



# Game-Theoretic Concept for Determining the Price of Time Series Data

J. Mayer<sup>1</sup>(✉), T. Kaufmann<sup>1</sup>, P. Niemietz<sup>1</sup>, and T. Bergs<sup>1,2</sup>

<sup>1</sup> Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany  
j.mayer@wzl.rwth-aachen.de

<sup>2</sup> Fraunhofer Institute for Production Technology IPT, Steinbachstraße 17, 52074 Aachen, Germany

**Abstract.** The digitization of manufacturing and the economic potential from the utilization of data sets result in the monetization and trading of process data. In data marketplaces, a suitable pricing mechanism ensures the definition of a fair value of data for the data consumer and the data owner. Game theoretic concepts ensure this fairness by determining the value functions of the participants in a transaction, which is based on the quality and quantity of data. In this paper, cooperative games are used to model data transactions. First, empirical methods are used to realistically determine the value of data sets. Using the example of a data set of a grinding machine, a function is set up depending on the quality dimensions, with the help of which declarations about the usability of the data set for regression models are possible. This is followed by the pricing of a data set based on the Kalai-Smorodinsky solution. It offers the advantage that the quality of a data set does not have to be assumed as a variable and that pricing is possible individually for already existing data sets whose quality is not to be adjusted further. Finally, the chosen approach is compared with the common method for price discovery, the Stackelberg game, with respect to the concession of the actors, based on the negotiation result to the maximum possible benefit. In the outlook, potential further developments of the approach based on Kalai-Smorodinsky are discussed.

**Keywords:** Data monetization · Data pricing · Game theory

## 1 Governance

The governance of a data marketplace deals with the applicable decision rights as well as the formal and informal control mechanisms on the sharing platform [1]. In addition to defining the technical characteristics of the platform, it regulates the cooperation and competition among marketplace participants through established principles [2]. The effective mechanisms of platform governance to maintain data security and integrity incentivize the sharing of authentic data. Monitoring of platform input and output and pricing of provided assets complete platform governance. The pricing of data is subject to various challenges, making it difficult for data sharing to reach market maturity. In this

paper, following the introduction to the challenges of data pricing and the fundamentals of game theory, a concept for pricing manufacturing data based on data quality for use in machine learning (ML) models is presented. This concept, which is presented on the example of data of an industrial grinding machine, forces the transferability of a data pricing method for any equipment of the production engineering for the use case of a ML model of the linear regression.

## 2 Basics of Pricing

Fair pricing in the context of industrial data requires provision for the benefits of both seller and buyer, fair consideration of the increments of benefits, respect for data quality and equal bargaining power of both parties, and maximum willingness to pay of the buyer and cost of the seller. Game-theoretic approaches allow observing the requirements. One method widely used is the Stackelberg game, which is briefly described in Sect. 2.1.

### 2.1 Stackelberg-Game

In the simplest case, the Stackelberg game is a strategic game with two players, in which the so-called Stackelberg leader (data owner) makes his decision first and the Stackelberg follower (data consumer) decides with reference to it [3]. Since multiple data owners participate in a data marketplace and they offer their data through the data marketplace operator, the data marketplace operator can be used as a Stackelberg leader to reduce complexity [4].

The Stackelberg game proceeds in two stages. In the first, the marketplace operator determines the pricing strategy, after which the data consumer defines its buying strategy in the second stage. The solution to the Stackelberg game represents a Nash equilibrium. Value functions are needed to determine the payoff of participants at different quantities and quality levels of data at different prices. A data consumer's value depends first on the price of the data set and second on the quantity and quality of the data. The value of the data owner is increased by the selling price and decreased by the cost of data aggregation and preparation.

The Stackelberg game has certain drawbacks when it comes to pricing. Designed to model a competitive situation in a market, the Stackelberg game is less well suited for cooperative interactions between data owner and data consumer [5], in which both parties each seek a contract conclusion, as opposed to strict competition. Without this, neither can generate value. Additionally, players are disadvantaged depending on the turn order. As soon as several players compete for the quantity of a good, the Stackelberg leader gets an advantage over the follower [6]. Another disadvantage is the assumption of rationality among the players. This rationality leads to an attempt to anticipate the strategy of the follower, so that suboptimal payoffs may result [7]. The cooperative game promises to remedy the drawbacks of the Stackelberg game and is described below.

