{y} = \min\limits_{x \in [\underline{x}, \overline{x}]} \underline{f}(x)$ and $\overline{y} = \max\limits_{x \in [\underline{x}, \overline{x}]} \overline{f}(x)$.

Usually, the better the original knowledge, the better the results. Thus, both function $\underline{f}(x)$ and $\overline{f}(x)$ should be increasing with x. So, the minimum of $\underline{f}(x)$ is attained at the smallest possible value $x = \underline{x}$ and the maximum of $\overline{f}(x)$ is attained at the largest possible value $x = \overline{x}$.

Thus, we arrive at the following formula:

Resulting formula. $f([\underline{x}, \overline{x}]) = \left[\underline{f}(\underline{x}), \overline{f}(\overline{x})\right]$.

5 Case of Interval Uncertainty: Algorithm

What we want to compute: a reminder. In the interval case, we want to find two functions $\underline{f}(x)$ and $\overline{f}(x)$ for real numbers to real numbers.

We consider the case of a linear dependence, so we assume that both functions are linear: $\underline{f} = \underline{m} \cdot x + \underline{a}$ and $\overline{f} = \overline{m} \cdot x + \overline{a}$.

These two linear functions must satisfy the following condition for all x: $\underline{f}(x) \le \overline{f}(x)$, i.e., $\underline{m} \cdot x + \underline{a} \le \overline{m} \cdot x + \overline{a}$. Since the functions are linear, it is sufficient to require that these conditions be satisfied for $x = 0$ and $x = t$, i.e., that $\underline{a} \le \overline{a}$ and $\underline{m} \cdot t + \underline{a} \le \overline{m} \cdot t + \overline{a}$.

Example. A graphical illustration of the corresponding two functions is given on Fig.5.

Comment. On this example, we see that it is possible to have $\underline{m} > \overline{m}$. Thus,

- while the corresponding estimates $\underline{a}$ and $\overline{a}$ for a do satisfy the inequality $\underline{a} \le \overline{a}$ and thus, form an interval $[\underline{a}, \overline{a}]$,
- the estimates $\underline{m}$ and $\overline{m}$ for m do not necessarily form an interval.

Estimating $\underline{m}$, $\underline{a}$, $\overline{m}$, and $\overline{a}$: a problem. We are given intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ and $\mathbf{y}_i = [\underline{y}_i, \overline{y}_i]$. We would like to find the values of the parameters $\underline{m}$, $\underline{a}$, $\overline{m}$, and $\overline{a}$ for which $\mathbf{y}_i \approx f(\mathbf{x}_i)$ for the corresponding interval-to-interval function f.

Due to the above representation of a general linear interval-to-interval function, this means that for every i, we must have $\underline{y}_i \approx \underline{m} \cdot \underline{x}_i + \underline{a}$ and $\overline{y}_i \approx \overline{m} \cdot \overline{x}_i + \overline{a}$.

Estimating $\underline{m}$, $\underline{a}$, $\overline{m}$, and $\overline{a}$: analysis. We see that we have, in effect, two independent sets of approximate equalities:

- to find $\underline{m}$ and $\underline{a}$, we use approximate equalities $\underline{y}_i \approx \underline{m} \cdot \underline{x}_i + \underline{a}$;
- to find $\overline{m}$ and $\overline{a}$, we use approximate equalities $\overline{y}_i \approx \overline{m} \cdot \overline{x}_i + \overline{a}$.

For each set of approximate equalities, we can apply the same Least Squares approach as we described for the case of crisp estimates. As a result, we arrive at the following formulas.