## 2.2 Cooperative Games

In cooperative games, players can make binding agreements among themselves [5]. An outside entity or contract ensures that these agreements are enforced and a strategy is chosen. A cooperative game is defined by the set of potential players  $N$ , the payoff space  $P$  with payoff vectors  $u = (u_1, \dots, u_n)$ , and the conflict point  $c$ . For the price negotiation for an economic good, this conflict point represents a lack of agreement and is defined as  $c = (0/0)$ . There is a bargaining problem if there exists a  $u \in P$  that promises a higher payoff than the conflict point  $c$  for all players. For a bargaining problem between two players with value functions  $u_1, u_2$  and the conflict point  $c = (c_1, c_2)$  with player-specific payoffs  $c_1, c_2$ , the problem can be represented two-dimensionally using the value limit  $H(P)$  of the bargaining space. The value limit  $H(P)$  is the set of all pareto-optimal payoff pairs in  $P$  on which the optimal payoff pair lies. This can be determined, for example, using the Kalai-Smorodinsky (KS) solution. This solution, unlike the Nash approach, allows the comparison of individual value gains [8]. Both the Stackelberg game and the cooperative game are based on the value-based pricing of data. The value of data in the present context refers to its use in ML models, the outcome of which depends heavily on the quality of the data with which the algorithm is trained, in addition to the quantity [9]. The quality of the prediction of an ML model increases when the data quality increases. The latter will be discussed in more detail in the following.

## 2.3 Data Quality

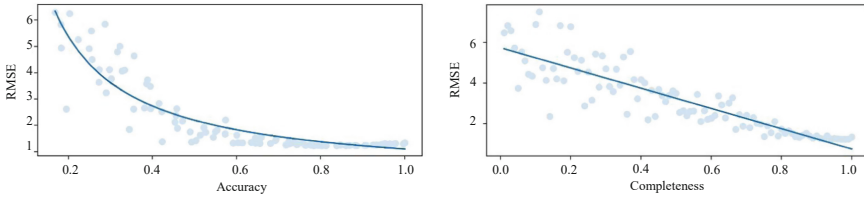
Data quality is characterized by multidimensionality. Quality dimensions (QD) can be divided into objective and subjective. Objective QD are considered measurable, while subjective dimensions depend heavily on the preferences of the data consumer and cannot be measured directly [10]. Objectively measurable criteria according to Wang and Strong include accuracy and completeness [11], which are the focus of this paper. There are different ways to model the overall quality from the different influencing factors. Yu and Zhang recommend the principle of integrated quality for the use case of pricing, in which the influence of the QD on each other is represented [12]. Assuming that the QDs negatively influence each other to a certain degree, the result is:

$$q_{im}^I = q_{i(m-1)}^I + q_{im}(1 - q_{i(m-1)}^I) \quad (1)$$

It is difficult to validate a relationship between QD and the resulting overall quality of a dataset because there is no metric for capturing or quantifying the overall quality of a dataset. Therefore, similar to Niyato et al. [3], this work makes the assumption that a data set is only as high quality as the results it can produce. To this end, the effects of data quality on the goodness of a regression model are examined. The underlying data set represents an external cylindrical grinding process, in which shafts were machined by plunge grinding. By varying selected grinding parameters (workpiece hardness, cooling lubricant exit velocity, bond hardness of the grinding tool), the effects on surface roughness, circularity deviation and shaft end diameter of the workpiece as well as radial wear of the grinding wheel were investigated.

To investigate the functional relationship between the QD and the quality of the ML model, specially developed for this application, the expression of the quality dimensions

is degraded incrementally and randomly. After each iteration, the model goodness is recalculated using the root mean square error (*RMSE*). The determined data points are visualized by the light blue color and the course of the point cloud is approximated by functions. The exact shape of the curves depends on parameters such as the original quality of the data set used and the correlation between the independent variables and the target variable (see Fig. 1). Therefore, no general correlation can be formed here.



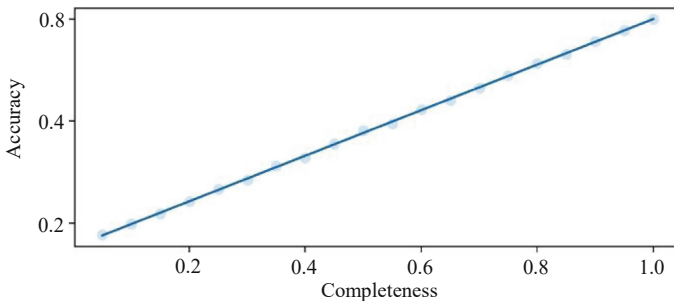
**Fig. 1.** Correlation between *RMSE* and QD accuracy and completeness

A hyperbolic course of the accuracy-*RMSE* or a linear course of the completeness-*RMSE* is assumed for the definition range of the QD on [0, 1]. The following functions with the constants  $\alpha, \beta, \gamma, \delta$  are obtained:

$$RMSE_A = \frac{1}{\alpha A + \beta} \tag{2}$$

$$RMSE_C = \gamma - \delta C \tag{3}$$

Before a functional dependence between the quality of the ML model and the QD can be found, the influence of the QD among each other shall be considered similar to Yu and Zhang [12]. For this purpose, the completeness is reduced gradually, and the resulting accuracy of the data set is calculated after each step. A linear relationship between accuracy and completeness follows (see Fig. 2).



**Fig. 2.** Correlation between accuracy and completeness

To map the functional relationship between the quality of the ML model and the QD of the data set, further conditions have to be defined. The quality of the regression model

shall be defined on the interval  $[0, 1]$ , where 1 represents a model of highest quality. This means that the functions on the interval  $[0, 1]$  should be monotonically increasing. Thus, a change of sign is required for the functions  $RMSE_A$  and  $RMSE_C$ , since the functions represent the course of the model quality when the quality dimension decreases, which is why they decrease monotonically. Consequently, the quality of the model should grow with increasing QD. Furthermore, for all partial functions  $g$  of which the quality function  $Q$  is composed,  $g(1) = 1$  shall hold. This is necessary, since the quality of the ML model should take a value of 1 at maximum QD. To ensure this, the function  $RMSE_A$  must be extended by a term  $\varphi$ , which shifts the function, but without influencing the slope of the function.

1. The following applies to the accuracy as a function of the completeness:

$$A(C) = C \tag{4}$$

2. For the quality of the regression model applies as a function of accuracy:

$$Q(A) = -\frac{1}{\alpha A + \beta} + \varphi \tag{5}$$

3. For the impact of completeness on the quality of the ML model holds:

$$Q_C(C) = C \tag{6}$$

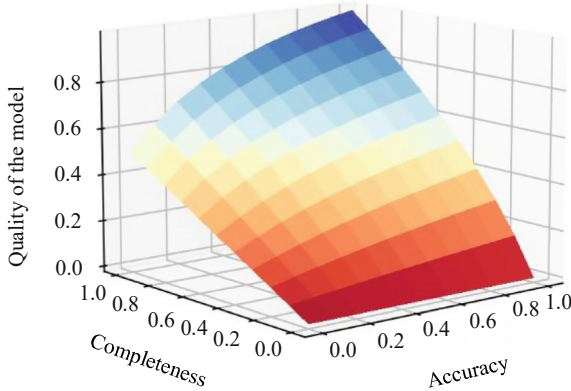
Consequently, the accuracy is negatively affected if the completeness is not equal to 1. For this reason, the expression  $A(1 - C)$  is subtracted from the accuracy of the data set. The result is a quality function of the ML model depending on the accuracy and completeness with the normalization factor of 0.5:

$$q(A, C) = \left[ \left( -\frac{1}{\alpha A - A(1 - C) + \beta} + \varphi \right) + C \right] \frac{1}{2} \tag{7}$$

This three-dimensional function is visualized in Fig. 3.

### 2.4 Research Gap

Pricing models can be distinguished as static and dynamic. Static methods are characterized by a lack of consideration of the individuality of data sets [13]. Dynamic methods, on the other hand, take into account not only the individual metrics of the various data sets, but also the demands on the data and the price of the data consumer and data owner. Game theoretic approaches, for example, can be classified as dynamic pricing models. A screening of existing scientific approaches illustrates the increasingly relevant topic of a data marketplace and the lack of a holistic evaluation of existing methods. Some authors push the correlation between data quality and ML model accuracy as the driving element of pricing [4], others the data quantity and model accuracy [14]. The Stackelberg game, for example, models the price of a data set after determining a value function [4]. Existing approaches currently lack reference to the pricing of manufacturing data, consideration



**Fig. 3.** Graphical illustration of the quality function

of multidimensional data quality for the use case in ML models, and multidimensionality of the value function with respect to data quality and quantity. Pricing is subject to challenges typical of mechanical engineering, such as the process-related boundary conditions, the accuracy of the measurement results of the sensors, sensor calibration and position, and the diversity of the technical data itself, which increase the comparability between individual data sets and thus the difficulty of pricing. In this paper, the first two aspects of the research gap (pricing of manufacturing data, multidimensionality of data quality) are addressed.