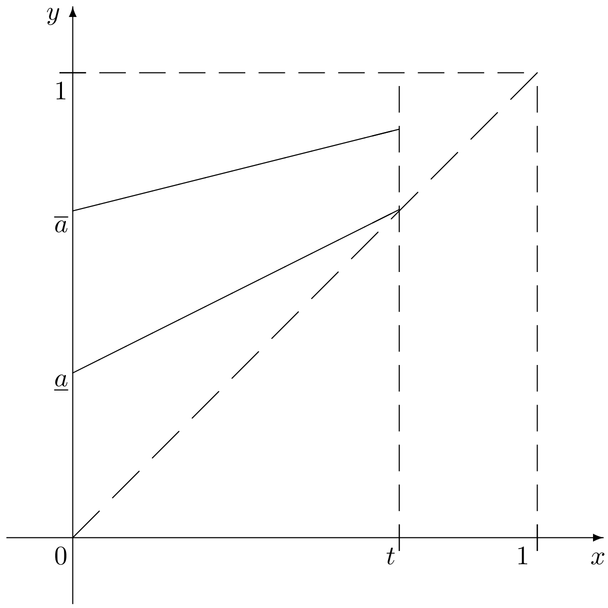

Figure 5: Case of interval uncertainty

Resulting algorithm. First, we compute $\underline{m} = C(\underline{x},\underline{y})/V(\underline{x})$, where

$$C(\underline{x},\underline{y}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}(\underline{x}_i - E(\underline{x}))\cdot(\underline{y}_i - E(\underline{y}));\quad V(\underline{x}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}(\underline{x}_i - E(\underline{x}))^2;$$

$$E(\underline{x}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}\underline{x}_i;\quad E(\underline{y}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}\underline{y}_i;\quad \underline{a} = E(\underline{y}) - \underline{m}\cdot E(\underline{x}).$$

Then, we compute $\overline{m} = C(\overline{x},\overline{y})/V(\overline{x})$, where

$$C(\overline{x},\overline{y}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}(\overline{x}_i - E(\overline{x}))\cdot(\overline{y}_i - E(\overline{y}));\quad V(\overline{x}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}(\overline{x}_i - E(\overline{x}))^2;$$

$$E(\overline{x}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}\overline{x}_i;\quad E(\overline{y}) \stackrel{\rm def}{=} \frac{1}{n}\cdot\sum_{i=1}^{n}\overline{y}_i;\quad \overline{a} = E(\overline{y}) - \overline{m}\cdot E(\overline{x}).$$

Acknowledgments

The authors are greatly thankful to all the participants of the IEEE World Congress on Computational Intelligence WCCI'2010 (Barcelona, Spain, July 18–23, 2010) and the 14th GAMM-IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics SCAN'2010 (Lyon, France, September 27–30, 2010) and for valuable discussions.

References

[1] Fincher, C., What is value-added education?, *Research in Higher Education*, vol.22, no.4, pp.395–398, 1985.

[2] Interval computations website http://www.cs.utep.edu/interval-comp.

[3] Jaulin, L., M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[4] Kreinovich, V., A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1998.

[5] McCaffrey, D.F., J.R. Lockwood, D.M. Koretz, and L.S. Hamilton, *Evaluating Value-added Models for Teacher Accountability*, RAND Corporation, Santa Monica, California, 2003.

[6] Moore, R.E., R.B. Kearfott, and M.J. Cloud, *Introduction to Interval Analysis*, SIAM Press, Philadelphia, Pennsylviania, 2009.

[7] Sanders, W.L., and S.R. Horn, The Tennessee Value-Added Assessment System (TVAAS): mixed-model methodology in educational assessment, *Journal of Personnel Evaluation in Education*, vol.8, no.9, pp.299–311, 1994.

[8] Sheskin, D.J., *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2004.

[9] Villaverde, K., and O. Kosheleva, Towards more detailed value-added teacher assessments, *Proceedings of the IEEE World Congress on Computational Intelligence WCCI'2010*, Barcelona, Spain, pp.3126–3133, 2010.

[10] Villaverde, K., and O. Kosheleva, Towards more adequate value-added teacher assessments: how intervals can help, *Abstracts of the 14th GAMM-IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics SCAN'2010*, Lyon, France, pp.140–141, 2010.