### 3 Value Function

Regardless of the game-theoretic pricing approach, value functions are needed for both data consumers and data owners. To define these functions, manufacturing companies were surveyed in several stages. In the first part of this survey, the value of the data consumer was determined using a method according to Halter and Mason [15]. The second part of the survey used selected questions to determine the respondents’ risk aversion to monetary gains from data sales. Due to the subjectivity of the respondents, a generally valid value function cannot be determined. Respondents were presented with different courses of action at different payoffs  $a, b, c, d$ , where  $a > b > c > d$ . Within the equally likely situations, an indifference value is to be chosen for which a decision maker considers two courses of action to be equivalent (Table 1). In the present case, the payoffs represent data of different quality levels from 0 to 1, where the data are identical in content, size, and all other metrics. The chosen indifference value is taken as a given value for the next game.

To generate a value function graphically requires two anchor points,  $[0, u(0)]$  and  $[c', u(c')]$ , which can be chosen arbitrarily, and a scaling point [15]. To determine the scaling point, the value of any indifference value can be freely chosen. It follows:

$$u(c) = \frac{1}{2}u(c') \tag{8}$$

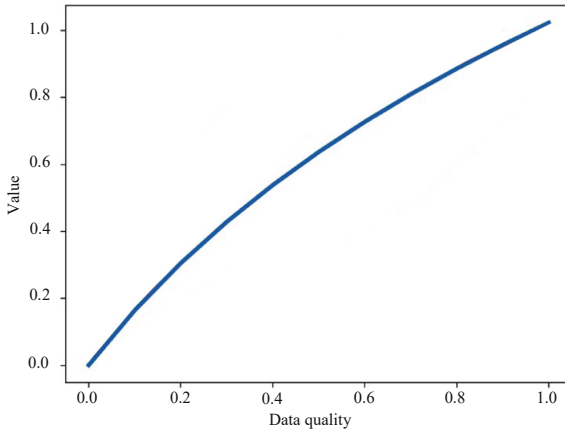
**Table 1.** First part of the survey

Game 1		Game 2	
Decision-making options		Decision-making options	
A1	A2	A1	A2
E1	d	E1	Indifference value 1
E2	a	E2	a
Equiprobableevents		Equiprobableevents	
Game 3		Game 4	
Decision-making options		Decision-making options	
A1	A2	A1	A2
E1	0	E1	Indifference value 1
E2	a'	E2	a''
Equiprobableevents		Equiprobableevents	
		Indifference value 2	
		Indifference value 3	

$$u(a') = u(c) + u(c') - u(d) \tag{9}$$

$$u(a'') = u(a') + u(c') - u(c) \tag{10}$$

Based on the determined four indifference values, the value functions can be created. It holds that  $u(d) = 0$  since the value  $d$  in the survey is zero. For the further procedure, the value functions are approximated by a function (dark blue graph), which is subsequently the basis for the value function of the data consumer (Fig. 4).



**Fig. 4.** Graphical illustration of the value functions of the first part of the survey

For the value function of the data owner, the value generated by monetary gains and consequently the Arrow-Pratt measure of risk aversion with regard to negotiations with data sets is relevant. With the help of a survey, individual negotiation situations are simulated at defined probabilities of occurrence (Table 2), in which the participants can either accept a fixed offer for the data set or negotiate with another company for the price. In the negotiation case, there is also the potential situation of a breakdown without monetary payout.

Calculating the expected value of a negotiation and comparing it to the fixed payoff serves to define risk behavior. Choosing a low expected value symbolizes risk affinity. In the underlying survey, all participants exhibited risk-neutral to risk-averse behavior (concave course of the value function). For the further procedure, the commonly used isoelastic value functions (CRRA) represent this concavity well [16]. It holds:

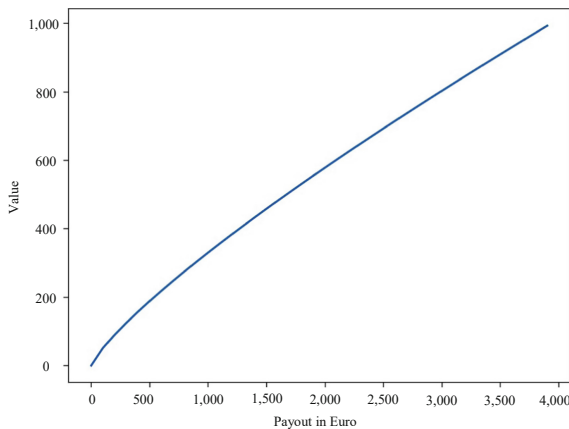
$$u(x) = x^{1-r}, r \geq 0 \tag{11}$$

Table 2 can now be used to calculate the measure of risk aversion of the respondents [17]. Due to the small size of the survey and the small distance between the individual values of risk aversion, the mean value is used ( $\bar{r} = 0.191$ ). In Fig. 5 the course of the function  $u(x) = x^{1-r}$  is drawn, which results for  $\bar{r}$ .



**Table 2.** Second part of the survey

Situation	Height fixed offer	Negotiation		Number of respondents		
		Successful	Not successful	Fixed offer	Negotiation	Indifferent
1	400 €	1000 € (50%)	0 € (50%)	–	5	1
2	300 €	1000 € (50%)	–250 € (50%)	–	6	–
3	100 €	1000 € (25%)	–250 € (75%)	6	–	–
4	6 €	1000 € (35%)	–400 € (65%)	1	5	–



**Fig. 5.** Graphical illustration of the value functions of the second part of the survey

To determine the progression and to maintain the comparability of the value functions of data owner and data consumer, the value is normalized to the interval [0, 1]. Otherwise, the price would influence the value functions excessively. In the normalization, the ideal value of the datum is ascribed the greatest possible value. In the function shown in the figure, this price is symbolically 4000 €. However, it can be captured realistically by a survey. The defined functions serve as a basis for finding the value functions for data owner and data consumer. When setting up the value functions, other aspects such as the costs of data aggregation are taken into account in addition to the empirically determined function curves. The following functions form the cornerstone for pricing using game-theoretic concepts of any kind.

### 3.1 The Value Function of the Data Owner

The value of the data owner increases with the level of the sales price. Since normalization also influences the price of the data set, an average value is formed from the ideal value from the data owner’s point of view and the data consumer’s maximum willingness to pay:

$$p_v = \frac{p_v^o + p_{max}}{2} \tag{12}$$

Furthermore, following Niyato et al. [3] and Liu et al. [4], the costs for aggregation and preparation of the data have to be considered. In these sources, these costs are always considered as constants. In this paper, however, we force quality-based pricing with a cost function to be normalized of the form  $k_i = kq_i$  with  $k_i$  for the cost and  $q_i$  for the quality of a dataset  $i$  [12]. Substituting the normalized price into the function  $u(x) = x^{1-r}$  and taking into account the costs, we obtain the value function of data owner  $i$  as a function of the price  $p_i$ :

$$u_i^o(p_i, q_i) = \left(\frac{p_i}{p_v}\right)^{1-r} - \frac{kq_i}{p_v} \tag{13}$$

### 3.2 The Value Function of the Data Consumer

The value of the data consumer is positively influenced by the quality of the data set and negatively influenced by the price. Using the value function of the quality determined from the survey values, the value function of the data consumer  $i$  is obtained:

$$u_i^c(p_i, q_i) = \gamma + \ln(\delta + \varepsilon q_i) - \frac{p_i}{p_v} \tag{14}$$

The parameters  $\gamma, \delta, \varepsilon$  serve to scale the empirically determined value function. The survey conducted for this paper resulted in the parameters  $\gamma = 0.8, \delta = 0.42, \varepsilon = 0.8$ . These values will be used in the further course.

## 4 Implementation of the Kalai-Smorodinsky Solution

According to KS, the negotiation problem between data owner and data consumer is solved with the point  $u^* = (u_i^{c*}; u_i^{o*})$  if applies to this point:

$$\frac{u_i^{o*} - c_2}{u_i^{c*} - c_1} = \frac{m_2 - c_2}{m_1 - c_1} \tag{15}$$

To obtain the coordinates of point  $m$ , a function is set up that describes the course of the boundary of the payoff space. The boundary of the bargaining space can be determined by the value functions of the data owner and data consumer. First, the value function of the data consumer is solved for  $p$ :

$$\frac{p_i}{p_v} = \gamma + \ln(\delta + \varepsilon q_i) - u_i^c \tag{16}$$

It follows by substituting into the value function of the data owner:

$$u_i^o(u_i^c) = (\gamma + \ln(\delta + \varepsilon q_i) - u_i^c)^{1-r} - \frac{kq_i}{p_v} \tag{17}$$

The straight line equation  $(c, m) = c_2 + \frac{m_2}{m_1} u^o$ , which runs through  $c = (c_1, c_2)$  and  $m = (m_1, m_2)$  can be determined by finding the coordinates of each point. The coordinates  $m_1$  and  $m_2$  of the point  $m = (m_1, m_2)$  can be determined by setting one of the two value functions to zero:

$$0 = (\gamma + \ln(\delta + \varepsilon q_i) - u_i^c)^{1-r} - \frac{kq_i}{p_v} \tag{18}$$

By rearranging the equation:

$$m_1 = u_i^c(u_i^o = 0) = \gamma + \ln(\delta + \varepsilon q_i) - \left(\frac{kq_i}{p_v}\right)^{\frac{1}{1-r}} \tag{19}$$

And analog for  $m_2$ :

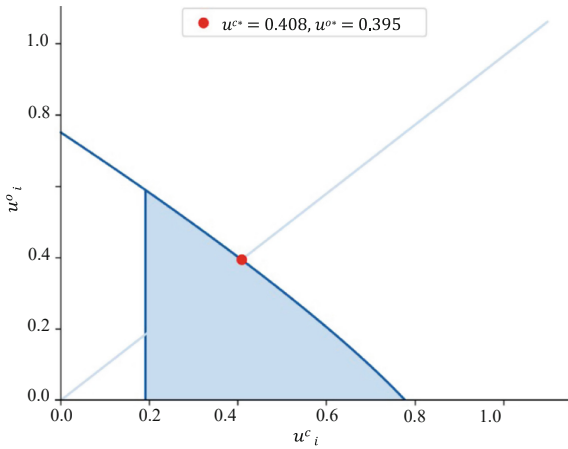
$$m_2 = u_i^o(u_i^c = 0) = (\gamma + \ln(\delta + \varepsilon q_i))^{1-r} - \frac{kq_i}{p_v} \tag{20}$$

For an exemplary calculation, a data quality of  $q_i = 0.8$  required by the Data Consumer, costs for data preparation and provision of  $k = 500 \text{ €}$  and a value assumption of  $p_v = 3000 \text{ €}$  are applied. Furthermore, it is applied  $p_{min} \leq p_i \leq p_{max}$  and  $p_{min} > k_i$ . The KS solution straight line equation is also known. The payoff space modeled in this way with solution  $u^*$  is shown in Fig. 6. A maximum willingness to pay of  $p_{max} = 2500 \text{ €}$  was chosen. The colored area represents the subset of the payoff space determined by the defined constraints. Here, irrational points (e.g., prices above  $p_{max}$ ) were considered for finding the solution of the bargaining game. Then, the solution was checked to see if it satisfied the defined constraints. This procedure reinforces the KS specific axiom of individual monotonicity as well as the fairness of the solution [18].

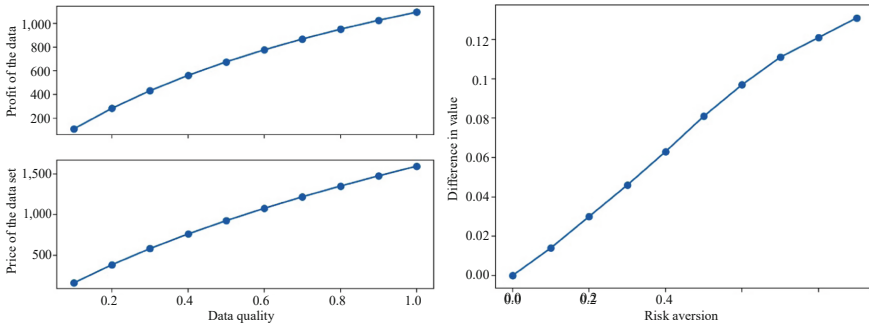
After the solution of the bargaining game is found, the value function of the data owner can be rearranged to  $p_i$  to determine the price. The result is:

$$p_i = \left(u_i^o + \frac{kq_i}{p_v}\right)^{\frac{1}{1-r}} p_v \tag{21}$$

From this follows with  $u_i^{o*} = 0.395$  and  $r_o \neq 1$  the price  $p_i^* = 1351.314 \text{ €}$ . After calculating the price for a fictitious dataset, the results of the negotiation when varying the parameters such as risk aversion of the data owner and quality of the dataset are considered in the following. The left side of Fig. 7 presents the price of a data set and the profit of the data owner depending on the quality of the data set. Both increase as a function of increasing quality, but level off due to the costs of data preparation, which also increase. The difference in value between data owner and data consumer at different levels of risk aversion of the data owner is plotted on the right side of Fig. 7. The lower the risk aversion of the data owner, the smaller the difference in value and the fairer the result.



**Fig. 6.** Graphical illustration of the KS solution



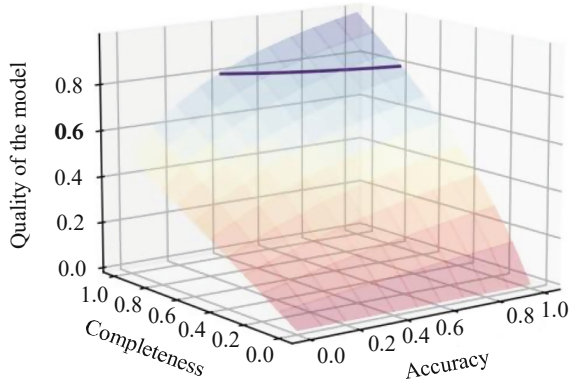
**Fig. 7.** Price and profit depending on data quality as well as the difference in value depending on different levels of risk aversion of the data owner

The dark blue line in Fig. 8 represents all combinations of QD accuracy and completeness with the overall data quality value  $q_i = 0.8$ . It can be seen that higher quality data are costly to produce and therefore have higher prices. For higher quality data, the dark blue line would become significantly shorter as the number of possible combinations of QD for a defined overall data quality value decreases.

### 5 Comparison to the Stackelberg Game

For a comparison of the two game-theoretic methods, data sharing must be modeled as a Stackelberg game. Except for the value functions, all parameters can be adopted. The Stackelberg game is solved by backward induction, so that the Stackelberg follower is considered first. The reason for the lack of suitability of the value function set up so far is explained on the basis of the data consumer:

$$\frac{\partial u_i^c}{\partial q_i} = \frac{\varepsilon}{\delta + \varepsilon q_i} = 0 \tag{22}$$



**Fig. 8.** All combinations of QD accuracy and completeness with the overall data quality value  $q_i = 0.8$

The problem of this function is the independence of the optimal response  $q_i^*$  of the Data Consumer from the pricing strategy of the Data Owner. Therefore, the equation must be modified so that the response function depends on the Data Owner’s price  $p_i$ :

$$u_i^o(p_i, q_i) = q_i \left( \left( \frac{p_i}{p_v} \right)^{1-r} - \frac{k}{p_v} \right) \tag{23}$$

$$u_i^c(p_i, q_i) = \gamma + \ln(\delta + \varepsilon q_i) - \frac{p_i}{p_v} q_i \tag{24}$$

With  $\frac{\partial u_i^c}{\partial q_i} = 0$ , the optimal response function of the data consumer now results with:

$$q_i^* = \frac{p_v}{p_i} - \frac{\delta}{\varepsilon} \tag{25}$$

By inserting the response function into the value function of the data owner and then deriving and searching for the zeros, the price  $p_i$  can be determined. For  $p_{min} = 500 \text{ €}$  and  $p_{max} = 2500 \text{ €}$  a price  $p^*$  of  $1142.3 \text{ €}$  follows. For the resulting quality, the reaction function of the data consumer is used. With the calculated pricing strategy,  $q_i^* = 0.72$ . For comparability with the KS solution, this must be formed at a data quality of  $q_i^* = 0.72$ , since this represents the optimal response of the Data Consumer to the Data Owner’s pricing strategy when using the Stackelberg game. The result is a price  $p^* = 1244.64 \text{ €}$ . Here the first advantage of the KS solution becomes apparent. The data consumer is able to purchase data with the desired quality. In the Stackelberg game, on the other hand, only a minimum quality requirement can be given. Thus, it is only conditionally suitable for quality-based pricing of data sets, since data sets have to be adjusted to the calculated quality after the price has been calculated. To experience the fairness of a negotiation, the concession of the players can be formed [19]:

$$\kappa_i = 1 - \frac{u_i}{\max u_i} \tag{26}$$

The concession using the KS solution is  $\kappa_i^c = 0.51$  for the data consumer and  $\kappa_i^o = 0.5$  for the data owner. Here, the maximum value of the KS solution is represented by the point  $m$ . In the case of the Stackelberg game, this results in a value for the Data Consumer  $u_i^c = 0.832$  with  $q_i = 1$  and  $p_i = p_{min}$ . The maximum value of the data owner is  $u_i^o = 0.345$  and is reached at  $q_i = q_{min}$  and  $p_i = p_{max}$ . Thus, the concessions are  $\kappa_i^c = 0.375$  and  $\kappa_i^o = 0.382$ . The difference in concessions is  $\Delta\kappa_i = 0.01$  when using the KS solution and  $\Delta\kappa_i = 0.07$  when using the Stackelberg game. Both methods are considered fair, since data consumer and data owner have to make similar concessions with respect to the outcome of the methods. In terms of price, the KS solution delivers a higher profit for the data owner.

## 6 Prospect

In this paper, a novel method and exemplary analysed in the field of industrial grinding data for pricing a dataset was presented which, compared to the Stackelberg game, offers the advantage that the quality of a dataset does not have to be assumed as a variable in quality-based pricing. This allows individual pricing for already existing data sets whose quality is not to be further adjusted. This allows SMEs with less available capital to acquire shares of datasets at a lower price and share in the economic benefits. Moreover, the data consumer can purchase data according to his needs. As a result of the KS solution, the price of a dataset can be determined for different quality levels. This concept enables the transferability of the procedure of a pricing of data of any equipment of the production engineering for the use case of the ML model of the linear regression. An extension of the methodology developed in this paper is to consider the data quantity when determining the price. The data quantity can be recorded either as a variable or as a constant. However, this increases the complexity of solving the negotiation game. On the other hand, data consumers with a low willingness to pay can also acquire data of high quality, since it is thus possible to buy shares of data sets. Furthermore, it has to be examined whether and in which dimensions the presented correlations change if other ML models are considered instead of regression models.

## References

1. Tiwana, A., Konsynski, B., Bush, A.: Research commentary—platform evolution: coevolution of platform architecture, governance, and environmental dynamics. *Inform. Syst. Res.* (2010). <https://doi.org/10.1287/isre.1100.0323>
2. Al-Ruithe, M., Benkhelifa, E., Hameed, K.: Data governance taxonomy: cloud vs. non-cloud. *Sustainability* (2018). <https://doi.org/10.3390/su10010095>
3. Niyato, D., Abu, M.A., Wang, P., Kim, D.I., Han, Z.: In: *IEEE International Conference on Communications (ICC)*, Kuala Lumpur (2016)
4. Liu, K., Qiu, X., Chen, W., Chen, X., Zheng, Z.: Optimal pricing mechanism for data market in blockchain-enhanced internet of things. *IEEE Internet Things J.* **6**(6), 9748–9761 (2019)
5. Holler, M.J.: *Einführung in die Spieltheorie*, 8. Auflage, Springer (2019), pp. 122 f., 199ff
6. Thon, M.P.: *First-Mover and Second-Mover Advantage under Uncertainty*. TU München (2018)

7. Rosenthal, R.: Games of perfect information, predatory pricing, and the chain store. *J. Econ. Theor.* **25**(1), 92–100 (1981)
8. Kalai, E., Smorodinsky, M.: Other solutions to Nash's bargaining problem. *Econometrica* **43**(3), 513–518 (1975)
9. Kessler, R., Gómez, J.M.: Implikationen von Machine Learning auf das Datenmanagement in Unternehmen. In: *HMD Praxis der Wirtschaftsinformatik: Ausgabe 5*, pp. 89–105. Springer (2020)
10. Zaveri, A., Rula, A., Maurino, A., Lehmann, J.: Quality assessment methodologies for linked open data. *Semantic Web J.* (2013)
11. Wang, R., Strong, M.: Beyond accuracy: what data quality means to data consumers. *J. Manage. Syst.* **12**(4), 5–33 (1996)
12. Yu, H., Zhang, M.: Data pricing strategy based on data quality. *Comput. Ind. Eng.* **112**, 1–10 (2017)
13. Sen, S., Joe-Wong, C., Ha, S., Chiang, M.: A survey of smart data pricing: past proposals, current plans, and future trends. *ACM Comput. Surv.* **46**(2), 1–37 (2013)
14. Lange, J., Stahl, F., Vossen, G.: Data marketplaces in different research disciplines: a review (2017). Retrieved from <https://doi.org/10.1007/s00287-017-1044-3>
15. Halter, A.N., Mason, R.: Value measurement for those who need to know. *Western J. Agric. Econ.* **3**(2), 99–109 (1978)
16. Wakker, P.P.: Explaining the characteristics of the power (CRRA) value family. *Health Econ.* **17**, 1329–1344 (2008)
17. Kimball Miles, S., Sahm Claudia, R., Shapiro, M.: Imputing risk tolerance from survey responses. *J. Am. Stat. Assoc.* **103**, 1028–1038 (2008)
18. Chen, M.A.: Individual monotonicity and the Leximin solution. *Econ. Theor.* 353–365 (2000)
19. Garcia, R.C., Contreras, J., De Lima Barbosa, M., De Silva Toledo, F.: Raiffa-Kalai-Smorodinsky Bargaining Solution for Bilateral Contracts in Electricity Markets (2020)